

## SPATIALLY INHOMOGENEOUS LINEAR INVERSE PROBLEMS WITH POSSIBLE SINGULARITIES

BY MARIANNA PENSKY<sup>1</sup>

*University of Central Florida*

The objective of the present paper is to introduce the concept of a spatially inhomogeneous linear inverse problem where the degree of ill-posedness of operator  $Q$  depends not only on the scale but also on location. In this case, the rates of convergence are determined by the interaction of four parameters, the smoothness and spatial homogeneity of the unknown function  $f$  and degrees of ill-posedness and spatial inhomogeneity of operator  $Q$ .

Estimators obtained in the paper are based either on wavelet–vaguelette decomposition (if the norms of all vaguelettes are finite) or on a hybrid of wavelet–vaguelette decomposition and Galerkin method (if vaguelettes in the neighborhood of the singularity point have infinite norms). The hybrid estimator is a combination of a linear part in the vicinity of the singularity point and the nonlinear block thresholding wavelet estimator elsewhere. To attain adaptivity, an optimal resolution level for the linear, singularity affected, portion of the estimator is obtained using Lepski [*Theory Probab. Appl.* **35** (1990) 454–466 and **36** (1991) 682–697] method and is used subsequently as the lowest resolution level for the nonlinear wavelet estimator. We show that convergence rates of the hybrid estimator lie within a logarithmic factor of the optimal minimax convergence rates.

The theory presented in the paper is supplemented by examples of deconvolution with a spatially inhomogeneous kernel and deconvolution in the presence of locally extreme noise or extremely inhomogeneous design. The first two problems are examined via a limited simulation study which demonstrates advantages of the hybrid estimator when the degree of spatial inhomogeneity is high. In addition, we apply the technique to recovery of a convolution signal transmitted via amplitude modulation.

### 1. Introduction.

1.1. *Formulation.* Let  $Q$  be a known linear operator on a Hilbert space  $H$  with inner product  $\langle \cdot, \cdot \rangle$ . The objective is to recover  $f \in H$  by observing

$$(1.1) \quad y(x) = (Qf)(x) + \sqrt{\varepsilon}W(x), \quad x \in \mathcal{X},$$

---

Received October 2012; revised August 2013.

<sup>1</sup>Supported in part by NSF Grant DMS-11-06564.

*MSC2010 subject classifications.* Primary 62C10, 65J10; secondary 62G05, 62G20.

*Key words and phrases.* Linear inverse problems, inhomogeneous, minimax convergence rates, singularity.

where  $W(x)$  is the white noise process and  $\sqrt{\varepsilon}$  is noise level. Assume that observations can be taken as functionals of  $y$

$$(1.2) \quad \langle y, g \rangle = \langle Qf, g \rangle + \sqrt{\varepsilon}\xi(g), \quad g \in H,$$

where  $\xi(g)$  is a Gaussian random variable with zero mean and variance  $\|g\|^2$  such that  $\mathbb{E}[\xi(g_1)\xi(g_2)] = \langle g_1, g_2 \rangle$ . In what follows,  $\|\cdot\|$  denotes the  $L^2$ -norm, all other norms are explicitly marked.

Model (1.1) is a common representation of a linear inverse problem with the Gaussian noise and has been studied by many authors [see, Abramovich and Silverman (1998), Bissantz et al. (2007), Cavalier and Golubev (2006), Cavalier et al. (2002), Cohen, Hoffmann and Reiß (2004), Hoffmann and Reiss (2008), Donoho (1995), Golubev (2010), Kalifa and Mallat (2003) and Mair and Ruymgaart (1996), among others]. A typical assumption in the problem above is that operator  $Q$  acts uniformly over the spaces of functions represented at a common scale, independently of the location of a function. In particular, consider a set of “test” functions  $\psi_{ha}(x) = h^{-1/2}\psi((x - a)/h)$  where  $\psi(x), x \in [0, 1]$ , has a bounded support  $(L_\psi, U_\psi)$  and unit  $L^2$ -norm  $\|\psi\| = 1$ . Then, functions  $\psi_{ha}(x)$  have scale  $h$ , supports concentrated around  $x = a$  and unit norms. Conditions which are commonly imposed on operator  $Q$  imply that it contracts the norms of all functions  $\psi_{ha}$  uniformly, that is, the value of  $\|Q\psi_{ha}\|$  depends considerably on the scale  $h$  but hardly at all on  $a$ . Moreover, if there exist  $(Q^*)^{-1}\psi_{ha}$ , where  $Q^*$  is the adjoint of operator  $Q$ , then values of  $\|(Q^*)^{-1}\psi_{ha}\|$  follow the same pattern. However, not all linear operators necessarily have those properties.

In order to illustrate the discussion above, consider linear operator  $Q$  with the adjoint  $Q^*$  given by

$$(1.3) \quad (Qf)(x) = \mu(x) \int_0^x f(t) dt, \quad (Q^*v)(x) = \int_x^1 \mu(z)v(z) dz,$$

where  $\mu(x)$  is a smooth function. Assume that function  $\psi$  is continuously differentiable and integrates to zero:  $\int_0^1 \psi(z) dz = 0$ . Denote  $\Psi(z) = \int_{L_\psi}^z \psi(x) dx$  and observe that  $\Psi(z) = 0$  whenever  $z \notin (L_\psi, U_\psi)$ . Then, direct calculations yield  $(Q\psi_{ha})(y) = h^{1/2}\mu(y)\Psi(\frac{y-a}{h})$ ,  $(Q^*)^{-1}\psi_{ha}(y) = -h^{-3/2}\mu^{-1}(y)\psi'(\frac{y-a}{h})$ , so that  $\|Q\psi_{ha}\|^2 = h^2 \int_{L_\psi}^{U_\psi} \mu^2(a + hz)\Psi^2(z) dz = h^2[\mu^2(a)\|\Psi\|^2 + o(1)]$  as  $h \rightarrow 0$  and  $\|(Q^*)^{-1}\psi_{ha}\|^2 = h^{-2} \int_{L_\psi}^{U_\psi} \mu^{-2}(a + hz)[\psi'(z)]^2 dz$ .

If  $\mu(y)$  is a constant or, at least,  $C_\mu^{-1} < \mu(y) < C_\mu$  for some relatively small  $C_\mu$ , then dependence of  $\|Q\psi_{ha}\|$  and  $\|(Q^*)^{-1}\psi_{ha}\|$  on  $a$  can be ignored, so equation (1.1) with  $Q$  given by (1.3) can be treated as a spatially homogeneous problem. However, if  $C_\mu$  is large, dependence on  $a$  becomes essential and equation (1.1) is a spatially inhomogeneous inverse problem.

Dependence on  $a$  becomes even more extreme if  $\mu(y)$  vanishes at some point  $x_0 \in (0, 1)$ , for example,  $\mu^2(x) = C_\alpha|x - x_0|^\alpha$ . Indeed, in this case,  $x_0$  is the *singularity point* and it is easy to show that  $\|Q\psi_{hx_0}\|^2 \asymp h^{2+\alpha}$  and

$$\|(Q^*)^{-1}\psi_{hx_0}\|^2 \asymp \begin{cases} h^{-(2+\alpha)}, & \text{if } \alpha < 1, \\ \infty, & \text{if } \alpha \geq 1. \end{cases}$$

Since wavelets provide an adequate tool for scale-location representations of functional spaces, it is convenient to introduce spatially inhomogeneous linear inverse problems using a wavelet–vaguelette decomposition proposed by Donoho (1995). In particular, in the case when  $H = L^2(\mathcal{D})$ ,  $\mathcal{D} \subset R$ , Donoho’s assumptions appear as follows:

(D1) There exist three sets of functions:  $\{\psi_{jk}\}$ , an orthonormal wavelet basis of  $H$ , and nearly orthogonal sets  $\{u_{jk}\}$  and  $\{v_{jk}\}$  such that  $Q\psi_{jk} = v_{jk}$ ,  $Q^*u_{jk} = \psi_{jk}$ ,  $\|v_{jk}\| = \lambda_j$ ,  $\|u_{jk}\| \asymp \lambda_j^{-1}$ , where  $\lambda_j$  depend on resolution index  $j$  but not on spatial index  $k$ .

(D2)  $u_{jk}$  and  $v_{jk}$  are such that  $\langle u_{j_1k_1}, v_{j_2k_2} \rangle = \delta_{j_1, j_2} \delta_{k_1, k_2}$ .

(D3) Sets  $\{u_{jk}\}$  and  $\{v_{jk}\}$  are nearly orthogonal, that is, for any sequence  $\{a_{jk}\} \in l^2$  one has

$$\left\| \sum_{j,k} a_{jk} \lambda_j u_{jk} \right\|^2 \asymp \sum_{j,k} a_{jk}^2, \quad \left\| \sum_{j,k} a_{jk} \lambda_j^{-1} v_{jk} \right\|^2 \asymp \sum_{j,k} a_{jk}^2.$$

Under conditions (D1)–(D3),  $f$  can be recovered using reproducing formula

$$(1.4) \quad f = \sum_{j,k} \langle Qf, u_{jk} \rangle \psi_{jk},$$

which is analogous to the reproducing formula for the SVD. Assumptions (D1)–(D3) are quite standard. Indeed, similar assumptions were introduced in Cavalier et al. (2002), Cavalier and Golubev (2006), Golubev (2010) and Knapik, van der Vaart and van Zanten (2011). The common premise is that operator  $Q$  acts “uniformly” over subspaces of  $H$ , so singular values or their surrogate equivalents depend on the resolution level only but not on location. If  $V_j = \text{Span}\{\psi_{i,k}, i \leq j, k \in \mathbb{Z}\}$  is the subspace of functions at resolution level  $j$ , the above assumptions reduce to a common assumption of Galerkin method [see, e.g., Cohen, Hoffmann and Reiß (2004) or Hoffmann and Reiss (2008)] that on subspace  $V_j$  operator  $Q$  has a bounded inverse with the norm dependent on  $j$  only, that is, there exist  $\lambda_j > 0$  such that

$$(1.5) \quad \sup_{j \geq 0} [\lambda_j \|Q^{-1}\|_{V_j \rightarrow Q^{-1}(V_j)}] < \infty,$$

which is very similar to combination of assumptions (D1) and (D3) above.

Note that both, assumptions (D1) and (1.5) imply that any function  $v \in V_j$  with  $\|v\| = 1$  has an inverse image, the norm of which is bounded by a constant which is

independent of the support of  $v$ . In this sense, operator  $Q$  is an ill-posed *spatially homogeneous* operator. In the present paper, we shall be interested in a different situation when assumptions (D1) and (D3) may not be true. In particular, we assume that the norms of the inverse images of  $\psi_{j,k}$  depend on the spatial index  $k$  and may be unbounded, that is, condition (D1) and possibly condition (D3) are violated. We shall refer to the such inverse linear problems as *spatially inhomogeneous* in comparison with spatially homogeneous problems which satisfy conditions (D1)–(D3) above.

1.2. *Motivation.* Spatially inhomogeneous ill-posed problems appear naturally in the case when either the noise level is spatially dependent or observations are irregularly spaced. Problems of this kind have been considered previously, both theoretically and in practical applications. Nevertheless, in former studies, it was always assumed that the noise level is uniformly bounded above or the design density of observations is bounded away from zero. The situations investigated in the present paper rather refer to *locally extreme noise* or *extremely inhomogeneous design* (which can be also described as *a local data loss*). Traditionally, in the first situation, measurements are treated as outliers and are removed from future analysis while the second one is dealt with using missing data techniques. Approach suggested in the present paper provides an alternative to those methodologies. Extreme noise or extremely inhomogeneous design occur in analysis of forensic data [see, e.g., Li and Satta (2011)] and deconvolution of LIDAR signals [see, e.g., Harsdorf and Reuter (2000) and Gurdev, Dreischuh and Stoyanov (2002)] or astronomical images [see, e.g., Starck and Pantin (2002)]. In addition, spatially inhomogeneous ill-posed problems arise whenever the kernel is spatially inhomogeneous, as, for example, in the case of the amplitude modulation. Below we consider some examples in more detail.

EXAMPLE 1 (Deconvolution of LIDAR signals). LIDAR (Light Detection And Ranging or Laser Imaging Detection And Ranging) is an optical remote sensing technology that can measure the distance to, or other properties of, targets by illuminating the target with laser light and analyzing the backscattered light. LIDAR technology has applications in archaeology, geography, geology, geomorphology, seismology, forestry, remote sensing, atmospheric physics. LIDAR data model is mathematically described by convolution equation  $P = R * P_\delta$  where  $P$  is the time-resolved LIDAR signal,  $P_\delta$  is the impulse response function and  $R$  is the system response function to be determined [see, e.g., Harsdorf and Reuter (2000) and Gurdev, Dreischuh and Stoyanov (2002)]. 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. This loss of precision becomes extreme when, for example, LIDAR is used to for emergency response and natural disaster management such as assessment of the

extent of damage due to volcanic eruptions or forest fires. In this situation, routinely, distances are calculated through filtering of the data set (removing outliers) and applying interpolation techniques. However, keeping all existing data and accounting for extreme noise may improve precision of the analysis of LIDAR signals.

**EXAMPLE 2 (Amplitude modulation).** Amplitude Modulation (AM) is a way of transmitting information in the form of electro-magnetic waves. In AM, a radio wave known as the “carrier” is modulated in amplitude by the signal, that is, to be transmitted, while the frequency remains constant [see, e.g., Miller, Vandome and McBrewster (2009)]. In video or image transmission (such as TV) where the base-band signal has inherent large bandwidth, AM is usually preferred to Frequency Modulation (FM) systems since the latter ones require additional bandwidth. Since in an AM, signal information is “stored” in amplitude which is affected by noise, AM is more susceptible to noise than FM. Mathematically, the problem reduces to multiplying the transmitted signal by the function  $\mu(x) = \cos(2\pi\omega x - \theta)$  with large  $\omega \approx n/2$  and  $\theta \in [0; 2\pi]$ . In Section 8.2, we provide an in-depth description of application of the methodology developed in the paper to recovery of a convolution signal transmitted via AM.

1.3. *Objectives and layout of the paper.* The objective of the present paper is to introduce the concept of a spatially inhomogeneous linear inverse problem which, to the best of the author’s knowledge, has never been considered previously in statistical framework. It turns out that spatially inhomogeneous problems exhibit properties which are very different from their spatially homogeneous counterparts. In particular, if the norms of vaguelettes  $u_{jk} = (Q^*)^{-1}\psi_{jk}$  are infinite in the vicinity of a singularity point, reproducing formula (1.4) cease working and the usual wavelet–vaguelette estimators cannot be applied. In this case, we propose a hybrid estimator which is based on combination of wavelet–vaguelette decomposition and Galerkin method. We study two application of the general theory, deconvolution with spatially inhomogeneous design and deconvolution with a spatially inhomogeneous kernel (the case of heterogeneous noise being a particular case of the latter).

Another interesting feature of the model is that the rates of convergence are determined by the interaction of four parameters, the smoothness and spatial homogeneity of the unknown function  $f$  and degrees of ill-posedness and spatial inhomogeneity of operator  $Q$ . In particular, if operator  $Q$  is weakly inhomogeneous, then the rates of convergence are not influenced by spatial inhomogeneity of operator  $Q$  and coincide with the rates which are usual for homogeneous linear inverse problems.

In what follows, we assume that operator  $Q$  in (1.1) is completely known. If, in practical applications, this is not true, one has to account for the extraneous errors which stem from the uncertainty in the operator  $Q$  by using, for example, ideas

of Hoffmann and Reiss (2008). Also, to simplify our considerations, we limit our study to the case when  $\mathcal{X} = [0, 1]$ ,  $H = L^2[0, 1]$  and  $k$  is a scalar. The theory presented below can be generalized to the case when  $H = L^2(\mathcal{D})$ ,  $\mathcal{D} \subset R^d$  and  $k$  is a  $d$ -dimensional vector. This extension should be relatively straightforward if one is dealing with isotropic Besov spaces but becomes much more interesting and involved in the case of anisotropic Besov spaces [see, e.g., Kerkyacharian, Lepski and Picard (2001)]. However, we leave those extensions for future investigations since considering them below will prevent us from focusing on the main objective of the paper.

The rest of the paper is organized as follows. Section 2 introduces the concept of a spatially inhomogeneous ill-posed problem and formulates major definitions and assumptions which are used throughout the paper. Section 3 presents the asymptotic minimax lower bounds for the  $L^2$ -risk of the estimators of the solution of the problem over a wide range of Besov balls. Section 4 talks about estimation strategies, in particular, about partitioning the unknown response function  $f$  and its estimator into the singularity-affected and the singularity-free parts, the main idea at the core of the hybrid estimator. Section 5 elaborates on the risk of the estimator constructed in the previous section when the lowest resolution level in the zero-affected portion of the estimator is fixed. Section 6 discusses the adaptive choice of the lowest resolution level resolution level and derives the asymptotic minimax upper bounds for the  $L^2$ -risk. In Section 7, we consider two examples of spatially inhomogeneous ill-posed problems, deconvolution with the spatially inhomogeneous operator (Section 7.1) which can be viewed as a version of a deconvolution equation with spatially inhomogeneous noise, and deconvolution based on irregularly spaced sample (Section 7.2). Section 8 presents a limited simulation study of deconvolution with heteroscedastic noise and also studies application of the hybrid estimator to recovery of a convolution signal transmitted via amplitude modulation. Section 9 concludes the paper with a discussion. Proofs of the statements are contained in the supplementary material [Pensky (2013)].

**2. Spatially inhomogeneous ill-posed problem: Assumptions and definitions.** Consider a scaling function  $\varphi$  and a corresponding wavelet  $\psi$  with bounded supports and form an orthonormal wavelet basis  $\{\psi_{jk}\}$  of  $L^2([0, 1])$ . We further impose the following set of assumptions on spatially inhomogeneous operator  $Q$ .

(A1) There exist functions  $\{u_{jk}\}$  and  $\{v_{jk}\}$  such that  $Q\psi_{jk} = v_{jk}$ ,  $Q^*u_{jk} = \psi_{jk}$ , where  $\|v_{jk}\| = \lambda_{jk} < \infty$ .

(A2) There exists a *singularity* point  $x_0 \in (0, 1)$  and a constant  $D \geq 0$  such that  $\|u_{jk}\| = \infty$  if  $|k - k_{0j}| < D$  and, for any  $j$  and any  $\{a_{jk}\}$ ,  $k = 0, \dots, 2^j - 1$ , one has

$$(2.1) \quad \left\| \sum_{|k-k_{0j}| \geq D} a_{jk} \lambda_{jk} u_{jk} \right\|^2 \leq C_u \sum_{|k-k_{0j}| \geq D} a_{jk}^2,$$

where  $C_u < \infty$  is independent of  $j$  and  $k_{0j} = 2^j x_0$  is the parameter corresponding to location  $x_0$  ( $k_{0j}$  is not necessarily an integer).

(A3) Functions  $v_{jk}$  are such that, for any  $j$ , inequality

$$(2.2) \quad \left\| \sum_{k=0}^{2^j-1} a_{jk} \lambda_{jk}^{-1} v_{jk} \right\|^2 \leq C_v \sum_{k=0}^{2^j-1} a_{jk}^2$$

holds for any  $\{a_{jk}\}, k = 0, \dots, 2^j - 1$ , where  $C_v < \infty$  is independent of  $j$ .

Note that assumptions (A1)–(A3) are weaker than assumptions (D1)–(D3) above. First,  $\lambda_{jk}$  depends not only on resolution level but also on location of the wavelet coefficient. Also, if  $D > 0$ , then, in the neighborhood of the singularity point  $x_0$ , wavelet coefficients cannot be recovered directly since  $\|u_{jk}\| = \infty$ , and we say that operator  $Q$  has a *singularity* at  $x_0$ .

Since one usually start wavelet expansion at some finite resolution level  $m$ , we need an extra assumption which mirrors assumption (A2) and can be derived from it:

(A4) There exist functions  $\{t_{mk}\}$  and positive constants  $C_t$  and  $D_0$  independent of  $m$  such that, for any  $\{a_{mk}\}, k = 0, \dots, 2^m - 1$ ,

$$(2.3) \quad Q^* t_{mk} = \varphi_{mk}, \quad \left\| \sum_{|k-k_{0j}| \geq D_0} a_{mk} \lambda_{mk} t_{mk} \right\|^2 \leq C_t \sum_{|k-k_{0j}| \geq D_0} a_{mk}^2.$$

If  $D = D_0 = 0$  in assumptions (A2) and (A4), then  $\|u_{jk}\| < \infty$  and  $\|t_{mk}\| < \infty$  for any  $k$ . Hence,  $f$  can be expressed using reproducing formula (1.4) which, in this case, becomes

$$(2.4) \quad f(x) = \sum_{k=0}^{2^m-1} a_{mk} \varphi_{mk}(x) + \sum_{j=m}^{\infty} \sum_{k=0}^{2^j-1} b_{jk} \psi_{jk}(x),$$

where  $a_{mk} = \langle Qf, t_{mk} \rangle$  and  $b_{jk} = \langle Qf, u_{jk} \rangle$ . If  $D > 0$ , reproducing formula (2.4) cease working and one needs an alternative solution to recovering  $f$ . Indeed, if  $Qf$  in expressions for  $a_{mk}$  and  $b_{jk}$  is replaced by  $y = Qf + \varepsilon W$ , then the variances of the wavelet coefficients in the vicinity of singularity  $x_0$  are infinite:  $\text{Var}\langle y, u_{jk} \rangle = \infty$  if  $|k - k_{0j}| < D$  and similar consideration applies to  $t_{mk}$ . For this reason, at each resolution level, we partition the set of all indices into the *singularity-affected* indices

$$K_{0m} = \{k = 0, \dots, 2^m - 1 : |k - k_{0m}| < D_0\},$$

$$K_{1j} = \{k = 0, \dots, 2^j - 1 : |k - k_{0j}| < D\}$$

and the *singularity-free* indices

$$K_{0m}^c = \{k : 0 \leq k \leq 2^m - 1, k \notin K_{0m}\},$$

$$K_{1j}^c = \{k : 0 \leq k \leq 2^j - 1, k \notin K_{1j}\}.$$

To be specific, in what follows, we assume that  $\lambda_{jk}$  are such that, for some positive constants  $\alpha, \beta, C_{\lambda_0}$  and  $C_\lambda$  independent of  $j$  and  $k$ , one has

$$(2.5) \quad C_{\lambda_0} 2^{-j(\alpha+\beta)} (1 + |k - k_{0j}|)^\alpha \leq \lambda_{jk}^2 \leq C_\lambda 2^{-j(\alpha+\beta)} (1 + |k - k_{0j}|)^\alpha.$$

We shall refer to coefficients  $\beta$  and  $\alpha$  in (2.5) as degrees of *ill-posedness* and *spatial inhomogeneity*, respectively. Observe that with  $\lambda_{jk}$  satisfying condition (2.5), the variances of the coefficients at the lower resolution levels may be significantly higher than the variances of the coefficients at higher resolution levels as long as the locations of the lower resolution level coefficients lie in a close proximity of a singularity point.

In the present paper, we consider estimation of a solution of inhomogeneous linear inverse problems in the case when the unknown function  $f$  is possibly spatially inhomogeneous itself, in particular,  $f$  belongs to a Besov ball  $B_{p,q}^s(A)$  of radius  $A$ . Interplay between spatial inhomogeneity of operator  $Q$  and properties of  $f$  lead to various very interesting phenomena. In particular, if  $\alpha$  is small or  $p$  and  $\beta$  are relatively large, spatial inhomogeneity does not affect convergence rates and  $f$  can be recovered as well as in the case of  $\alpha = 0$ .

REMARK 1 (Multiple singularity points). Note that one can consider a spatially inhomogeneous problems with multiple singularity points  $x_{0,1} < x_{0,2} < \dots < x_{0,L}$  and corresponding constants  $D_1, \dots, D_L$  where  $L < \infty$  and  $x_{0,i} - x_{0,i-1} \geq \delta > 0$  for some fixed positive  $\delta$ . The theory developed below can be easily extended to this case, with the convergence rates of the estimators determined by the “worst case scenario” among singular points  $x_{0,i}, i = 1, \dots, L$ .

**3. Minimax lower bounds for the risk over Besov balls.** Before constructing an estimator of the unknown function  $f$  under model (1.1), we derive the asymptotic minimax lower bounds for the  $L^2$ -risk over a wide range of Besov balls.

Recall that for an  $r_0$ -regular multiresolution analysis [see, e.g., Meyer (1992), pp. 21–25], with  $0 < s < r_0$ , and for a Besov ball  $B_{p,q}^s(A) = \{f \in L^p([0, 1]) : f \in B_{p,q}^s, \|f\|_{B_{p,q}^s} \leq A\}$ , of radius  $A > 0$  with  $1 \leq p, q \leq \infty$  and  $s' = s + 1/2 - 1/p$ , one has

$$\|f\|_{B_{p,q}^s} = \left( \sum_{k=0}^{2^m-1} |a_{mk}|^p \right)^{1/p} + \left( \sum_{j=m}^{\infty} 2^{js'q} \left( \sum_{k=0}^{2^j-1} |b_{jk}|^p \right)^{q/p} \right)^{1/q}$$

with respective sum(s) replaced by maximum if  $p = \infty$  and/or  $q = \infty$  [see, e.g., Johnstone et al. (2004)].

In what follows, we use the symbol  $C$  for a generic positive constant, which takes different values at different places and is independent of the noise level  $\varepsilon$ . The following statement provides the asymptotic minimax lower bounds for the  $L^2$ -risk over Besov balls  $B_{p,q}^s(A)$ .



**THEOREM 1.** *Let  $1 \leq p, q \leq \infty$  and  $s > \max(1/p, 1/2)$ . Then, under assumptions (A1)–(A3), as  $\varepsilon \rightarrow 0$ ,*

$$(3.1) \quad R_\varepsilon(B_{p,q}^s(A)) = \inf_{\tilde{f}} \sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\tilde{f} - f\|^2 \geq C \Delta(\varepsilon),$$

where the infimum is taken over all possible square-integrable estimators  $\tilde{f}$  of  $f$  based on  $y$  from model (1.1) and

$$(3.2) \quad \Delta(\varepsilon) = \begin{cases} A^{(2(\alpha+\beta))/(2s'+\alpha+\beta)} \varepsilon^{2s'/(2s'+\alpha+\beta)}, \\ \text{if } 2s(\alpha - 1) \geq (\beta + 1)(1 - 2/p), \\ A^{(2(\beta+1))/(2s+\beta+1)} \varepsilon^{2s/(2s+\beta+1)}, \\ \text{if } 2s(\alpha - 1) < (\beta + 1)(1 - 2/p). \end{cases}$$

**REMARK 2 (Convergence rates).** As we show below, the minimax global convergence rates in Theorem 1 are attainable up to a logarithmic factor. The rates are determined by the interaction of four parameters,  $s$ ,  $p$ ,  $\alpha$  and  $\beta$ . Parameters  $s$  and  $p$  describe, respectively, smoothness and spatial homogeneity of the unknown function  $f$ , while  $\beta$  and  $\alpha$ , defined in (2.5), are referred to as degrees of ill-posedness and spatial inhomogeneity of operator  $Q$ . If the value of  $\alpha$  is large, in particular,  $2s\alpha > 2s' + \beta(1 - 2/p)$ , convergence rate is significantly affected by the degree of spatial inhomogeneity  $\alpha$  of  $Q$ . On the other hand, if  $2s\alpha < 2s' + \beta(1 - 2/p)$ , spatial inhomogeneity of operator  $Q$  does not affect convergence rate which is determined entirely by the degree of ill-posedness  $\beta$ .

**4. Estimation strategies in the presence of a singularity.** To be more specific, consider a periodized version of the wavelet basis on the unit interval

$$(4.1) \quad \{\varphi_{mk}, \psi_{jk} : j \geq m, k = 0, 1, \dots, 2^j - 1\},$$

where  $\varphi_{mk}(x) = 2^{m/2} \varphi(2^m x - k)$ ,  $\psi_{jk}(x) = 2^{j/2} \psi(2^j x - k)$ ,  $x \in [0, 1]$ . Note that the latter requires that the resolution level  $m$  is high enough, in particular,  $m \geq m_1$ , where  $m_1$  is such that

$$(4.2) \quad 2^{m_1} > \max(\mathcal{L}_{\varphi^*}, \mathcal{L}_{\psi^*}).$$

Here,  $\mathcal{L}_{\varphi^*}$  and  $\mathcal{L}_{\psi^*}$  are the lengths of supports of the mother and father wavelets,  $\varphi^*$  and  $\psi^*$ , that generate periodized wavelet basis. Then, for any  $m \geq m_1$ , the set (4.1) forms an orthonormal wavelet basis for  $L^2([0, 1])$  and, hence, any  $f \in L^2([0, 1])$ , can be expanded using formula (2.4). Under assumptions (A1), (A2) and (A4), one can construct unbiased estimators of coefficients  $a_{mk}$  and  $b_{jk}$

$$(4.3) \quad \hat{a}_{mk} = \langle y, t_{mk} \rangle, \quad \hat{b}_{jk} = \langle y, u_{jk} \rangle.$$

If  $k \in K_{0m}^c$  and  $k \in K_{1j}^c$ , respectively, then estimators  $\hat{a}_{mk}$  and  $\hat{b}_{jk}$  have finite variances

$$(4.4) \quad \begin{aligned} \text{Var}(\hat{a}_{mk}) &\asymp \lambda_{mk}^{-2}, & k \in K_{0m}^c, \\ \text{Var}(\hat{b}_{jk}) &\asymp \lambda_{jk}^{-2}, & k \in K_{1j}^c \end{aligned}$$

and have infinite variances otherwise. In order to account for the latter, for any  $m \geq m_1$ , we partition  $f$  into the sum of singularity-affected and singularity-free parts

$$f(x) = f_{0,m}(x) + f_{c,m}(x), \quad x \in [0, 1],$$

where

$$(4.5) \quad f_{0,m}(x) = \sum_{k \in K_{0m}} a_{mk} \varphi_{mk}(x) + \sum_{j=m}^{\infty} \sum_{k \in K_{1j}} b_{jk} \psi_{jk}(x), \quad x \in [0, 1],$$

$$(4.6) \quad f_{c,m}(x) = \sum_{k \in K_{0m}^c} a_{mk} \varphi_{mk}(x) + \sum_{j=m}^{\infty} \sum_{k \in K_{1j}^c} b_{jk} \psi_{jk}(x), \quad x \in [0, 1].$$

We then construct estimators  $\hat{f}_{0,m}$  and  $\hat{f}_{c,m}$  of  $f_{0,m}$  and  $f_{c,m}$ , respectively, and estimate  $f$  by a hybrid estimator

$$(4.7) \quad \hat{f}_m(x) = \hat{f}_{0,m}(x) + \hat{f}_{c,m}(x), \quad x \in [0, 1].$$

In particular, we shall use a linear estimator with the resolution level  $m$  estimated from the data as  $\hat{f}_{0,m}$  and a nonlinear block thresholding wavelet estimator as  $\hat{f}_{c,m}$ , where the lowest resolution level  $m$  in  $\hat{f}_{c,m}$  is determined by the linear part  $\hat{f}_{0,m}$ . In what follows, we shall consider estimation of  $f_{0,m}$  and  $f_{c,m}$  separately.

First, we construct a block thresholding wavelet estimator  $\hat{f}_{c,m}$  of  $f_{c,m}$ . For this purpose, we divide the wavelet coefficients at each resolution level into  $l_{j\varepsilon}^L$  blocks of length  $\ln(\varepsilon^{-1})$  to the left of  $(k_{0j} - D)$  and  $l_{j\varepsilon}^R$  blocks to the right of  $(k_{0j} + D)$ , where

$$(4.8) \quad \begin{aligned} l_{j\varepsilon}^L &= (k_{0j} - D) / \ln(\varepsilon^{-1}), \\ l_{j\varepsilon}^R &= (2^j - D - 1 - k_{0j}) / \ln(\varepsilon^{-1}), \\ l_{j\varepsilon} &= \max(l_{j\varepsilon}^L, l_{j\varepsilon}^R). \end{aligned}$$

Define blocks  $U_{jl}^L$  and  $U_{jl}^R$  of indices  $k$  to the left of  $(k_{0j} - D)$  and to the right of  $(k_{0j} + D)$ , respectively, as

$$\begin{aligned} U_{jl}^L &= \{k : -D - l \ln(\varepsilon^{-1}) < k - k_{0j} \leq -D - (l - 1) \ln(\varepsilon^{-1})\}, & l \in U_j^L, \\ U_{jl}^R &= \{k : D + (l - 1) \ln(\varepsilon^{-1}) < k - k_{0j} \leq D + l \ln(\varepsilon^{-1})\}, & l \in U_j^R, \end{aligned}$$

where

$$\begin{aligned}
 U_j^L &= \{l : 1 \leq l \leq l_{j\varepsilon}^L\}, \\
 U_j^R &= \{l : 1 \leq l \leq l_{j\varepsilon}^R\}, \\
 U_j &= U_j^L \cup U_j^R.
 \end{aligned}
 \tag{4.9}$$

To simplify the narrative, we shall write  $l \in U_j$  and  $k \in U_{jl}$  without a specific reference whether a block lies to the right or to the left of  $k_{0j}$ . Denote

$$B_{jl} = \sum_{k \in U_{jl}} b_{jk}^2, \quad \widehat{B}_{jl} = \sum_{k \in U_{jl}} \widehat{b}_{jk}^2,
 \tag{4.10}$$

$$R_{jl\varepsilon} = \frac{\varepsilon \ln(\varepsilon^{-1}) 2^j(\alpha+\beta)}{|D + (l - 1) \ln(\varepsilon^{-1})|^\alpha} \asymp \varepsilon \sum_{k \in U_{jl}} \lambda_{jk}^{-2}.
 \tag{4.11}$$

For any  $m \geq m_1$ , estimate  $f_{c,m}$  by

$$\begin{aligned}
 \widehat{f}_{c,m}(x) &= \sum_{k \in K_{0m}^c} \widehat{a}_{mk} \varphi_{mk}(x) \\
 &+ \sum_{j=m}^{J-1} \sum_{l \in U_j} \sum_{k \in U_{jl}} \widehat{b}_{jk} \mathbb{I}(\widehat{B}_{jl} \geq \tau^2 R_{jl\varepsilon}) \psi_{jk}(x),
 \end{aligned}
 \tag{4.12}$$

where  $\mathbb{I}(\Omega)$  is the indicator function of the set  $\Omega$ , the value of  $\tau$  will be defined later and

$$2^J = \varepsilon^{-2/(\alpha+\beta+2)}.
 \tag{4.13}$$

Now, consider estimation of the singularity-affected part. Since the estimators  $\widehat{a}_{mk}$  of  $a_{mk}$ , given in (4.3), have infinite variances when  $k \in K_{0m}$ , we estimate those coefficients by solving a system of linear equations. Denote  $w_{mk} = Q\varphi_{mk}$  and observe that, for a given  $m$ ,  $m_1 \leq m \leq J - 1$ , one has  $f = f_m + R_m$ . Here

$$\begin{aligned}
 f_m &= \sum_{k \in K_{0m}} a_{mk} \varphi_{mk} + \sum_{k \in K_{0m}^c} a_{mk} \varphi_{mk}, \\
 R_m &= \sum_{j=m}^{\infty} \sum_{k=0}^{2^j-1} b_{jk} \psi_{jk}
 \end{aligned}
 \tag{4.14}$$

and, hence,

$$Qf = \sum_{k \in K_{0m}} a_{mk} w_{mk} + \sum_{k \in K_{0m}^c} a_{mk} w_{mk} + QR_m.
 \tag{4.15}$$

Taking scalar products of both sides of (4.15) with  $w_{ml}, l \in K_{0m}$ , obtain

$$(4.16) \quad \begin{aligned} \langle w_{ml}, Qf \rangle &= \sum_{k \in K_{0m}} a_{mk} \langle w_{ml}, w_{mk} \rangle \\ &+ \sum_{k \in K_{0m}^c} a_{mk} \langle w_{ml}, w_{mk} \rangle + \langle w_{ml}, QR_m \rangle, \quad l \in K_{0m}. \end{aligned}$$

Introduce matrices  $\mathbf{A}^{(m)}$  and  $\mathbf{B}^{(m)}$  and vectors  $\mathbf{c}^{(m)}, \hat{\mathbf{c}}^{(m)}, \mathbf{r}^{(m)}, \mathbf{z}^{(m)}, \mathbf{h}^{(m)}$  and  $\hat{\mathbf{h}}^{(m)}$  with elements

$$(4.17) \quad A_{lk}^{(m)} = \langle w_{ml}, w_{mk} \rangle, \quad B_{lv}^{(m)} = \langle w_{ml}, w_{mv} \rangle,$$

$$(4.18) \quad c_l^{(m)} = \langle w_{ml}, Qf \rangle, \quad \hat{c}_l^{(m)} = \langle w_{ml}, y \rangle,$$

$$(4.19) \quad r_l^{(m)} = \langle w_{ml}, QR_m \rangle, \quad z_k^{(m)} = a_{mk},$$

$$(4.20) \quad h_v^{(m)} = a_{mv}, \quad \hat{h}_v^{(m)} = \hat{a}_{mv} = \langle y, t_{mv} \rangle,$$

where  $k, l \in K_{0m}, v \in K_{0m}^c$  and  $\hat{a}_{mk}$  with  $k, l \in K_{0m}$ , are defined in (4.3). Then, one can rewrite an exact system of linear equations (4.16) as  $\mathbf{c}^{(m)} = \mathbf{A}^{(m)}\mathbf{z}^{(m)} + \mathbf{B}^{(m)}\mathbf{h}^{(m)} + \mathbf{r}^{(m)}$  and obtain its approximate version

$$(4.21) \quad \hat{\mathbf{c}}^{(m)} = \mathbf{A}^{(m)}\hat{\mathbf{z}}^{(m)} + \mathbf{B}^{(m)}\hat{\mathbf{h}}^{(m)}.$$

Since matrix  $\mathbf{A}^{(m)}$  is a nonnegative definite matrix of a finite size, in order to guarantee that it is nonsingular, it is sufficient to impose the following almost trivial assumption:

(A5) Functions  $w_{mk} = Q\varphi_{mk}, k \in K_{0m}$  are linearly independent.

Under assumption (A5), one has

$$(4.22) \quad \begin{aligned} \mathbf{z}^{(m)} &= (\mathbf{A}^{(m)})^{-1}(\mathbf{c}^{(m)} - \mathbf{B}^{(m)}\mathbf{h}^{(m)} - \mathbf{r}^{(m)}), \\ \hat{\mathbf{z}}^{(m)} &= (\mathbf{A}^{(m)})^{-1}(\hat{\mathbf{c}}^{(m)} - \mathbf{B}^{(m)}\hat{\mathbf{h}}^{(m)}). \end{aligned}$$

Finally, for a given  $m$ , we set  $\hat{a}_{mk} = \hat{z}_k^{(m)}, k \in K_{0m}$  and estimate  $f_{0,m}$  by the following wavelet linear estimator

$$(4.23) \quad \hat{f}_{0,m}(x) = \sum_{k \in K_{0m}} \hat{a}_{mk} \varphi_{mk}(x), \quad x \in [0, 1].$$

REMARK 3 (Relation to nonparametric regression estimation based on spatially inhomogeneous data). We need to touch upon the relationship between the present paper and the paper by Antoniadis, Pensky and Sapatinas (2013) which considered nonparametric regression estimation based on irregularly spaced data, in particular, in the case when design density has zeros. The latter problem is the well-known formulation which has been already studied extensively [see, e.g.,

Gaïffas (2005, 2007a, 2007b, 2009)] and, indeed, can be considered as a trivial case of deconvolution with the spatially inhomogeneous design studied in Section 7.2 with  $Q$  being an identity operator. For this reason, the hybrid estimator was proposed in Antoniadis, Pensky and Sapatinas (2013). However, due to the fact that, in the regression set up, one observes function  $f$  directly, construction of the hybrid estimator is much more involved in the case of an inverse problem than in the case of nonparametric regression. In addition, the present paper provides the implementation of the hybrid estimator and studies its performance via simulations which has never been done previously since Antoniadis, Pensky and Sapatinas (2013) considered only theoretical construction of the hybrid estimator.

**5. The risks of the estimators of the singularity-free and the singularity-affected parts.** In this section, we shall provide asymptotic expressions for the risks of estimators (4.12) and (4.23) when resolution level  $m$  is a fixed, nonrandom quantity possibly dependent on  $\varepsilon : m = m(\varepsilon)$ .

Let us first construct an asymptotic upper bound for the singularity-free portion (4.12) of the estimator. Denote

$$(5.1) \quad \lambda_m^{-2} = \sum_{k \in K_{0m}^c} \lambda_{mk}^{-2}$$

and observe that, under condition (2.5), there exist positive constants  $C_{\lambda\alpha 0}$  and  $C_{\lambda\alpha}$  independent of  $m$  such that

$$(5.2) \quad C_{\lambda\alpha 0} 2^{m(\beta + \max(1, \alpha))} m^{\mathbb{I}(\alpha=1)} \leq \lambda_m^{-2} \leq C_{\lambda\alpha} 2^{m(\beta + \max(1, \alpha))} m^{\mathbb{I}(\alpha=1)}.$$

LEMMA 1. *Let  $1 \leq p, q \leq \infty$ ,  $s \geq \max(1/2, 1/p)$ ,  $s^* = \min(s, s')$  and assumptions (A1)–(A4) hold. Let  $\hat{f}_{c,m}$  be given by (4.12) where the nonrandom quantity  $m = m(\varepsilon)$  is such that  $m_1 \leq m \leq J - 1$ , with  $J$  defined in (4.13). If*

$$(5.3) \quad \tau^2 = 4C_u C_\lambda (\sqrt{2}\chi + 1)^2 \quad \text{and} \quad \chi^2 \geq \frac{4(\beta + \max(1, \alpha))}{2 + \alpha + \beta},$$

where  $C_u$  and  $C_\lambda$  are defined in (2.1) and (2.5), respectively, then, as  $\varepsilon \rightarrow 0$ ,

$$(5.4) \quad \sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{c,m} - f_{c,m}\|^2 \leq C(\lambda_m^{-2} \varepsilon + \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho).$$

Here  $\Delta(\varepsilon)$  is defined in (3.2),

$$(5.5) \quad \rho = \begin{cases} \mathbb{I}\left(\frac{2s}{1 - 2/p} = \frac{\beta + 1}{\alpha - 1}\right) + \frac{(1 - \alpha)(2 - p)}{2 - \alpha p} \mathbb{I}(\alpha < 1) \mathbb{I}(p < 2), & \text{if } \alpha \neq 1, \\ (2s^* + \beta + 1)^{-1} 2s^*, & \text{if } \alpha = 1 \end{cases}$$

and  $\mathbb{I}$  is the indicator function. Moreover, as  $\varepsilon \rightarrow 0$ ,

$$(5.6) \quad \sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{c,m} - f_{c,m}\|^4 = o(\varepsilon^{-2}).$$

Now, we find upper bounds for the singularity-affected portion of the estimator  $\hat{f}_{0,m}(x)$ . Recall that  $w_{mk} = Q\varphi_{mk}$  and let  $\rho_m$  be such that

$$(5.7) \quad C_{w1}\rho_m \leq \|w_{mk}\| \leq C_{w2}\rho_m \quad \text{if } k \in K_{0m}.$$

Note that, since the set  $K_{0m}$  contains at most  $(2D_0)$  indices,  $\rho_m$  satisfying condition (5.7) can always be found. The advantage of using the system of equations (4.21) rests upon the fact that matrix  $\mathbf{A}^{(m)}$  is a finite-dimensional positive definite matrix with all eigenvalues of order  $\rho_m^2$ . In particular,

$$(5.8) \quad \|\mathbf{A}^{(m)}\| \leq C_{A1}\rho_m^2, \quad \|(\mathbf{A}^{(m)})^{-1}\| \leq C_{A2}\rho_m^{-2}$$

for some positive constants  $C_{A1}$  and  $C_{A2}$  independent of  $m$ , as it is shown in the proof of the following lemma which states the rate of convergence of the singularity-affected portion of the estimator.

LEMMA 2. *Let  $1 \leq p, q \leq \infty$  and  $s \geq \max(1/2, 1/p)$ . Let assumptions (A1)–(A5) hold and there exists  $C_{\rho\lambda}$ ,  $0 < C_{\rho\lambda} < \infty$ , independent of  $m$ , such that*

$$(5.9) \quad \rho_m^2 \geq C_{\rho\lambda}\lambda_m^2,$$

where  $\lambda_m$  and  $\rho_m$  are defined in (5.1) and (5.7), respectively. Let  $\alpha \geq 1$  and also

$$(5.10) \quad \rho_m^{-4} \max_{l \in K_{0m}} \sum_{k \in K_{0m}^c} \lambda_{mk}^{-2} \langle w_{ml}, w_{mk} \rangle^2 \leq K_1 \lambda_m^{-2},$$

$$(5.11) \quad \max_{l \in K_{0m}} \sum_{j=m}^{\infty} \sum_{k=0}^{2^j-1} |\langle w_{ml}, v_{jk} \rangle| \leq K_2 \rho_m^2$$

for some absolute constants  $K_1$  and  $K_2$  independent of  $m$ . Let estimator  $\hat{f}_{0,m}$  of  $f_{0,m}$  be given by (4.23). Then, for any  $m$ ,  $m_1 \leq m \leq J - 1$ , and some constant  $C$  independent of  $m$  and  $\varepsilon$ , as  $\varepsilon \rightarrow 0$ , one has

$$(5.12) \quad \sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{0,m} - f_{0,m}\|^2 \leq C(\varepsilon \lambda_m^{-2} + 2^{-2ms'}),$$

$$(5.13) \quad \sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{0,m} - f_{0,m}\|^4 = o(\varepsilon^{-2}).$$

Note that in order  $\hat{f}_m = \hat{f}_{c,m} + \hat{f}_{0,m}$  estimates  $f$ , one needs to start the estimator  $\hat{f}_{c,m}$  in (4.12) at exactly the same resolution level at which the linear estimator  $\hat{f}_{0,m}$  in (4.23) is constructed. Thus, the choice of the lowest resolution level in (4.12) is driven by the choice of  $m$  in (4.23).

Let  $m_0$  be such that

$$(5.14) \quad 2^{m_0} = (\varepsilon [\ln(\varepsilon^{-1})]^{\mathbb{I}(\alpha=1)})^{-1/(2s'+\alpha+\beta)},$$

so that, for  $\alpha \geq 1$ , one has

$$(5.15) \quad \varepsilon \lambda_{m_0}^{-2} + 2^{-2m_0 s'} \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho.$$

The following statement delivers the total squared risk of the estimator (4.7) of  $f$  if  $DD_0 > 0$ .

**THEOREM 2.** *Let  $1 \leq p, q \leq \infty$  and  $s \geq \max(1/2, 1/p)$ . Let conditions (5.9)–(5.11) and assumptions (A1)–(A5) hold with  $\alpha \geq 1$  and  $DD_0 > 0$  in assumptions (A2) and (A4). Consider estimator (4.7) of  $f$  where  $\hat{f}_{c,m}$  and  $\hat{f}_{0,m}$  are given by formulae (4.12) and (4.23), respectively. Let  $m = m_0$  where  $m_0$  is defined in (5.14),  $J$  be defined in (4.13) and let positive constants  $\tau$  and  $\chi$  satisfy condition (5.3). Then, as  $\varepsilon \rightarrow 0$ , one has*

$$(5.16) \quad \sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{m_0} - f\|^2 \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho,$$

where  $\Delta(\varepsilon)$  and  $\rho$  are defined in (3.2) and (5.5), respectively.

Validity of Theorem 2 follows directly from Lemmas 1 and 2 and inequality (5.15).

Note that the value of  $m_0$  depends on the unknown parameters  $s$  and  $p$  of the Besov space, hence, in general, estimator  $\hat{f}_{m_0}$  in (5.16) is not adaptive. However, if  $D = D_0 = 0$ , then  $f_{0,m} = \hat{f}_{0,m} \equiv 0$ ,  $\hat{f} = \hat{f}_{c,m}$  and one can choose  $m = m_1$  in  $\hat{f}_{c,m}$  using formula (4.2), so that the estimator is adaptive. In this case, convergence rates of  $\hat{f}$  are given entirely by Lemma 1. In particular, the following corollary is valid.

**COROLLARY 1.** *Let  $D = D_0 = 0$  and assumptions of Lemma 1 hold. Consider estimator  $\hat{f}_m = \hat{f}_{c,m}$  given by (4.12) with  $m = m_1$ . Then*

$$(5.17) \quad \sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{m_1} - f\|^2 \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^\rho \quad (\varepsilon \rightarrow 0),$$

where  $\Delta(\varepsilon)$  and  $\rho$  are defined in (3.2) and (5.5), respectively.

If  $DD_0 > 0$ , then it is necessary to construct an adaptive estimator of  $f$ . Note that this is not an easy task. Expanding the system of equations in (4.16) so that it includes not only the scaling but also the wavelet coefficients will compromise uniformity of eigenvalues of matrix  $\mathbf{A}^{(m)}$  [see (5.8)] which are ensured by positive-definiteness and finite size of  $\mathbf{A}^{(m)}$ . On the other hand, introducing a penalty on the solution does not help either since, for any  $m$ , the system involves the unknown bias term  $\mathbf{r}^{(m)}$ . For this reason, in order to choose parameter  $m$ , we apply Lepski’s method since it allows us to eliminate the bias inherent to the system of equations (4.21).

**6. Adaptive estimation in the presence of singularity.** In order to construct an adaptive estimator of  $f$  in the presence of singularity, we shall use the technique of optimal tuning parameter selection pioneered by Lepski (1990, 1991) and further exploited in Lepski and Spokoiny (1997) and Lepski, Mammen and Spokoiny (1997). The idea behind this technique is to construct estimators for various values of the tuning parameter in question ( $m$ , in our case), and then choose an optimal value of the tuning parameter by regulating the differences between the estimators constructed with different values of the parameter.

In particular, for various values of  $m$ , we construct versions of the system of equations, obtain values of  $\hat{\mathbf{z}}^{(m)}$  in (4.22) and use them as  $\hat{a}_{mk} = \hat{z}_k^{(m)}$ ,  $k \in K_{0m}$ , in (4.23). Then, for various values of  $m$ , we obtain estimators  $\hat{f}_m$  of  $f$  using formula (4.7) where  $\hat{f}_{0,m}$  and  $\hat{f}_{c,m}$  are of the forms (4.23) and (4.12), respectively, and  $m$  is the lowest resolution level of  $\hat{f}_{c,m}$ . After that, we choose the “best possible” resolution level  $\hat{m}$  and consider estimator  $\hat{f}_{\hat{m}}$  as the final estimator. The choice of the resolution level  $\hat{m}$  is driven by the singularity-affected portion of  $f$  rather than the zero-free portion as it is described below.

For any resolution level  $m > 0$ , we define a neighborhood  $\Omega_m$  of  $x_0$  as

$$(6.1) \quad \Omega_m = \{x : \min(L_\varphi - D_0, L_\psi - D) < 2^m(x - x_0) < \max(U_\varphi + D_0, U_\psi + D)\},$$

where  $\text{supp } \varphi = (L_\varphi, U_\varphi)$  and  $\text{supp } \psi = (L_\psi, U_\psi)$ . Observe that  $\Omega_m$  is designed so that  $\text{supp}(f_{0,m}) \subseteq \Omega_m$ ,  $\text{supp}(\hat{f}_{0,m}) \subseteq \Omega_m$  and  $\Omega_j \subset \Omega_m$  if  $j > m$ .

Choose  $m = \hat{m}$  such that  $m_1 \leq m \leq J - 1$ , where  $J$  is defined in (4.13) and

$$(6.2) \quad \hat{m} = \min\{m : \|(\hat{f}_m - \hat{f}_j)\mathbb{I}(\Omega_m)\|^2 \leq \kappa^2 \varepsilon \ln(\varepsilon^{-1}) \lambda_j^{-2} \text{ for all } j \in [m, J - 1]\},$$

where  $\kappa > 0$  is a constant which is defined below.

The construction of  $\hat{m}$  is based on the following idea. Note that when  $m = \hat{m} \leq m_0$ , one has

$$(6.3) \quad \mathbb{E} \|\hat{f}_{\hat{m}} - f\|^2 \leq 2[\mathbb{E} \|\hat{f}_{\hat{m}} - \hat{f}_{m_0}\|^2 + \mathbb{E} \|\hat{f}_{m_0} - f\|^2].$$

The first component in (6.3) is small due to the definition of the resolution level  $\hat{m}$  while the second component is calculated at the optimal resolution level  $m_0$  and, hence, tends to zero at the optimal convergence rate (up to a logarithmic factor). On the other hand, if  $m = \hat{m} > m_0$ , then there should exist  $j > m$  such that  $\|(\hat{f}_m - \hat{f}_j)\mathbb{I}(\Omega_m)\|^2 > \kappa^2 \varepsilon \ln(\varepsilon^{-1}) \lambda_j^{-2}$ . The following Lemma shows that the probability of this event is infinitesimally small provided  $\kappa$  is large enough.

**LEMMA 3.** *Let  $m_0$  and  $\hat{m}$  be given by expressions (5.14) and (6.2) and the conditions of Theorem 2 hold. Let  $C_\kappa = 2^{11} D_0^2 C_{A2}^2 \max(C_{\rho\lambda}^{-1} C_{w2}^2, C_t K_1)$  where constants  $C_{A2}$ ,  $C_{\rho\lambda}$ ,  $C_{w2}$ ,  $C_t$  and  $K_1$  are defined in (5.8), (5.9), (5.7), (2.3) and (5.10),*



respectively. If

$$(6.4) \quad \kappa \geq \max(d^2 C_\kappa, 2^6(d^2 + 1)C_t), \quad \chi^2 \geq d^2 + 2/(2 + \alpha + \beta),$$

then, as  $\varepsilon \rightarrow 0$ ,

$$(6.5) \quad \mathbb{P}(\hat{m} > m_0) \leq C\varepsilon^{d^2}.$$

Lemma 3 confirms that indeed  $m = \hat{m}$  can be chosen as the lowest resolution level in the nonlinear portion of the estimator, so that we estimate  $f$  by

$$(6.6) \quad \hat{f}(x) = \hat{f}_{0,\hat{m}}(x) + \hat{f}_{c,\hat{m}}(x), \quad x \in [0, 1],$$

where  $\hat{f}_{0,m}(x)$  and  $\hat{f}_{c,m}(x)$  are defined in (4.23) and (4.12), respectively. The following statement confirms that the wavelet nonlinear estimator  $\hat{f}$  given by (6.6) indeed attains (up to a logarithmic factor) the asymptotic minimax lower bounds obtained in Theorem 1.

**THEOREM 3.** *Let  $1 \leq p, q \leq \infty$  and  $s \geq \max(1/2, 1/p)$ . Let conditions (5.9)–(5.11) and assumptions (A1)–(A5) hold with  $\alpha \geq 1$  and  $DD_0 > 0$  in assumptions (A2) and (A4). Consider the estimator (4.7) of  $f$  where  $\hat{f}_{c,m}$  and  $\hat{f}_{0,m}$  are given by formulae (4.12) and (4.23), respectively. Let  $m = \hat{m}$  where  $\hat{m}$  is defined in (6.2). Let  $J$  be defined in (4.13) and  $\kappa$  and  $\chi$  be such that*

$$(6.7) \quad \kappa \geq \max(4C_\kappa, 320C_t), \quad \chi^2 \geq 4 + 2/(2 + \alpha + \beta),$$

where  $C_\kappa$  and  $C_t$  are defined in Lemma 3 and formula (2.3), respectively. Then,

$$(6.8) \quad \sup_{f \in B_{p,q}^s(A)} \mathbb{E} \|\hat{f}_{\hat{m}} - f\|^2 \leq C \Delta(\varepsilon) [\ln(\varepsilon^{-1})]^{1+\mathbb{I}(\alpha=1)} \quad (\varepsilon \rightarrow 0),$$

where  $\Delta(\varepsilon)$  is defined in (3.2).

**REMARK 4** (Logarithmic factor in convergence rates). Note that in (5.5),  $\rho \neq 0$  only if  $2s(\alpha - 1) = (\beta + 1)(1 - 2/p)$ , or  $\alpha = 1$ , or  $\alpha < 1$  and  $p \leq 2$ . The latter shows that the lower bounds for the risk in Theorem 1 cannot be made tighter, at least, in the case when  $\alpha > 1$ . Theorems 1 and 2 and Corollary 1 demonstrate that estimator (4.7) attains the asymptotically optimal convergence rates if  $2s(\alpha - 1) \neq (\beta + 1)(1 - 2/p)$  and  $\alpha > 1$ , or if  $\alpha < 1$  and  $p > 2$ . Otherwise, estimator (4.7) is asymptotically near-optimal up to a logarithmic factor. However, in Theorem 3 the risk of the adaptive estimator is always within a logarithmic factor  $[\ln(\varepsilon^{-1})]^{1+\mathbb{I}(\alpha=1)}$  of the minimax risk. The latter is due to application of Lepski method. Note that in spite of the fact that we are using the integrated mean squared error, Lepski method is applied locally and, hence, leads to an extra log-factor in the risk, as it usually happens with application of Lepski method to pointwise estimation.

**7. Examples.**

7.1. *Deconvolution with a spatially inhomogeneous kernel.* Consider problem (1.1) with operator  $Q$  of the form

$$(7.1) \quad (Qf)(x) = \mu(x) \int_0^1 q(x-t)f(t) dt, \quad x \in [0, 1],$$

where functions  $\mu(x)$ ,  $q(x)$  and  $f(x)$  are periodic and both  $q(x)$  and  $\mu(x)$  are completely known. Problem (1.1) is equivalent to the following statistical problem:

$$(7.2) \quad y_i = \mu(i/n) \int_0^1 q(i/n-t)f(t) dt + \sigma \xi_i, \quad i = 1, \dots, n,$$

where  $\xi_i$  is a white Gaussian noise and  $\varepsilon = \sigma^2/n$ . Equation of the form (7.1) can appear when one observes a convolution  $Y(x)$  of the known kernel  $q$  with the unknown function of interest  $f$  and a known heteroscedastic noise  $\sqrt{\varepsilon}\gamma(x)W(x)$ , so that  $\mu(x) = [\gamma(x)]^{-1}$  and  $Y(x) = \gamma(x)y(x) = y(x)/\mu(x)$ . In this case, equation (7.2) takes the form

$$(7.3) \quad Y_i = \int_0^1 q(i/n-t)f(t) dt + \sigma \gamma(i/n)\xi_i, \quad i = 1, \dots, n.$$

If  $\mu(x)$  is uniformly bounded above and below, in principle, spatial inhomogeneity of operator  $Q$  in (7.1) can be ignored. Below we consider the case when the former is not true since  $\mu(x)$  vanishes at some point  $x_0 \in (0, 1)$ , in particular,

$$(7.4) \quad C_{\mu 1}|x - x_0|^\alpha \leq \mu^2(x) \leq C_{\mu 2}|x - x_0|^\alpha$$

for some for some positive constants  $\alpha$ ,  $C_{\mu 1}$  and  $C_{\mu 2}$  independent of  $x_0$  and  $x$ . Therefore, the version of the problem studied in the present paper can be described as locally extreme noise which occurs when the degree of spatial inhomogeneity is high. Direct calculations show that

$$(Q^*h)(z) = \int_0^1 q(x-z)h(x)\mu(x) dx.$$

Hence, functions  $u_{jk}$  and  $t_{mk}$  are solutions of the equations

$$Q^*u_{jk} = \psi_{jk}, \quad Q^*t_{mk} = \varphi_{mk}.$$

Consider  $\omega \in \mathbb{Z}$  and let  $e_\omega(t) = e^{i2\pi\omega t}$ ,  $t \in [0, 1]$ , be the periodic Fourier basis on  $[0, 1]$ . Denote  $U_{jk}(x) = u_{jk}(x)\mu(x)$  and, similarly,  $T_{mk}(x) = t_{mk}(x)\mu(x)$  and introduce Fourier coefficients  $q_\omega = \langle q, e_\omega \rangle$ ,  $U_{jk\omega} = \langle U_{jk}, e_\omega \rangle$ ,  $T_{mk\omega} = \langle T_{mk}, e_\omega \rangle$ ,  $\psi_{jk\omega} = \langle \psi_{jk}, e_\omega \rangle$  and  $\varphi_{mk\omega} = \langle \varphi_{mk}, e_\omega \rangle$ . Then,  $U_{jk\omega} = [\bar{q}_\omega]^{-1}\psi_{jk\omega}$ ,  $T_{mk\omega} = [\bar{q}_\omega]^{-1}\varphi_{mk\omega}$  and

$$(7.5) \quad \begin{aligned} U_{jk}(x) &= \sum_{\omega \in \mathbb{Z}} [\bar{q}_\omega]^{-1} \psi_{jk\omega} e_\omega(x), \\ T_{mk}(x) &= \sum_{\omega \in \mathbb{Z}} [\bar{q}_\omega]^{-1} \varphi_{mk\omega} e_\omega(x), \end{aligned}$$

where  $\bar{q}_\omega$  is the complex conjugate of  $q_\omega$ . Moreover, estimators  $\hat{a}_{mk}$  and  $\hat{b}_{jk}$  in (4.3) can be constructed using Fourier wavelet transform suggested in [Johnstone et al. \(2004\)](#). Indeed, if  $Y(x) = y(x)/\mu(x)$  and  $Y_\omega$  are Fourier coefficients of function  $Y(x)$ , then

$$(7.6) \quad \begin{aligned} \hat{b}_{jk} &= \langle y, u_{jk} \rangle = \langle Y, U_{jk} \rangle = \sum_{\omega \in \mathbb{Z}} [\bar{q}_\omega]^{-1} \psi_{jk\omega} \bar{Y}_\omega, \\ \hat{a}_{mk} &= \sum_{\omega \in \mathbb{Z}} [\bar{q}_\omega]^{-1} \varphi_{mk\omega} \bar{Y}_\omega. \end{aligned}$$

In the case of the statistical experiment described in formula (7.2), Fourier coefficients are replaced by the discrete Fourier transform.

Note that application of the wavelet-vaguellete methodology to deconvolution with heteroscedastic noise (7.3) appears very reasonable. Really, if noise level  $\gamma(x)$  is such that  $\int \gamma^2(x) dx < \infty$ , then formula (7.6) implies that wavelet coefficients are estimated using Fourier transform of the measured signal  $Y$  in (7.3) and then thresholded taking into account the local noise level. Indeed, it is easy to observe that  $\lambda_{jk}$  in (7.10) is such that  $\lambda_{jk}^{-2} \asymp \|U_{jk}\|^2 \int \gamma^2(x) \mathbb{I}(x \in \text{supp } U_{jk}) dx$ . If  $\int \gamma^2(x) dx = \infty$  in the vicinity of some point  $x_0$ , the natural strategy suggested above cease working and one needs another means for estimating scaling and wavelet coefficients in the neighborhood of  $x_0$ . In this situation one has to apply the hybrid estimator constructed in Sections 4 and 6. In Section 8, we provide a detailed description of the computational algorithm for this task.

If function  $\mu(x)$  is not completely known and is estimated from data, matrices  $\mathbf{A}^{(m)}$  and  $\mathbf{B}^{(m)}$  as well as vector  $\hat{c}^{(m)}$  will be subjected to additional errors which have to be accounted for by using regularization techniques designed for the inverse problems with errors in the operator [see, e.g., [Engl, Hanke and Neubauer \(1996\)](#) and [Hoffmann and Reiss \(2008\)](#)].

In order to find  $\lambda_{jk} \asymp \|v_{jk}\|$  in assumption (A1) and verify assumptions (A1)–(A5), we impose the following conditions on the kernel  $q$  and mother and father wavelets  $\psi$  and  $\varphi$ .

(E1) Kernel  $q(x)$  is  $(r - 2)$  times continuously differentiable on  $[0, 1]$  and  $r_1 > r \geq 1$  times differentiable outside the neighborhood of jump discontinuities of  $q^{(r-1)}$  with  $q^{(r)}$  and  $q^{(r_1)}$  uniformly bounded. The value  $r = 1$  corresponds to the case when  $q$  itself has jump discontinuities.

(E2) Fourier coefficients  $q_\omega$  of  $q$  are such that  $C_{q1}(|\omega| + 1)^{-r} \leq |q_\omega| \leq C_{q2}(|\omega| + 1)^{-r}$  for some positive constants  $C_{q1}$  and  $C_{q2}$  independent of  $\omega$ .

(E3) Let  $\psi$  be  $r_0$ -regular,  $r_0$  times continuously differentiable wavelet function with the bounded support, where  $r_0 > \max(r, r_1)$ .

(E4) Kernel  $q$  is such that functions  $U_{jk}$  and  $T_{jk}$  defined in (7.5) have bounded supports of the lengths proportional to  $2^{-j}$  and centered at  $2^{-j}k$ :  $\text{supp}(U_{jk}) = (2^{-j}(k - d_U), 2^{-j}(k + d_U))$  and  $\text{supp}(T_{jk}) = (2^{-j}(k - d_T), 2^{-j}(k + d_T))$ .

There are many functions  $q$  satisfying conditions above, among them, for example,

$$(7.7) \quad \begin{aligned} q_1(x) &= \sum_{k \in \mathbb{Z}} \exp(-\lambda|x+k|) \quad \text{and} \\ q_2(x) &= \sum_{k \geq 0} \exp(-\lambda(x+k))(x+k)^N \end{aligned}$$

with  $r = 2$  for  $q_1(x)$  and  $r = N + 1$  for  $q_2(x)$ . Note that under assumptions (E1)–(E4), one has  $\|u_{jk}\|^2 = \infty$  if  $|k - k_{0j}| < d_U$  and  $\alpha \geq 1$  and

$$(7.8) \quad \begin{aligned} \|u_{jk}\|^2 &\asymp \mu^{-2}(2^{-j}k)\|U_{jk}\|^2 \asymp |2^{-j}k - x_0|^{-\alpha} \sum_{\omega \in \mathbb{Z}} |\bar{q}_\omega|^{-2} |\psi_{jk\omega}|^2 \\ &\leq C2^{j\alpha} [|k - k_{0j}|^\alpha + 1]^{-1} \sum_{\omega \in \mathbb{Z}} (|\omega|^{2r} + 1) |\psi_{jk\omega}|^2, \end{aligned}$$

otherwise. Due to conditions (E3) and (E4) and periodicity of  $\psi_{jk}$ , using integration by parts  $r$  times, we obtain

$$\begin{aligned} \psi_{jk\omega} &= \int_0^1 \psi_{jk}(x) e^{i2\pi\omega x} dx \\ &= 2^{j(r+1/2)} (-2\pi i\omega)^r \int_0^1 \psi^{(r)}(2^j x - k) e^{i2\pi\omega x} dx, \end{aligned}$$

so that  $2^{-2jr} \sum_{\omega \in \mathbb{Z}} (|\omega|^{2r} + 1) |\psi_{jk\omega}|^2 \leq C(2^{-2jr} \|\psi\|^2 + \|\psi^{(r)}\|^2)$ . Therefore,

$$(7.9) \quad \begin{aligned} \|u_{jk}\|^2 &\leq C2^{j(\alpha+2r)} [|k - k_{0j}|^\alpha + 1]^{-1} \asymp \lambda_{jk}^{-2}, \\ &\text{if } |k - k_{0j}| \geq d_U \text{ or } \alpha < 1 \end{aligned}$$

and  $\|u_{jk}\| = \infty$  otherwise. Similar inequality can be proved for  $\|t_{mk}\|^2$ .

The following proposition shows that, indeed,  $\lambda_{jk}$  is defined by expression (7.9) and that, under conditions (E1)–(E4), assumptions (A1)–(A5) hold.

**PROPOSITION 1.** *Let  $\mu(x)$  satisfy condition (7.4). Then, under assumptions (E1)–(E4) with  $r_1$  such that  $2r_1 + 1 > 2r + \alpha$ , one has*

$$(7.10) \quad \lambda_{jk}^2 \asymp 2^{-j(2r+\alpha)} [|k - k_{0j}|^\alpha + 1].$$

*Furthermore, assumptions (A1)–(A5) hold with  $D = D_0 = 0$  if  $\alpha < 1$ , and with  $D = d_U$  and  $D_0 = d_T$  if  $\alpha \geq 1$ . Conditions (5.9)–(5.11) of Lemma 2 are also valid.*

Due to Proposition 1, all statements and constructions in Sections 3–6 can be applied to equation (1.1) with operator  $Q$  given in (7.1). In particular, Theorems 1–3 can be utilized with  $\beta = 2r$ . By direct comparison with, for example, Johnstone et al. (2004), one can see that if  $\alpha = 0$ , so that the problem is spatially homogeneous, then the rates of convergence in Theorems 1–3 coincide with the usual convergence rates exhibited in deconvolution problems with white noise.

7.2. *Deconvolution with spatially inhomogeneous design.* Consider the problem of deconvolution when measurements are irregularly spaced. In particular, let  $g$  be a sampling p.d.f. with the corresponding c.d.f.  $G$ . Due to irregular design, operator  $Q$  can be presented as

$$(7.11) \quad (Qf)(x) = \int_0^1 q(G^{-1}(x) - t)f(t) dt,$$

where  $G^{-1}$  is the inverse of  $G$ . In this case, equation (1.1) can be viewed as an idealized version of the equation

$$(7.12) \quad y_i = \int_0^1 q(x_i - t)f(t) dt + \sigma \xi_i, \quad i = 1, \dots, n,$$

where  $\varepsilon = \sigma^2/n$ ,  $\xi_i$  is a white Gaussian noise, observation points  $x_i, i = 1, \dots, n$ , are such that  $G(i/n) = x_i$  and  $x_i$ 's and  $\xi_j$ 's are independent. Then,  $G^{-1}(i/n) = x_i$ , and the right-hand side of (1.1) with operator  $Q$  given by (7.11) provides a continuous equivalent of the statistical problem (7.12).

In what follows, we assume that functions  $q(x)$  and  $f(x)$  are periodic and both  $q(x)$  and  $g(x)$  are completely known. In this case, the conjugate operator  $Q^*$  is of the form

$$(Q^*h)(z) = \int_0^1 q(G^{-1}(x) - z)h(x) dx.$$

It is pretty straightforward to show that  $u_{jk}(x) = U_{jk}(G^{-1}(x))$  and  $t_{mk}(x) = T_{mk}(G^{-1}(x))$  where, as before,  $U_{jk}(\cdot)$  and  $T_{mk}(\cdot)$  are given by formula (7.5). Wavelet coefficients  $\hat{b}_{jk}$  and  $\hat{a}_{mk}$  can also be estimated in a manner similar to Example 1. Indeed, if  $Y(x) = y(G(x))$  and  $Y_\omega$  are Fourier coefficients of  $Y$ , then  $\hat{b}_{jk}$  and  $\hat{a}_{mk}$  can be evaluated using formula (7.6).

If design density  $g$  is unknown, then both  $g(x)$  and  $G(x)$  have to be estimated from observations  $x_i, i = 1, \dots, n$ . The latter will lead to additional errors in estimating wavelet coefficients  $\hat{b}_{jk}$  and  $\hat{a}_{mk}$  as well as entries of matrices  $\mathbf{A}^{(m)}$  and  $\mathbf{B}^{(m)}$  and vector  $\hat{\mathbf{c}}^{(m)}$ . The issue of additional errors has to be addressed by using, for example, regularization techniques [see, e.g., Engl, Hanke and Neubauer (1996) and Hoffmann and Reiss (2008)].

We assume that design density  $g(x)$  has a single zero of order  $\alpha$  at  $x_0$ , that is,  $g(x_0 + x)|x|^{-\alpha} \rightarrow C_g$  as  $x \rightarrow 0$ . The latter implies that there exist some absolute constants  $C_{g1}$  and  $C_{g2}$  such that, for any  $x$ , one has

$$(7.13) \quad C_{g1}|x - x_0|^\alpha \leq g(x) \leq C_{g2}|x - x_0|^\alpha.$$

Thus, we are considering the case of extremely inhomogeneous design which can be described also as a local data loss.

In this case, under conditions (E1)–(E4), similarly to Example 1, one has  $\|u_{jk}\|^2 = \int g^{-1}(x)U_{jk}^2(x) dx$ . Hence, identically to (7.9), one has  $\|u_{jk}\|^2 \leq$

$C2^{j(\alpha+2r)}|k - k_{0j}|^{-\alpha}$  if  $|k - k_{0j}| \geq d_U$  or  $\alpha < 1$ , and  $\|u_{jk}\|^2 = \infty$  if  $|k - k_{0j}| < d_U$  and  $\alpha \geq 1$ . Moreover, by simple modifications of the proof of Proposition 1, it easy to show that the following statement is valid.

**PROPOSITION 2.** *Let  $g(x)$  satisfy condition (7.13). Then, under assumptions (E1)–(E4) with  $r_1$  such that  $2r_1 + 1 > 2r + \alpha$ , the value of  $\lambda_{jk}$  is given by formula (7.10). Furthermore, assumptions (A1)–(A5) hold with  $D = D_0 = 0$  if  $\alpha < 1$ , and with  $D = d_U$  and  $D_0 = d_T$  if  $\alpha \geq 1$ . Conditions (5.9)–(5.11) of Lemma 2 are also valid.*

Again, analogously to Section 7.1, due to Proposition 2, all statements and constructions in Sections 3–6 can be applied to equation (1.1) with operator  $Q$  given in (7.11). In particular, Theorems 1–3 can be used with  $\beta = 2r$ . If  $q$  is the Dirac delta function, then  $(Qf)(x) = f(G^{-1}(x))$ , then  $r = 0$  and the problem reduces to regression estimator based on spatially inhomogeneous data studied in Antoniadis, Pensky and Sapatinas (2013). In this case, the rates of convergence coincide with the minimax convergence rates derived therein.

**REMARK 5** (Irregularly spaced observations and heterogeneous noise). It follows from examples in Sections 7.1 and 7.2 that there is a direct correspondence between deconvolution with irregularly spaced measurements and deconvolution with heterogeneous noise. In particular, as far as convergence rates are concerned, the squared noise level acts in a similar way to the inverse of the design density and both are equivalent, in some way, to a multiplicative factor in the convolution operator.

**8. Simulation study and real data application.**

8.1. *Simulation study.* In order to assess finite sample properties of the proposed methodology and, in particular, performance of the hybrid estimator, we carried out a small simulation study. We limited our attention to the deconvolution in the presence of heteroscedastic noise described in Section 7.1. Specifically, we considered  $q(x) = q_1(x)$  with  $\lambda = 5$  where  $q_1(x)$  is defined in (7.7). We chose one of the standard test functions, `blip`, as the true function  $f(x)$ . Function  $\mu(x)$  in (7.1) is of the form  $\mu(x) = |h^{-1}(x - x_0)|^{\alpha/2} \mathbb{I}(|x - x_0| \leq h) + \mathbb{I}(|x - x_0| > h)$  with  $x_0 = 1/3$  and  $h = 1/6$ , so that condition (7.4) holds. We generated data using equation (7.3) with  $\gamma(x) = \mu^{-1}(x)$  and  $\sigma = 0.02$ , in particular,

$$Y_i = H(i/n) + \sigma \gamma(i/n) \xi_i, \quad i = 1, \dots, n$$

with

$$H(x) = \int_0^1 q(x - t) f(t) dt.$$

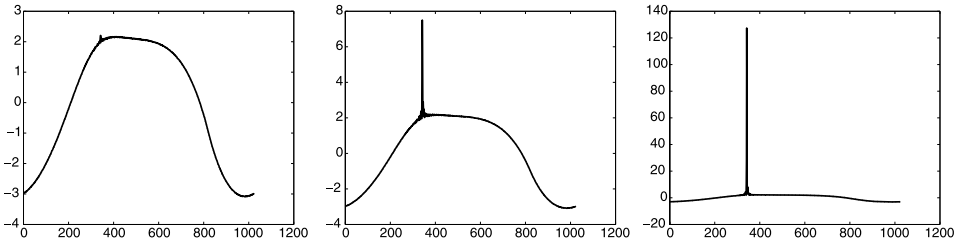


FIG. 1. True function  $H$  (dotted line) and observed data (solid line) with  $\alpha = 1$  (left),  $\alpha = 2$  (middle) and  $\alpha = 3$  (right). Here,  $\text{SNR} = 0.8848$  for  $\alpha = 1$ ,  $\text{SNR} = 0.0808$  for  $\alpha = 2$ ,  $\text{SNR} = 0.0183$  for  $\alpha = 2.5$  and  $\text{SNR} = 0.0040$  for  $\alpha = 3$ .

We evaluate noise intensity by the common in signal processing, signal-to-noise ratio (SNR) which is defined as  $\text{SNR} = \sqrt{n} \text{std}(f) / (\|\gamma\| * \sigma)$  where  $\|\gamma\|$  is the  $L^2$ -norm of  $\gamma$  and  $\text{std}(p)$  is the standard deviation of  $p$  for any function  $p(x)$ .

We used WaveLab package for Matlab and carried out simulations using degree 8 Daubechies wavelets and  $n = 1024$ . In order to obtain estimators of wavelet and scaling coefficients, we generated wavelet and scaling functions  $\psi_{jk}$  and  $\varphi_{mk}$  using `MakeWavelet` command and obtained the respective matrix of the Fourier coefficients. Subsequently, we found estimators of wavelet and scaling coefficients using formula (7.6) with  $Y_\omega$  being discrete Fourier transform of vector  $Y$  in (7.3). We generated values of  $\lambda_{jk}^{-2}$  using equation (7.8) and used them for hard thresholding. Due to relatively small value of  $n$ , we did not use block thresholding described in Section 4. By applying inverse wavelet transform to the thresholded wavelet coefficients, we obtained deconvolution estimator  $\hat{f}$ .

We evaluated performance of the estimators for  $n = 1024$  and different values of  $\alpha$ . As it is expected, when  $\alpha$  is growing, the SNRs are decreasing and the quality of observed data is declining. Figure 1 demonstrates observed data for various values of  $\alpha$ . The corresponding signal-to-noise ratios are  $\text{SNR} = 0.8848$  for  $\alpha = 1$ ,  $\text{SNR} = 0.0808$  for  $\alpha = 2$  and  $\text{SNR} = 0.0040$  for  $\alpha = 3$ .

Figure 2 shows wavelet–vaguelette deconvolution estimators (4.12) obtained for  $\alpha = 0, \alpha = 1$  and  $\alpha = 2$ . Note that for moderate values of  $\alpha$ , the wavelet–vaguelette

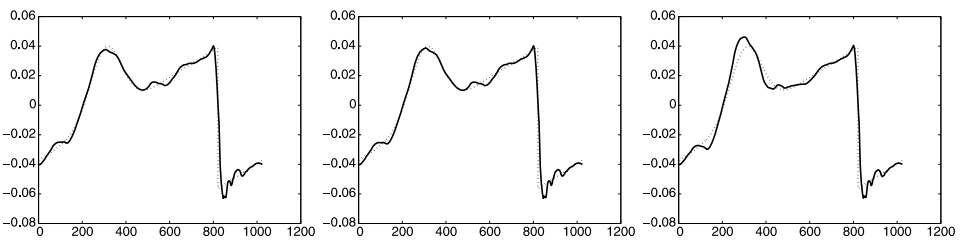


FIG. 2. Thresholded wavelet–vaguelette deconvolution estimator with  $n = 1024$ . True regression  $f$  (dotted line) and its estimated value (solid line) with  $\alpha = 0$  (left),  $\alpha = 1$  (middle) and  $\alpha = 2$  (right).

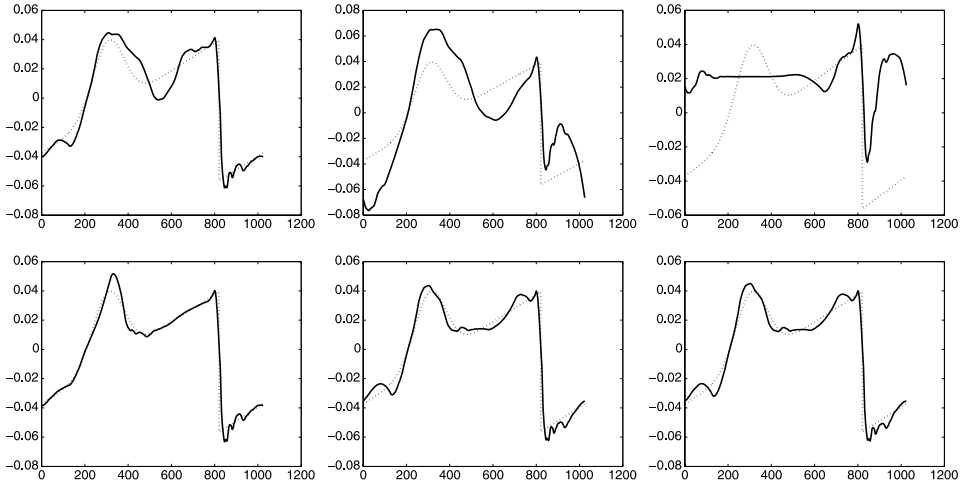


FIG. 3. Wavelet–vaguelette (top row) and hybrid (bottom row) estimators for various values of  $\alpha$ :  $\alpha = 2.5$  (left),  $\alpha = 3$  (middle) and  $\alpha = 4$  (right). True  $f$  (dotted line), wavelet–vaguelette estimator or hybrid estimator (solid line).

estimator adjusts to spatially inhomogeneous noise quite well. Indeed, fluctuations at the right end of the graph appear even when  $\alpha = 0$  (upper left) and are due to the relatively crude choice of threshold in formula (4.12). Actually, for  $\alpha = 0$  the noise cease to be inhomogeneous and estimator (4.12) reduces to Fourier-wavelet estimator of Johnstone et al. (2004).

As the values of  $\alpha$  grow, SNR declines and the wavelet–vaguelette estimators (4.12) deteriorate. If  $\alpha = 3$ , as Figure 3 shows, the wavelet–vaguelette reconstruction has little resemblance to the regression function which it estimates. For large values of  $\alpha$ , we construct hybrid estimators described in Sections 4 and 6. Construction of the adaptive hybrid estimator consists of the following steps.

1. Fix the lowest resolution level  $m_1$  and the highest resolution level  $J$ . For each value of  $m = m_1, \dots, J - 1$ , repeat steps 2–6.
2. Obtain the wavelet–vaguelette estimator of  $f$  with the lowest resolution level  $m$  using formula (4.12).
3. Identify sets  $K_{0m}, K_{0m}^c, K_{1j}$  and  $K_{1j}^c$  for  $j = m, \dots, J - 1$ . Also, find set  $\Omega_m$ .
4. Form matrices  $\mathbf{A}^{(m)}$  and  $\mathbf{B}^{(m)}$  and vector  $\hat{\mathbf{c}}^{(m)}$  using formulae (4.17) and (4.18), respectively, and obtain solution  $\hat{\mathbf{z}}^{(m)}$  of the system of equations (4.21). Use vector  $\hat{\mathbf{z}}^{(m)}$  as coefficients  $\hat{a}_{mk}$ ,  $k \in K_{0m}$ , in the zero-affected portion of the estimator (4.23).
5. Replace estimators of the scaling coefficients (if  $k \in K_{0m}$ ) and wavelet coefficients (if  $k \in K_{1j}$ ,  $j = m, \dots, J - 1$ ) by zeros to obtain the zero-free portion of the estimator (4.12).



6. Combine wavelet coefficients in steps 4 and 5 to obtain wavelet coefficients of  $\hat{f}_m$ . Recover  $\hat{f}_m$  using inverse wavelet transform. Set

$$\lambda_m^{-2} = \sigma^2 \sum_k \mu^{-2} (2^{-mk}) \|T_{mk}\|^2.$$

7. For each  $m = m_1, \dots, J - 1$  and  $j = m + 1, \dots, J - 1$ , evaluate matrix of the adjusted differences

$$(8.1) \quad \mathcal{L}_{mj} = \|(\hat{f}_m - \hat{f}_j)\mathbb{I}(\Omega_m)\|^2 / (\sigma^2 n^{-1} \log n \lambda_j^{-2}).$$

Choose  $\hat{m}$  in (6.2) by comparing maximum value of row  $m$  of matrix  $\mathcal{L}$  with a constant  $\kappa^2$

$$(8.2) \quad \hat{m} = \min\{m : \mathcal{L}_{mj} \leq \kappa^2 \text{ for all } j, m \leq j \leq J - 1\}.$$

In our simulations, we used  $n = 1024$ ,  $m_1 = 1$  and  $J = 7$  and carried out hybrid estimation with  $\alpha = 2.5$ ,  $\alpha = 3$  and  $\alpha = 4$ . Simulation results for these three cases are presented in Figure 3. The upper and the lower rows present reconstructions of  $f$  by wavelet–vaguelette and hybrid estimator, respectively. Observe that for  $\alpha = 2.5$  the wavelet–vaguelette estimator still generally follows the true function  $f$  but for  $\alpha = 3$  or  $\alpha = 4$  it bears little resemblance to  $f$ . The hybrid estimator allows to account for inhomogeneity of the noise and to significantly improve reconstruction of  $f$ .

Lepski procedure provides a choice of resolution level  $\hat{m}$  for each of the values of  $\alpha$ . Figure 4 demonstrates hybrid estimators with the various lowest resolution

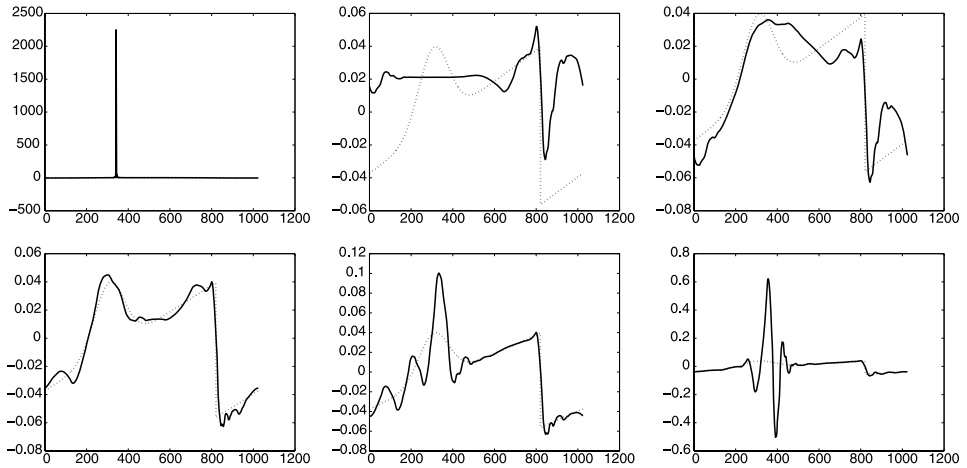


FIG. 4. Hybrid estimation with  $\alpha = 4$ . True  $H$  (dotted line) and observed data (solid line) (upper left), true  $f$  (dotted line) and wavelet–vaguelette estimator (solid line) (upper middle), true  $f$  (dotted line) and hybrid estimators (solid line) with  $m = 2$  (upper right),  $m = 3$  (lower left),  $m = 4$  (lower middle) and  $m = 5$  (lower right). Lepski method selects estimator with  $\hat{m} = 3$  (lower left).

levels  $m$  when  $\alpha = 4$ . In this case, maximum over  $j$  of  $\mathcal{L}_{mj}$  is very large for  $m \leq 2$  and is below 3 for  $m \geq 3$ , so that  $\hat{m} = 3$ . Figure 4 confirms that Lepski procedure makes the correct choice.

8.2. *Real data application.* Below, we consider application of the hybrid estimator developed in the paper to recovery of a convolution signal transmitted via amplitude modulation described in Example 2. Mathematically, the problem reduces to deconvolution with a spatially inhomogeneous kernel in Section 7.1 and appears in the form of equation (7.2) with  $\mu(x) = \cos(2\pi\omega x + \theta)$ , where  $\omega \approx n/2$  and  $\theta \in [0; \pi]$ . We chose  $\omega = n/2 + 1/2$ , so that  $\mu(x)$  has two zeros of order  $\alpha = 2$ ,  $x_{01} \leq 1/2$  and  $x_{02} = x_{01} + 1/2$  in  $[0, 1]$ . In particular,  $x_{01} = 0.5 - \theta/\pi$  if  $\theta < \pi/2$  and  $x_{01} = 1.0 - \theta/\pi$  if  $\pi/2 \leq \theta \leq \pi$ .

For simplicity, we considered the same set up as in simulation example, that is, we used  $q(x) = q_1(x)$  with  $\lambda = 5$  where  $q_1(x)$  is defined in (7.7) and one of the standard test functions, `blip`, as the true function  $f(x)$ . We carried out simulations with  $n = 512$ ,  $\sigma = 0.01$  and degree 8 Daubechies wavelets. The locations of zeros were estimated from the data. For example, when  $\theta = \pi/6$ ,  $x_{01} = 1/3$  and  $x_{02} = 5/6$ , locations of zeros were estimated as  $\hat{x}_{01} = 0.33496$  and  $\hat{x}_{02} = 0.83496$ .

Figure 5 presents signal  $y$  with uniform noise generated according to equation (7.2) as well as signals  $Y$  with heteroscedastic noise obtained according to equation (7.3) by dividing equation (7.2) by  $\mu(i/n)$ . Due to the limited resolution of the pictures, in Figure 5, we plotted only the first  $n/2 = 256$  values of the original signals  $y$ .

Figure 6 shows wavelet–vaguelette deconvolution estimators and hybrid estimators for various values of  $\theta$ :  $\theta = 0.49\pi$  ( $x_{01} = 0.010$  and  $x_{02} = 0.510$ ),  $\theta = \pi/6$  ( $x_{01} = 1/3$  and  $x_{02} = 5/6$ ) and  $\theta = \pi/10$  ( $x_{01} = 0.400$  and  $x_{02} = 0.900$ ). The

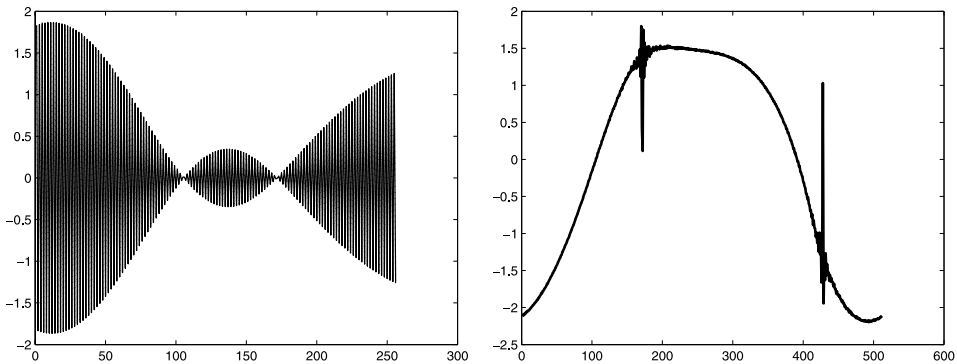


FIG. 5. Observed values of the signal and the true signal. Left: the left half (first  $n/2 = 256$  points) of the true signal (solid line) and observed data with homogeneous noise (dashed line). Right: true signal divided by  $\mu$  (dotted line) and observed data with heteroscedastic noise (solid line). Here, noise level  $\sigma = 0.01$ ,  $\theta = \pi/6$ ,  $x_{01} = 1/3$  and  $x_{02} = 5/6$ .

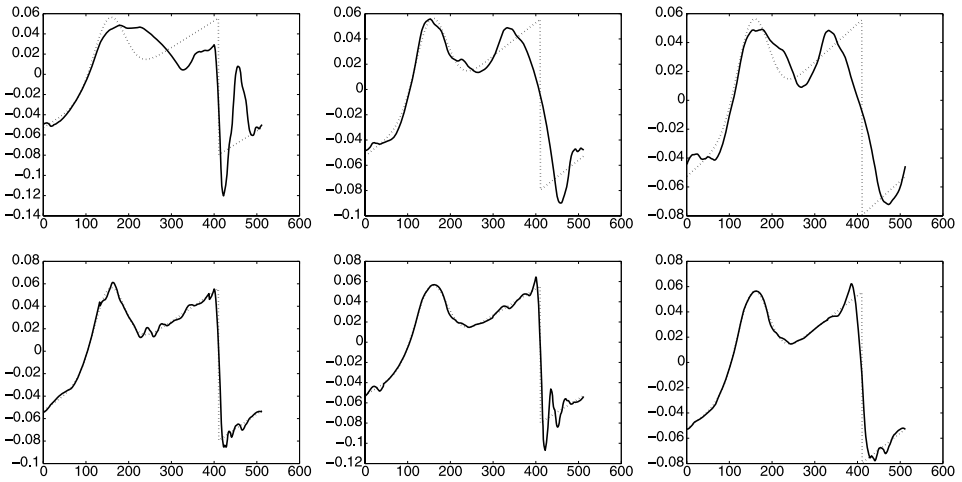


FIG. 6. Estimators of the signal. Top: true function  $f$  (dotted line) and wavelet–vaguelette estimator (solid line). Bottom: true function  $f$  (dotted line) and hybrid estimator (solid line). Here,  $\theta = 0.49\pi$  ( $x_{01} = 0.010$  and  $x_{02} = 0.510$ ) (left),  $\theta = \pi/6$  ( $x_{01} = 1/3$  and  $x_{02} = 5/6$ ) (middle) and  $\theta = \pi/10$  ( $x_{01} = 0.400$  and  $x_{02} = 0.900$ ) (right); noise level  $\sigma = 0.01$ .

upper and the lower rows present the wavelet–vaguelette and the hybrid estimators, respectively. In all three cases, the wavelet–vaguelette estimators deteriorate due to high value of  $\alpha$  while hybrid estimators deliver satisfactory recovery of the true function  $f$ . Note that the hybrid estimator has the worst performance when  $\theta = \pi/6$  due to proximity of zeros,  $x_{01} = 1/3$  and  $x_{02} = 5/6$ , of  $\mu(x)$  to discontinuities in the derivatives of the underlying signal  $f$ .

**9. Discussion.** In the present paper, we consider estimation of a solution of a spatially inhomogeneous linear inverse problem (1.1) with possible singularities. The special feature of problems like this is that the degree of ill-posedness depends not only on the scale but also on location. In spite of a huge number of publications devoted to linear inverse problems, to the best of our knowledge, this type of problems has never been treated before. We consider a version of a spatially inhomogeneous problem where there exists a singularity point  $x_0$  such that the norm of the solution grows when the right-hand side is localized in the vicinity of  $x_0$ . We characterize ill-posedness and spatial inhomogeneity of operator  $Q$  in terms of wavelet–vaguelette decomposition. The novel feature here is that the norms of vaguelettes depend on location and may be infinite in the vicinity of a singularity point, so that SVD-type solutions cease to work.

For this reason, estimators obtained in the paper are based either on wavelet–vaguelette decomposition (if the norms of all vaguelettes are finite) or on a hybrid of wavelet–vaguelette decomposition and Galerkin method (if vaguelettes in the neighborhood of the singularity point have infinite norms). We show that, up to a

logarithmic factor, the hybrid estimator attains the asymptotically optimal convergence rates.

The theory presented in the paper is supplemented by examples of deconvolution with a spatially inhomogeneous kernel, deconvolution in the presence of locally extreme noise or extremely inhomogeneous design. In addition, we apply the technique to recovery of a convolution signal transmitted via amplitude modulation.

We note that the wavelet-based estimation procedure presented in the paper is motivated by the need of constructing an asymptotically optimal estimator in the case when the unknown function  $f$  is spatially inhomogeneous. The estimator uses relatively crude thresholding procedure which can be improved by applying more sensitive thresholding techniques. Moreover, one can possibly find more efficient computational procedures than the hybrid estimator if establishing asymptotic optimality is not a priority.

The paper assumes that the operator  $Q$  in (1.1) is completely known. However, if this is not true, it will be interesting to investigate how uncertainty about  $Q$  affects the rates of convergence. Also, although the hybrid estimator works adequately when  $Q$  is completely known, it would require appropriate modifications if  $Q$  is partially unknown.

Finally, in the paper, we consider only the simplest case when the unknown function is univariate and is defined on an interval. The problem can be naturally extended to the case of multivariate function  $f$  which belongs to an isotropic or anisotropic Besov space. However, all these extensions will be a matter of future investigation.

**Acknowledgements.** The author would like to thank the Editor, the Associate Editor and two anonymous referees for their suggestions and support that helped to significantly improve the paper

## SUPPLEMENTARY MATERIAL

**Proofs** (DOI: [10.1214/13-AOS1166SUPP](https://doi.org/10.1214/13-AOS1166SUPP); .pdf). Supplement contains proofs of the statements in the manuscript.

## REFERENCES

- ABRAMOVICH, F. and SILVERMAN, B. W. (1998). Wavelet decomposition approaches to statistical inverse problems. *Biometrika* **85** 115–129. [MR1627226](#)
- ANTONIADIS, A., PENSKY, M. and SAPATINAS, T. (2013). Nonparametric regression estimation based on spatially inhomogeneous data: Minimax global convergence rates and adaptivity. *ESAIM Probab. Stat.* To appear.
- BISSANTZ, N., HOHAGE, T., MUNK, A. and RUYMGAART, F. (2007). Convergence rates of general regularization methods for statistical inverse problems and applications. *SIAM J. Numer. Anal.* **45** 2610–2636. [MR2361904](#)

- CAVALIER, L. and GOLUBEV, Y. (2006). Risk hull method and regularization by projections of ill-posed inverse problems. *Ann. Statist.* **34** 1653–1677. [MR2283712](#)
- CAVALIER, L., GOLUBEV, G. K., PICARD, D. and TSYBAKOV, A. B. (2002). Oracle inequalities for inverse problems. *Ann. Statist.* **30** 843–874. [MR1922543](#)
- COHEN, A., HOFFMANN, M. and REISS, M. (2004). Adaptive wavelet Galerkin methods for linear inverse problems. *SIAM J. Numer. Anal.* **42** 1479–1501 (electronic). [MR2114287](#)
- DONOHO, D. L. (1995). Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Appl. Comput. Harmon. Anal.* **2** 101–126. [MR1325535](#)
- ENGL, H. W., HANKE, M. and NEUBAUER, A. (1996). *Regularization of Inverse Problems*. Kluwer Academic, Dordrecht. [MR1408680](#)
- GAÏFFAS, S. (2005). Convergence rates for pointwise curve estimation with a degenerate design. *Math. Methods Statist.* **14** 1–27. [MR2158069](#)
- GAÏFFAS, S. (2007a). Sharp estimation in sup norm with random design. *Statist. Probab. Lett.* **77** 782–794. [MR2369683](#)
- GAÏFFAS, S. (2007b). On pointwise adaptive curve estimation based on inhomogeneous data. *ESAIM Probab. Stat.* **11** 344–364 (electronic). [MR2339297](#)
- GAÏFFAS, S. (2009). Uniform estimation of a signal based on inhomogeneous data. *Statist. Sinica* **19** 427–447. [MR2514170](#)
- GOLUBEV, Y. (2010). On universal oracle inequalities related to high-dimensional linear models. *Ann. Statist.* **38** 2751–2780. [MR2722455](#)
- GURDEV, L. L., DREISCHUH, T. N. and STOYANOV, D. V. (2002). High-range-resolution velocity-estimation techniques for coherent Doppler lidars with exponentially shaped laser pulses. *Appl. Opt.* **41** 1741–1749.
- HARSDORF, S. and REUTER, R. (2000). Stable deconvolution of noisy lidar signals. In *Proceedings of EARSeL-SIG-Workshop LIDAR*. Dresden, FRG.
- HOFFMANN, M. and REISS, M. (2008). Nonlinear estimation for linear inverse problems with error in the operator. *Ann. Statist.* **36** 310–336. [MR2387973](#)
- JOHNSTONE, I. M., KERKYACHARIAN, G., PICARD, D. and RAIMONDO, M. (2004). Wavelet deconvolution in a periodic setting. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **66** 547–573. [MR2088290](#)
- KALIFA, J. and MALLAT, S. (2003). Thresholding estimators for linear inverse problems and deconvolutions. *Ann. Statist.* **31** 58–109. [MR1962500](#)
- KERKYACHARIAN, G., LEPSKI, O. and PICARD, D. (2001). Nonlinear estimation in anisotropic multi-index denoising. *Probab. Theory Related Fields* **121** 137–170. [MR1863916](#)
- KNAPIK, B. T., VAN DER VAART, A. W. and VAN ZANTEN, J. H. (2011). Bayesian inverse problems with Gaussian priors. *Ann. Statist.* **39** 2626–2657. [MR2906881](#)
- LEPSKI, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466. [MR1091202](#)
- LEPSKI, O. V. (1991). Asymptotically minimax adaptive estimation. I. Upper bounds. Optimally adaptive estimates. *Theory Probab. Appl.* **36** 682–697. [MR1147167](#)
- LEPSKI, O. V., MAMMEN, E. and SPOKOINY, V. G. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929–947. [MR1447734](#)
- LEPSKI, O. V. and SPOKOINY, V. G. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512–2546. [MR1604408](#)
- LI, C. T. and SATTA, R. (2011). On the location-dependent quality of the sensor pattern noise and its implication in multimedia forensics. In: *4th International Conference on Imaging for Crime Detection and Prevention (ICDP 2011)*. London, UK.
- MAIR, B. A. and RUYMGAART, F. H. (1996). Statistical inverse estimation in Hilbert scales. *SIAM J. Appl. Math.* **56** 1424–1444. [MR1409127](#)
- MEYER, Y. (1992). *Wavelets and Operators*. Cambridge Univ. Press, Cambridge. [MR1228209](#)

- MILLER, F. P., VANDOME, A. F. and MCBREWSTER, J. (2009). *Amplitude Modulation*. AlphaScript Publishing.
- PENSKY, M. (2013). Supplement to “Spatially inhomogeneous linear inverse problems with possible singularities.” DOI:[10.1214/13-AOS1166SUPP](https://doi.org/10.1214/13-AOS1166SUPP).
- STARCK, J. L. and PANTIN, E. (2002). Deconvolution in astronomy: A review. *Publ. Astron. Soc. Pac.* **114** 1051–1069.

DEPARTMENT OF MATHEMATICS  
UNIVERSITY OF CENTRAL FLORIDA  
ORLANDO, FLORIDA 32816-1354  
USA  
E-MAIL: [Marianna.Pensky@ucf.edu](mailto:Marianna.Pensky@ucf.edu)