

# On the problem of local adaptive estimation in tomography

LAURENT CAVALIER

*Université Aix-Marseille I, UMR CNRS 6632, CMI, 39 rue Joliot-Curie, 13453 Marseille cedex 13, France. E-mail: cavalier@cmi.univ-mrs.fr*

The principle of tomography is to reconstruct a multidimensional function from observations of its integrals over hyperplanes. We consider here a model of stochastic tomography where we observe the Radon transform  $Rf$  of the function  $f$  with a stochastic error. Then we construct a ‘data-driven’ estimator which does not depend on any a priori smoothness assumptions on the function  $f$ . Considering pointwise mean-squared error, we prove that it has (up to a log) the same asymptotic properties as an oracle. We give an example of Sobolev classes of functions where our estimator converges to  $f(x)$  with the optimal rate of convergence up to a log factor.

*Keywords:* adaptive methods; Radon transform

## 1. Introduction

The aim of tomography is to reconstruct an image by use of its integrals over hyperplanes; see, for example, Herman (1980) and Natterer (1986). This problem appears in different fields (such as medical image processing, nuclear medicine and radar theory) and with various models. The model considered here is the stochastic tomography problem defined in Korostelev and Tsybakov (1991; 1993), where we have  $n$  observations containing stochastic errors. This model corresponds to the case of X-ray tomography, one of the earliest tomographic techniques. We wish to produce an image of the internal structure of a human body from observations of integrals over X-ray beams.

Johnstone and Silverman (1990), Korostelev and Tsybakov (1991; 1993), Donoho and Low (1992) and Donoho (1995) have already described a statistical approach to the problems of positron emission tomography and X-ray tomography. These papers propose estimators which attain the optimal rate of convergence for the Sobolev and Besov classes of functions. However, the definition of these estimators requires knowledge of the smoothness of the function  $f$ .

Thus the problem of adaptation arises quite naturally. We wish to define estimators without any a priori assumptions concerning smoothness properties of the function  $f$ . The problem of minimax adaptive estimation of a nonparametric function has been studied in different papers; see, for example, Efroimovich and Pinsker (1984), Lepskii (1990), Golubev and Nussbaum (1992) and Donoho *et al.* (1995). Here we are interested in spatial adaptivity, as in Donoho and Johnstone (1994) and Lepski and Spokoiny (1995). The aim of spatial adaptive methods is to construct ‘data-driven’ estimators that adapt to pointwise

properties and spatially inhomogeneous smoothness of the function  $f$ . For a more general view and references on spatial adaptation, see Lepski *et al.* (1997).

The motivation for using spatially adaptive methods in tomography is that the spatial inhomogeneities in this context seem important. For example, in a medical application of a scanner, the image  $f$  of a human brain will present spatial irregularity. Thus, using spatial adaptive estimators in the problem of tomography is of real interest. Moreover, tomography is a good setting for asymptotic results since the number of observations may be large (of the order of 10 million).

From a mathematical point of view, the construction of estimators by adaptive methods in the framework of tomography may be viewed as part of a larger problem. Indeed, in the problem of tomography we are confronted with indirect observations; we observe integrals of a function  $f$  and we want to reconstruct  $f$  in the whole space. Such a problem of reconstruction from indirect observations is called an inverse problem. Inverse problems appear in fields as different as medical image processing and astronomy; for a statistical overview, see O'Sullivan (1986). Thus, adaptation in inverse problems is different from the normal run of adaptive results which are concerned with estimation of a nonparametric regression or a probability density which is directly observed. There are other results for adaptation in inverse problems; see Efromovich (1997) and Goldenshluger (1999) for the problem of deconvolution, and Johnstone (1999) for different inverse problems using wavelet–vaguelette decomposition. One point to note here is that we consider a difficult inverse problem and that the results are obtained in a ‘regression-type’ model and not in an idealized Gaussian white noise model. Thus, keeping in mind that adaptive procedures are more convenient in applications, results obtained in a realistic model are clearly of interest.

Lepski *et al.* (1997) proposed a variable bandwidth kernel estimator for the problem of estimating a univariate function in a Gaussian white noise. In this spirit, but in a different context, we define in Section 2 a ‘data-driven’ estimator in the problem of X-ray tomography. Then, for a very wide class of functions  $f$ , we prove in Theorem 1 that there exists some unknown oracle which attains (up to a log) the optimal risk. In Theorem 2 we prove that our ‘data-driven’ estimator mimics the asymptotic properties of the oracle and thus presents (in some sense) the best possible choice of bandwidth. In Theorem 3, we study as an example the estimation on the Sobolev classes of functions  $H(\beta, C)$  for unknown parameters  $\beta$  and  $C$ . On these classes, the estimator is proved to be nearly optimal (up to a log factor). In Section 3 we give the proofs.

## 2. Model and results

Let  $f: \mathbb{R}^d \rightarrow \mathbb{R}$ ,  $f \in L_1(\mathbb{R}^d)$ ,  $d \geq 2$ . Define the Radon transform  $Rf$  of  $f$  as

$$Rf(s, u) = \int_{w: \langle w, s \rangle = u} f(w) dw, \quad (1)$$

where  $u \in \mathbb{R}$ ,  $s \in S^{d-1}$ ,  $S^{d-1} = \{w \in \mathbb{R}^d, |w| = 1\}$  is the unit sphere in  $\mathbb{R}^d$ ,  $|w|$  is the Euclidean norm of  $w$ , and  $\langle \cdot, \cdot \rangle$  is the scalar product in  $\mathbb{R}^d$ . The function  $Rf(s, u)$  is defined on the cylinder  $\mathcal{Z} = S^{d-1} \times \mathbb{R}$ . The Radon transform  $Rf(s, u)$  represents the integral of  $f$

over the hyperplane  $\{w \in \mathbb{R}^d, \langle w, s \rangle = u\}$ . Thus, tomography is a particular inverse problem where we want to invert the Radon transform. For general properties of the Radon transform, we refer to Natterer (1986).

Here we only give the following important theorem, which can be found in Natterer (1986, p. 11).

**Lemma 1. (Projection theorem).** *We have for  $f \in L_1(\mathbb{R}^d) \cap L_2(\mathbb{R}^d)$ ,  $\forall s \in S^{d-1}$ , and  $t \in \mathbb{R}$ ,*

$$\widehat{Rf}(s, t) = \hat{f}(ts),$$

where  $\widehat{Rf}$  is the Fourier transform of  $Rf$  over the second argument only.

Define here the class  $B(L)$ ,  $L > 0$ , of functions  $f: \mathbb{R}^d \rightarrow \mathbb{R}$  such that

$$f(w) = 0, \quad |w| > 1,$$

and

$$|f(w)| \leq L, \quad \forall w \in \mathbb{R}^d.$$

For  $f \in B(L)$ , the Radon transform  $Rf$  is well defined.

Consider the statistical model

$$Y_i = Rf(s_i, u_i) + \varepsilon_i, \quad i = 1, \dots, n, \quad (2)$$

where  $(s_i, u_i)$  are independent random variables uniformly distributed on  $S^{d-1} \times [-1, 1]$ ,  $\varepsilon_i$  are independent and identically distributed (i.i.d.) random variables,  $\varepsilon_i \sim \mathcal{N}(0, \sigma^2)$ , with  $\sigma^2 > 0$ , and  $(\varepsilon_1, \dots, \varepsilon_n)$  are independent of  $((s_1, u_1), \dots, (s_n, u_n))$ . The problem is to reconstruct  $f$  from noisy observations  $Y_i$  of  $Rf$ . This setting models the problem of X-ray tomography which appears, for example, in medical image processing. This model has been studied in Korostelev and Tsybakov (1991; 1993) for the case of functions that belong to a Sobolev ball. Here, we have very weak conditions on  $f$ . Indeed,  $f \in B(L)$  is only supposed to be compactly supported and bounded, and in tomography observed functions are compactly supported.

Our aim is to construct an estimator which does not depend on any smoothness assumptions on  $f$ . The idea is to define a class of estimators which depend on a bandwidth parameter  $\delta > 0$ . Then we define a way to choose  $\delta$  adaptively.

Let

$$K_\delta(u) = 2\rho_d(2\pi)^{-d} \int_0^{1/\delta} r^{d-1} \cos(ur) dr, \quad (3)$$

where  $\rho_d$  is the surface area of the sphere  $S^{d-1}$ ,  $\delta > 0$  is a bandwidth,  $u \in \mathbb{R}$ . The Fourier transform of  $K_\delta$  is

$$\hat{K}_\delta(t) = \int_{\mathbb{R}} K_\delta(u) e^{iut} du = \rho_d(2\pi)^{1-d} |t|^{d-1} I_\delta(t), \quad (4)$$

where  $t \in \mathbb{R}$  and  $I_\delta(t) = I(|t| \leq 1/\delta)$  is the indicator function of the interval  $|t| \leq 1/\delta$ . This function  $K_\delta$  will be used as a kernel of the estimator. One may remark that  $K_\delta$  does not

satisfy all the standard properties of kernels. In fact,  $K_\delta$  is called a band-limited filter; it has a compactly supported Fourier transform. This filter has already been used in the context of tomography in Natterer (1986) and Korostelev and Tsybakov (1991; 1993). Using Lemma 1, one may understand that the factor  $|t|^{d-1}$  has been chosen in order to give  $K_\delta$  the property of approximately inverting the Radon transform.

The estimators studied will be of the form

$$\bar{f}_\delta(x) = \frac{1}{n} \sum_{i=1}^n K_\delta(\langle s_i, x \rangle - u_i) Y_i, \quad (5)$$

for  $\delta > 0$ . Thus, this ‘kernel’ estimator  $\bar{f}_\delta(x)$  will be able to approximately invert the Radon transform and to estimate  $f$  from observations of  $Rf$ .

A natural idea in minimax estimation would be to choose the ‘optimal’ bandwidth in order to balance the bias term and the stochastic term entering the risk of the estimator  $\bar{f}_\delta(x)$ . This choice would depend on the smoothness of the function  $f$  being reconstructed. In our problem we do not assume any smoothness for the function  $f$ . Thus, we wish to define a ‘data-driven’ method of bandwidth selection and not a selection based on a priori assumptions on the function  $f$ .

Furthermore, in order to adapt to pointwise properties of the function  $f$ , this bandwidth selector will be local. This means that for each point  $x \in \mathbb{R}^d$  the selected bandwidth will possibly be different in order to take into account the spatially inhomogeneous smoothness of the function  $f$ .

Thus, we propose a local bandwidth selector  $\bar{\delta}(x)$ . We suppose that  $\bar{\delta}(x)$  takes its values in the geometrical grid

$$\Delta_n = \{\delta \in [\delta_n^-, \delta_n^*] : \delta = a^{-j} \delta_n^*, j = 0, 1, 2, \dots\}, \quad (6)$$

where  $a > 1$  is a fixed constant,  $\{\delta_n^*\}$  is a sequence such that  $a \log(\delta_n^* n) / n \leq \delta_n^* \leq 1$  and  $\delta_n^- = \log(\delta_n^* n) / n$ . For example, for  $a = 2$  and  $\delta_n^* = 1$ , this grid contains the values  $(1, \frac{1}{2}, \frac{1}{4}, \dots)$ . The number of points in the grid is a log of the observations.

We need some definitions and properties of the estimators of the form (5). Write

$$f_\delta(x) = E_f \bar{f}_\delta(x) = \frac{1}{2\rho_d} \int_{S^{d-1}} \int_{-1}^1 K_\delta(\langle s, x \rangle - u) Rf(s, u) du ds \quad (7)$$

and

$$v^2(\delta) = c^* \frac{\delta^{1-2d}}{n}, \quad (8)$$

where  $c^* = (\mu_d^2 L^2 + \sigma^2) \rho_d^2 (2\pi)^{1-2d} / (2d - 1)$ ,  $\mu_d$  is the area of the unit ball in  $\mathbb{R}^{d-1}$  and  $E_f$  ( $\text{var}_f$ ) denotes the mathematical expectation (variance) with respect to the observations if the true function is  $f$ . We have:

**Lemma 2.** *If  $f \in B(L)$  then the variance  $\text{var}_f \bar{f}_\delta(x)$  of  $\bar{f}_\delta(x)$  satisfies*

$$\text{var}_f \bar{f}_\delta(x) \leq v^2(\delta),$$

where  $v^2(\delta)$  is defined in (8).

**Proof.** By definition,

$$\begin{aligned}\text{var}_f \bar{f}_\delta(x) &= \text{var}_f \left( \frac{1}{n} \sum_{i=1}^n K_\delta(\langle s_i, x \rangle - u_i) Y_i \right) \\ &= \text{var}_f \left( \frac{1}{n} \sum_{i=1}^n K_\delta(\langle s_i, x \rangle - u_i) Rf(s_i, u_i) \right) + \text{E}_f \left( \frac{1}{n^2} \sum_{i=1}^n K_\delta^2(\langle s_i, x \rangle - u_i) \varepsilon_i^2 \right).\end{aligned}$$

The fact that the function  $f$  belongs to  $B(L)$  implies that  $|Rf(s, u)| \leq \mu_d L$ . Then,

$$\text{var}_f \bar{f}_\delta(x) \leq \frac{\mu_d^2 L^2 + \sigma^2}{n} \text{E}_f(K_\delta^2(\langle s_i, x \rangle - u_i)) \leq \frac{\mu_d^2 L^2 + \sigma^2}{2n} \int_{\mathbb{R}} K_\delta^2(u) du.$$

Using the Parseval equality and (4), we obtain Lemma 2.  $\square$

Now define, for  $\delta > \eta$ ,

$$\psi(\delta, \eta) = v(\delta)\lambda(\delta) + v(\delta, \eta)\lambda(\eta). \quad (9)$$

Here

$$\lambda(\delta) = \max \left( 1, \left( c_2 \log \frac{\delta_n^*}{\delta} \right)^{1/2} \right), \quad (10)$$

where  $c_2 > 8(2d - 1)$ , and

$$v^2(\delta, \eta) = \frac{\mu_d^2 L^2 + \sigma^2}{2n} \int_{\mathbb{R}} (K_\delta(u) - K_\eta(u))^2 du. \quad (11)$$

Similarly to Lemma 2,

$$\text{var}_f(\bar{f}_\delta(x) - \bar{f}_\eta(x)) \leq v^2(\delta, \eta). \quad (12)$$

Define, for any fixed  $x \in \mathbb{R}^d$ , the bandwidth

$$\bar{\delta}(x) = \max\{\delta \in \Delta_n : |\bar{f}_\delta(x) - \bar{f}_\eta(x)| \leq \psi(\delta, \eta), \text{ for all } \eta \leq \delta, \eta \in \Delta_n\}; \quad (13)$$

clearly  $\bar{\delta}(x)$  is well defined. The choice of  $\bar{\delta}(x)$  means that we take the largest bandwidth  $\delta$  such that the estimators  $\bar{f}_\delta(x)$  and  $\bar{f}_\eta(x)$  are not too different, in some sense, for all  $\eta \leq \delta$ . The idea is that there is no point in choosing a finer resolution, i.e. a smaller bandwidth, if we cannot see more specific details. Thus, as long as the estimators are ‘close’ we can take a larger bandwidth. Here ‘close’ roughly means when the difference between the estimators is smaller than the error in estimation. This idea may be found in Lepskii (1990).

With this choice of bandwidth, instead of the usual deterministic bandwidth, we define the estimator of  $f(x)$ :

$$f^*(x) = \bar{f}_{\bar{\delta}(x)}(x) = \frac{1}{n} \sum_{i=1}^n K_{\bar{\delta}(x)}(\langle s_i, x \rangle - u_i) Y_i. \quad (14)$$

One may remark that this estimator does not belong to the class defined in (5), since the bandwidth  $\bar{\delta}(x)$  is not deterministic and depends on the observations. In particular,  $f^*(x)$  is not a linear estimator. Furthermore, the estimator  $f^*(x)$  does not depend on any assumptions on the smoothness of the unknown function  $f$ . Observe that  $\bar{\delta}(x)$  is defined for each point  $x$  and is not a global bandwidth. In this sense, the choice of  $\bar{\delta}(x)$  is said to be spatially adaptive; it will be a local choice independent of any a priori smoothness.

Define

$$\delta_n = \delta_n(x, f) = \max \left\{ \delta \in \Delta_n : |f_\eta(x) - f(x)| \leq \frac{v(\delta)\lambda(\delta)}{2} \text{ for all } \eta \in \Delta_n, \eta \leq \delta \right\}. \quad (15)$$

If the set whose maximum is taken is empty, then we set  $\delta_n = \delta_n^-$ . In the proof of Theorem 2 we show that, for  $d > 2$  and  $n$  large enough, the set is non-empty. If  $d = 2$ , the choice  $\delta_n^-$  does not modify the results.

This local bandwidth  $\delta_n$  clearly depends on the unknown smoothness of  $f$  at point  $x \in \mathbb{R}^d$ . Indeed, when the smoothness of  $f$  increases, the bias term becomes smaller and thus  $\delta_n$  increases. In fact,  $\delta_n$  is some unknown ‘ideal’ bandwidth such that the bias of the estimator  $\bar{f}_\eta(x)$  is smaller than the ‘rate of convergence’  $v(\delta_n)\lambda(\delta_n)$  for all  $\eta \leq \delta_n$ .

The idea now is to study the properties of  $\delta_n$ . Observe that the usual decomposition of the mean-squared risk gives

$$E_f(\bar{f}_\delta(x) - f(x))^2 = (f_\delta(x) - f(x))^2 + \text{var}_f \bar{f}_\delta(x) \leq (f_\delta(x) - f(x))^2 + c^* \frac{\delta^{1-2d}}{n}. \quad (16)$$

The usual way to obtain the optimal rate of convergence is to choose the bandwidth which balances the two terms of the right-hand side of (16). This trade-off depends on the smoothness of the function  $f$  at point  $x$ , which is unknown here. Define

$$r_n(x, f) = \inf_{\delta_n^- \leq \delta \leq 1} \left\{ \sup_{0 \leq \eta \leq \delta} (f_\eta(x) - f(x))^2 + c^* \delta^{1-2d} \frac{\log n}{n} \right\},$$

which is the right-hand term of (16) modified by a log factor. It is known from Lepskii (1990) and Brown and Low (1996) that in pointwise estimation, one incurs an additional cost in the form of a logarithmic factor when the smoothness of  $f$  is unknown. Then, we have:

**Theorem 1.** *Let  $\delta_n^* = 1$ . For any  $f \in B(L)$  we have, as  $n \rightarrow \infty$ ,*

$$E_f[(\bar{f}_{\delta_n}(x) - f(x))^2] \leq \frac{5}{4} v(\delta_n)\lambda(\delta_n) \leq C(a)r_n(x, f),$$

where  $C(a)$  is a positive constant depending on  $a$  defined in (6).

In the setting of a signal observed with additive Gaussian noise, Lepski and Spokoiny (1997) and Lepski *et al.* (1997) call the term  $r_n(x, f)$  the adaptive rate of convergence. We can conjecture that in the tomography problem,  $r_n(x, f)$  may also be considered as an

adaptive rate. Hence, Theorem 1 would mean that the ‘estimator’  $\bar{f}_{\delta_n}$  has a risk bounded by the best possible risk in the context of adaptive estimation  $r_n(x, f)$ . Thus,  $\delta_n$  would realize (up to a log) the perfect trade-off in the risk, and might be considered as some optimal choice of bandwidth.

The main problem is that  $\bar{f}_{\delta_n}$  is not an estimator but what is called an oracle, since it depends on the function  $f$  itself.

Our aim is to show that the estimator  $f^*(x)$  constructed with the bandwidth  $\bar{\delta}(x)$  has a risk bounded by the same value as the risk of the oracle  $\bar{f}_{\delta_n}(x)$ .

We obtain the following theorem, which is quite general since the function is only supposed to be in  $B(L)$ .

**Theorem 2.** *We have uniformly over  $f \in B(L)$ ,  $L > 0$  and  $x \in \mathbb{R}^d$ ,*

$$E_f[(f^*(x) - f(x))^2] \leq c(a)(v(\delta_n)\lambda(\delta_n))^2,$$

*and thus, for  $\delta_n^* = 1$ , we have*

$$E_f[(f^*(x) - f(x))^2] \leq c'(a)r_n(x, f),$$

*as  $n \rightarrow \infty$ , where  $c(a)$  and  $c'(a)$  are positive constants depending on  $a$  defined in (6).*

Theorem 2 proves that the estimator  $f^*(x)$  mimics the asymptotic properties of the oracle. Thus, for each  $f$  our estimator has (up to a log term) the same risk as the risk of the estimator constructed with the bandwidth  $\delta_n$  which may be considered as optimal. Furthermore, following Lepskii (1990) and Donoho and Johnstone (1994), we know that in the standard setting of a signal observed with additive Gaussian noise, the log term in the rate of convergence is the price paid for pointwise adaptation. Thus, in the setting of tomography, there is no reason for this log not to disappear.

This is of special interest in applications, since we have constructed an estimator which is, for each  $f$  with minimal assumptions, close to the best possible oracle. Thus the estimator adapts to a particular function  $f$  and not to classes of functions, as usual in minimax estimation.

We conjectured that the results from direct models could be transposed to the indirect setting of tomography. This point should have to be proved. However, even if Theorem 2 meant that the estimator is in some sense the ‘best’ possible estimator for a given function  $f$ , we did not know if it is at least a ‘good’ estimator. Indeed, without any assumptions on  $f$  we cannot conclude anything concerning the asymptotic properties of the oracle  $\bar{f}_{\delta_n}$ . In particular, it is impossible to obtain minimax results on the class  $B(L)$ , which is too large.

Thus, the evaluation of the estimator  $f^*(x)$  using a conjecture may be enough, since, even with a more mathematical result, we could not say that  $f^*(x)$  is a good estimator.

A more standard way to appreciate the accuracy of the estimator  $f^*(x)$ , is to compare it with other estimators. We consider as an example the case of functions in some Sobolev ball. These classes of functions are of interest in the problem of tomography (see Natterer 1980; 1986). One reason for studying Sobolev classes in tomography is that smoothness conditions in term of Fourier transforms are of natural interest in this problem. However, other classes of functions need to be studied (Besov classes, analytic functions).

A function  $f \in H(\beta, C)$ ,  $\beta > d/2$ ,  $C > 0$ , is such that

$$f(w) = 0, \quad |w| > 1,$$

and

$$\int_{\mathbb{R}^d} (1 + |\omega|^2)^\beta |\hat{f}(\omega)|^2 d\omega \leq C,$$

where  $\hat{f}$  is the Fourier transform of  $f$ ,  $\hat{f}(\omega) = \int_{\mathbb{R}^d} f(x) e^{i\omega \cdot x} dx$ . On this class, Korostelev and Tsybakov (1991; 1993) proved that the optimal rate of convergence for the mean-squared error is  $n^{-(2\beta-d)/(2\beta+d-1)}$ . Here, we have the following result.

**Theorem 3.** *Let  $L > 0$  and let  $\delta_n^* = 1$ . Then for any fixed  $x \in \mathbb{R}^d$ , we have, for any  $\beta > d/2$  and  $C > 0$  unknown,*

$$\limsup_{n \rightarrow \infty} \sup_{f \in H(\beta, C) \cap B(L)} \left( \frac{n}{\log n} \right)^{(2\beta-d)/(2\beta+d-1)} E_f[(f^*(x) - f(x))^2] < \infty.$$

Theorem 3 says that  $f^*(x)$  has almost the optimal rate of convergence for each  $\beta$  and  $C$ . This means that  $f^*(x)$  has (up to a log term) the best rate of convergence among all the estimators. Thus, in the standard case of functions belonging to some Sobolev ball,  $f^*(x)$  is almost optimal.

Furthermore,  $f^*(x)$  does not depend on  $\beta$  and  $C$  and it is then more realistic to construct it in practical cases than classical estimators which need the knowledge of an a priori smoothness. The only cost incurred is the loss of a logarithmic factor in the rate of convergence. In the direct case, Tsybakov (1998) proved that for the pointwise risk and the Sobolev classes, the loss due to adaptation in the rate of convergence is a log term. Once more, we can conjecture that this log term is also a cost incurred in the tomography problem. Nevertheless, what is important is that we only lose a log term in the rate of convergence which will be negligible for applications.

Tomography is a specific inverse problem. Thus, it would be natural to extend the methods described in this paper to other inverse problems, especially all the problems described in Donoho (1995) – Abel transforms, integration, etc. This would be of interest for comparison between ‘adaptive kernels’ and wavelet–vaguelette decomposition methods. We can say roughly here that in wavelet–vaguelette decomposition we make a projection on the vaguelette basis and then we obtain the empirical wavelet coefficient, which is related to the value  $w_{j,k} = \bar{f}_{2^{-j}\delta_n^*}(x) - \bar{f}_{2^{-j-1}\delta_n^*}(x)$ . Thus, when using thresholding, we are making a comparison between close resolution levels, and then important features may exist at each level.

### 3. Proofs

In order to prove Theorem 1 we require the following lemma:

**Lemma 3.** *We have, for all  $\delta$ ,*



$$f_\delta(x) = (2\pi)^{-d} \int_{\mathbb{R}^d} \hat{f}(\omega) I_\delta(|\omega|) e^{-i\langle \omega, x \rangle} d\omega.$$

**Proof of Lemma 3.** We have that  $Rf(s, u) = 0$  for  $|u| > 1$  since  $f \in B(L)$ . Thus, using properties of the inverse Fourier transform, the convolution product and Lemma 1, we obtain

$$\begin{aligned} f_\delta(x) &= \frac{1}{2\rho_d} \int_{S^{d-1}} \int_{\mathbb{R}} K_\delta(\langle s, x \rangle - u) Rf(s, u) du ds = \frac{(2\pi)^{-1}}{2\rho_d} \int_{S^{d-1}} \int_{\mathbb{R}} \hat{K}_\delta(t) \widehat{Rf}(s, t) e^{-it\langle s, x \rangle} dt ds \\ &= (2\pi)^{-d} \int_{S^{d-1}} \int_0^\infty t^{d-1} I_\delta(t) \hat{f}(ts) e^{-i\langle ts, x \rangle} dt ds = (2\pi)^{-d} \int_{\mathbb{R}^d} \hat{f}(\omega) e^{-i\langle \omega, x \rangle} I_\delta(|\omega|) d\omega. \end{aligned}$$

□

**Proof of Theorem 1.** In order to prove Theorem 1, we need first to prove that  $\delta_n$  is well defined, i.e. that the maximum in (15) is taken over a non-empty set. Let  $\delta = \delta_n^-$ . By use of Lemma 3, we obtain

$$\begin{aligned} |f_{\delta_n^-}(x) - f(x)|^2 &\leq 2|f_{\delta_n^-}(x)|^2 + 2|f(x)|^2 \\ &\leq 2(2\pi)^{-2d} \left| \int_{\mathbb{R}^d} \hat{f}(\omega) I(|\omega| \leq 1/\delta_n^-) e^{-i\langle \omega, x \rangle} d\omega \right|^2 + 2|f(x)|^2 \\ &\leq 2(2\pi)^{-2d} \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 d\omega \int_{\mathbb{R}^d} I(|\omega| \leq 1/\delta_n^-) d\omega + 2|f(x)|^2. \end{aligned}$$

Since,  $f \in B(L)$ , this implies that

$$|f_{\delta_n^-}(x) - f(x)|^2 \leq c_3 (\delta_n^-)^{-d},$$

where  $c_3$  is a positive constant. Using (8) and (10), observe that

$$v^2(\delta_n^-) \lambda^2(\delta_n^-) = c^* c_2 (\delta_n^-)^{2-2d}.$$

Thus, for  $d > 2$ , for  $n$  large enough,

$$|f_{\delta_n^-}(x) - f(x)| \leq \frac{1}{2} v(\delta_n^-) \lambda(\delta_n^-).$$

Therefore,  $\delta_n$  is well defined in (15) for  $d > 2$  since the set is non-empty.

For  $d = 2$ , one may observe that  $\delta_n$  is not always defined. This is of no importance since  $(f_{\delta_n^-}(x) - f(x))$  and  $v(\delta_n^-) \lambda(\delta_n^-)$  are of the same order. Thus, we fix  $\delta_n = \delta_n^-$  if  $\delta_n$  is not defined.

Using the standard decomposition of the risk as in (16), the definition of  $\delta_n$  in (15) and Lemma 2, we obtain

$$\begin{aligned} E_f(\bar{f}_{\delta_n}(x) - f(x))^2 &= (f_{\delta_n}(x) - f(x))^2 + \text{var}_f \bar{f}_{\delta_n}(x) \leq \frac{v^2(\delta_n) \lambda^2(\delta_n)}{4} + v^2(\delta_n) \\ &\leq \frac{5}{4} v^2(\delta_n) \lambda^2(\delta_n). \end{aligned} \tag{17}$$

Now, suppose that the infimum in the definition of  $r_n(x, f)$  is attained for  $\delta = \delta_0$ . If  $\delta_0 < a\delta_n$  then, for  $C$  large enough,

$$Cr_n(x, f) \geq Cc^* \delta_n^{1-2a} \frac{\log n}{n} \geq \frac{5}{4} v^2(\delta_n) \lambda^2(\delta_n). \quad (18)$$

If  $\delta_0 \geq a\delta_n$  then, by use of the definition of  $\delta_n$  in (15), we obtain

$$\sup_{0 \leq \eta \leq \delta_0} (f_\eta(x) - f(x))^2 \geq \sup_{0 \leq \eta \leq a\delta_n} (f_\eta(x) - f(x))^2 > \frac{v^2(a\delta_n) \lambda^2(a\delta_n)}{4},$$

and thus

$$Cr_n(x, f) \geq \frac{5}{4} v^2(\delta_n) \lambda^2(\delta_n). \quad (19)$$

Using (18) and (19) in (17), we have the result for  $C$  large enough.  $\square$

**Proof of Theorem 2.** Decompose the quadratic risk as follows:

$$E_f[(f^*(x) - f(x))^2] = E_f[(f^*(x) - f(x))^2 I(\bar{\delta} \geq \delta_n)] + E_f[(f^*(x) - f(x))^2 I(\bar{\delta} < \delta_n)],$$

where we have written  $\bar{\delta}$  for  $\bar{\delta}(x)$ . We have two cases to study.

*Case 1:*  $\{\bar{\delta} \geq \delta_n\}$ . Note that  $\delta' \geq \delta$  implies that  $v(\delta) \geq v(\delta')$  and  $\lambda(\delta) \geq \lambda(\delta')$ . Thus,

$$\psi(\delta', \delta) \leq v(\delta) \lambda(\delta) + v(\delta', \delta) \lambda(\delta) \leq 3v(\delta) \lambda(\delta).$$

By the use of (9) and the definition of  $\bar{\delta}$  in (13), we have

$$|f^*(x) - \bar{f}_{\delta_n}(x)| I(\bar{\delta} \geq \delta_n) \leq \psi^*(\delta_n) \leq 3v(\delta_n) \lambda(\delta_n), \quad (20)$$

where  $\psi^*(\delta) = \sup\{\psi(\delta', \delta) : \delta' \in \Delta_n, \delta' > \delta\}$ . Thus,

$$\begin{aligned} E_f[(f^*(x) - f(x))^2 I(\bar{\delta} \geq \delta_n)] \\ \leq 2(E_f[(f^*(x) - \bar{f}_{\delta_n}(x))^2 I(\bar{\delta} \geq \delta_n)] + E_f[\bar{f}_{\delta_n}(x) - f_{\delta_n}(x)]^2 + (f_{\delta_n}(x) - f(x))^2). \end{aligned}$$

Using Lemma 2, (15) and (20), we obtain

$$E_f[(f^*(x) - f(x))^2 I(\bar{\delta} \geq \delta_n)] \leq 2 \left( 9v^2(\delta_n) \lambda^2(\delta_n) + v^2(\delta_n) + \frac{1}{4} v^2(\delta_n) \lambda^2(\delta_n) \right).$$

Finally,

$$E_f[(f^*(x) - f(x))^2 I(\bar{\delta} \geq \delta_n)] \leq c_4 v^2(\delta_n) \lambda^2(\delta_n), \quad (21)$$

where  $c_4$  is a positive constant.

*Case 2:*  $\{\bar{\delta} < \delta_n\}$ . Write

$$B_n(x, \delta, \eta) = \{|\bar{f}_\delta(x) - \bar{f}_\eta(x)| > \psi(\delta, \eta)\},$$

where  $\eta, \delta \in \Delta_n, \delta > \eta$ . By the definition of  $\bar{\delta}$ ,

$$\forall \delta \in \Delta_n, \quad \{\bar{\delta} = a^{-1}\delta\} \subset \bigcup_{\eta \in \Delta_n(\delta)} B_n(x, \delta, \eta),$$

where  $\Delta_n(\delta) = \{\eta \in \Delta_n, \eta < \delta\}$ . Therefore,

$$\{\bar{\delta} < \delta_n\} \subset \bigcup_{\delta \in \Delta_n(a\delta_n)} \bigcup_{\eta \in \Delta_n(\delta)} B_n(x, \delta, \eta).$$

Thus,

$$\begin{aligned} \mathbb{E}_f[(f^*(x) - f(x))^2 I(\bar{\delta} < \delta_n)] &\leq \sum_{\delta \in \Delta_n(a\delta_n)} \mathbb{E}_f[(\bar{f}_{a^{-1}\delta}(x) - f(x))^2 I(\bar{\delta} = a^{-1}\delta)] \\ &\leq \sum_{\delta \in \Delta_n(a\delta_n)} \sum_{\eta \in \Delta_n(\delta)} \mathbb{E}_f[(\bar{f}_{a^{-1}\delta}(x) - f(x))^2 I(B_n(x, \delta, \eta))]. \end{aligned}$$

By the definition of  $\delta_n$  in (15),  $a^{-1}\delta < \delta_n$  implies that

$$|f_{a^{-1}\delta}(x) - f(x)| \leq \frac{1}{2}v(\delta_n)\lambda(\delta_n) \leq \frac{1}{2}v(\delta)\lambda(\delta).$$

And, for any  $\eta < \delta \leq \delta_n$ ,

$$|f_\eta(x) - f(x)| \leq \frac{1}{2}v(\delta_n)\lambda(\delta_n) \leq \frac{1}{2}v(\delta)\lambda(\delta).$$

This implies that

$$\begin{aligned} B_n(x, \delta, \eta) &= \{|\bar{f}_\delta(x) - \bar{f}_\eta(x) - (f_\delta(x) - f_\eta(x)) + f_\delta(x) - f_\eta(x) - f(x) + f(x)| > \psi(\delta, \eta)\} \\ &\subset \left\{ v(\delta)\lambda(\delta) + \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| > \psi(\delta, \eta) \right\}, \end{aligned}$$

where  $\xi_i = (K_\delta(\langle s_i, x \rangle - u_i) - K_\eta(\langle s_i, x \rangle - u_i))Y_i - (f_\delta(x) - f_\eta(x))$ . Observe by (12) that the  $\xi_i$  are i.i.d. random variables with zero mean and variance  $\text{var}_f \xi_1 \leq nv^2(\delta, \eta)$ .

Using (9), simplify the previous display to give

$$B_n(x, \delta, \eta) \subset \left\{ \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| > v(\delta, \eta)\lambda(\eta) \right\}.$$

Therefore,

$$\begin{aligned} \mathbb{E}_f[(f^*(x) - f(x))^2 I(\bar{\delta} < \delta_n)] \\ \leq \sum_{\delta \in \Delta_n(a\delta_n)} \sum_{\eta \in \Delta_n(\delta)} \mathbb{E}_f \left[ (\bar{f}_{a^{-1}\delta}(x) - f(x))^2 I \left( \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| > v(\delta, \eta)\lambda(\eta) \right) \right]. \end{aligned} \quad (22)$$

Using the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned} & \mathbb{E}_f \left[ (\bar{f}_{a^{-1}\delta}(x) - f(x))^2 I \left( \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| > v(\delta, \eta) \lambda(\eta) \right) \right] \\ & \leq (\mathbb{E}_f [(\bar{f}_{a^{-1}\delta}(x) - f(x))^4])^{1/2} \left( \mathbb{P}_f \left( \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| > v(\delta, \eta) \lambda(\eta) \right) \right)^{1/2}. \end{aligned} \quad (23)$$

We have

$$\mathbb{P}_f \left( \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| > v(\delta, \eta) \lambda(\eta) \right) \leq \mathbb{P}_f \left( \frac{1}{n} \sum_{i=1}^n \xi_i > v(\delta, \eta) \lambda(\eta) \right) + \mathbb{P}_f \left( -\frac{1}{n} \sum_{i=1}^n \xi_i > v(\delta, \eta) \lambda(\eta) \right). \quad (24)$$

In fact, we only study the first probability on the right-hand side of (24). Using Chebychev's exponential inequality, we obtain

$$\mathbb{P}_f \left( \frac{1}{n} \sum_{i=1}^n \xi_i > v(\delta, \eta) \lambda(\eta) \right) \leq \mathbb{E}_f \left[ \exp \left( \frac{z}{n} \sum_{i=1}^n \xi_i \right) \right] \exp(-zv(\delta, \eta) \lambda(\eta)), \quad \forall z \geq 0.$$

Since the  $\xi_i$  are i.i.d., we have

$$\mathbb{E}_f \left[ \exp \left( \frac{z}{n} \sum_{i=1}^n \xi_i \right) \right] = \left( \mathbb{E}_f \left[ \exp \frac{z}{n} \xi_1 \right] \right)^n.$$

Now use the conditional expectation. Let  $(S, U)$  be a uniform random variable on  $S^{d-1} \times [-1, 1]$ . Write  $K_\delta^\eta = K_\delta(\langle S, x \rangle - U) - K_\eta(\langle S, x \rangle - U)$ . We have

$$\mathbb{E}_f \left[ \exp \left( \frac{z}{n} \xi_1 \right) | (S, U) \right] = \exp \left( \frac{z}{n} (K_\delta^\eta Rf(S, U) - f_\delta(x) + f_\eta(x)) \right) \mathbb{E}_f \left[ \exp \left( \frac{z}{n} K_\delta^\eta \varepsilon_1 \right) | (S, U) \right].$$

Since  $\varepsilon_1$  is a zero-mean Gaussian random variable independent of  $(S, U)$ , we finally obtain

$$\begin{aligned} \mathbb{E}_f \left[ \exp \left( \frac{z}{n} \xi_1 \right) | (S, U) \right] &= \exp \left( \frac{z}{n} (K_\delta^\eta Rf(S, U) - f_\delta(x) + f_\eta(x)) + \frac{z^2 \sigma^2}{2n^2} (K_\delta^\eta)^2 \right) \\ &\leq \exp(U_1 + U_2) \exp \left( \frac{z^2 \sigma^2}{2n^2} E(K_\delta^\eta)^2 \right), \end{aligned}$$

where  $U_1 = z/n(K_\delta^\eta Rf(S, U) - f_\delta(x) + f_\eta(x))$  and  $U_2 = z^2 \sigma^2 ((K_\delta^\eta)^2 - E(K_\delta^\eta)^2)/(2n^2)$ , where  $E$  denotes the expectation with respect to the uniform random variable  $(S, U)$ . Observe that  $EU_1 = EU_2 = 0$ ,  $\text{var}_f U_1 \leq (z^2 \mu_d^2 L^2 / 2n^2) \int_{\mathbb{R}} (K_\delta(u) - K_\eta(u))^2 du$  and  $\text{var}_f U_2 \leq z^4 \sigma^4 n^2 v^2(\delta, \eta) / (4n^4)$ . Finally, we obtain for  $z = \rho \lambda(\eta) / v(\delta, \eta)$ , where  $\rho$  is a positive constant, since the term  $U_2$  is small for  $n$  large enough,

$$\left( \mathbb{E}_f \left[ \exp \left( \frac{z}{n} \xi_1 \right) \right] \right)^n \leq \exp \left( \frac{\rho^2 (2\mu_d^2 L^2 + \sigma^2) \lambda^2(\eta)}{4nv^2(\delta, \eta)} \int_{\mathbb{R}} (K_\delta(u) - K_\eta(u))^2 du (1 + o(1)) \right).$$

Thus, by definition of  $v(\delta, \eta)$  in (11), we have, as  $n \rightarrow \infty$ ,

$$P_f \left( \frac{1}{n} \sum_{i=1}^n \xi_i > v(\delta, \eta) \lambda(\eta) \right) \leq \exp(((1 - \kappa)\rho^2 - \rho)\lambda^2(\eta)),$$

where  $\kappa$  is a small positive constant. Finally, we can obtain, as  $n \rightarrow \infty$ ,

$$P_f \left( \left| \frac{1}{n} \sum_{i=1}^n \xi_i \right| > v(\delta, \eta) \lambda(\eta) \right) \leq 2 \exp\left(-\frac{\lambda^2(\eta)}{4}\right) \leq 2 \exp\left(-\frac{c_2}{4} \log \frac{\delta_n^*}{\eta}\right); \quad (25)$$

if  $\eta$  is not small enough, then the result may easily be obtained directly from (22) since  $\eta < \delta \leq \delta_n \leq \delta_n^*$ .

Now study the first term on the right-hand side of (23),

$$\begin{aligned} E_f[(\bar{f}_{a^{-1}\delta}(x) - f(x))^4] &\leq 4E_f[(\bar{f}_{a^{-1}\delta}(x) - f_{a^{-1}\delta}(x))^4] + 4(f_{a^{-1}\delta}(x) - f(x))^4 \\ &\leq 4E_f \left[ \left( \frac{1}{n} \sum_{i=1}^n Z_i \right)^4 \right] + \frac{v^4(\delta_n) \lambda^4(\delta_n)}{4}, \end{aligned}$$

where  $Z_i = \bar{f}_{a^{-1}\delta}(x) - f_{a^{-1}\delta}(x)$ , by the definition of  $\delta_n$  in (15). Observe that the  $Z_i$  are centred random variables with  $\text{var}_f(Z_1) \leq nv^2(a^{-1}\delta)$ . Thus, as  $n \rightarrow \infty$ ,

$$E_f \left[ \left( \frac{1}{n} \sum_{i=1}^n Z_i \right)^4 \right] \leq \frac{E_f Z_1^4}{n^3} + \frac{3(E_f(Z_1^2))^2}{n^2} \leq c'_5 v^4(a^{-1}\delta), \quad (26)$$

where  $c'_5$  is a positive constant.

Finally, using (25) and (26) in (22), we obtain

$$E_f[(f^*(x) - f(x))^2 I(\bar{\delta} < \delta_n)] \leq c_5 \sum_{\delta \in \Delta_n(a\delta_n)} \sum_{\eta \in \Delta_n(\delta)} (v^4(a^{-1}\delta) + v^4(\delta_n) \lambda^4(\delta_n))^{1/2} e^{-\lambda^2(\eta)/8},$$

where  $c_5$  is a positive constant. Using (8), we have  $v^2(a^{-1}\delta) = a^{2d-1} c^* \delta^{1-2d}/n$ . And  $\eta < \delta \leq \delta_n$  implies  $\lambda(\delta_n) \leq \lambda(\eta)$ ,  $v(\delta_n) < v(a^{-1}\delta)$ . Assume that  $c_2/8 > 2d - 1$ . Thus,

$$E_f[(f^*(x) - f(x))^2 I(\bar{\delta} < \delta_n)] \leq c_6 \sum_{\delta \in \Delta_n(a\delta_n)} \sum_{\eta \in \Delta_n(\delta)} \left( \frac{\delta^{1-2d}}{n} \right) \lambda^2(\eta) \left( \frac{\eta}{\delta_n^*} \right)^{c_2/8}, \quad (27)$$

where  $c_6$  is a positive constant which depends on  $a$ . Now we wish to bound the number of elements of the set  $\Delta_n$ . If  $\eta = a^{-j} \delta_n^* \in \Delta_n$  then  $j \leq [\log(\delta_n^*/\eta)] = N_n$ , where  $[\cdot]$  denotes the integer part. Thus, if  $\alpha > 0$  small enough (in fact  $2\alpha \leq c_2/8 - 2d + 1$ ), then

$$\begin{aligned} \sum_{\eta \in \Delta_n(\delta)} \lambda^2(\eta) \left( \frac{\eta}{\delta_n^*} \right)^{c_2/8} &\leq \left( \frac{\delta}{\delta_n^*} \right)^{(c_2/8) - \alpha} \sum_{\eta \in \Delta_n(\delta)} c_2 \log \left( \frac{\delta_n^*}{\eta} \right) \left( \frac{\eta}{\delta_n^*} \right)^\alpha \\ &\leq \left( \frac{\delta}{\delta_n^*} \right)^{(c_2/8) - \alpha} c_2 \sum_{j=0}^{N_n} j \left( \frac{1}{a^\alpha} \right)^j \log a. \end{aligned}$$

This series is convergent when  $a > 1$ ; therefore

$$\sum_{\eta \in \Delta_n(\delta)} \lambda^2(\eta) \left( \frac{\eta}{\delta_n^*} \right)^{c_2/8} \leq c_7 \left( \frac{\delta}{\delta_n^*} \right)^{(c_2/8)-a}, \quad (28)$$

where  $c_7$  is a positive constant. By using (28) in (27) we obtain,

$$E_f[(f^*(x) - f(x))^2 I(\bar{\delta} < \delta_n)] \leq \frac{c_8 \delta_n^{*2d-1}}{n} \sum_{\delta \in \Delta_n(a\delta_n)} \left( \frac{\delta}{\delta_n^*} \right)^{(c_2/8)-a-2d+1},$$

where  $c_8$  is a positive constant. Finally, as the number of elements of the set  $\Delta_n(a\delta_n)$  can be bounded by  $N_n$  and the series converges,

$$E_f[(f^*(x) - f(x))^2 I(\bar{\delta} < \delta_n)] \leq \frac{c_8 \delta_n^{*2d-1}}{n} \sum_{j=0}^{N_n} \left( \frac{1}{a^j} \right)^j \leq c v^2(\delta_n) \lambda^2(\delta_n), \quad (29)$$

where  $c$  is a constant large enough which depends on  $a$ . The theorem follows from (21) and (29).  $\square$

**Proof of Theorem 3.** We have, using Lemma 3,

$$\begin{aligned} (f_\eta(x) - f(x))^2 &= (2\pi)^{-2d} \left( \int_{\mathbb{R}^d} \hat{f}(\omega) I\left(|\omega| > \frac{1}{\eta}\right) e^{-i\langle \omega, x \rangle} d\omega \right)^2 \\ &\leq (2\pi)^{-2d} \left( \int_{\mathbb{R}^d} |\hat{f}(\omega)| \frac{2(\eta|\omega|)^\beta}{1 + (\eta|\omega|)^\beta} d\omega \right)^2 \\ &\leq 4(2\pi)^{-2d} \eta^{2\beta} \int_{\mathbb{R}^d} |\hat{f}(\omega)|^2 |\omega|^{2\beta} d\omega \cdot \int_{\mathbb{R}^d} \frac{d\omega}{1 + (\eta|\omega|)^\beta}. \end{aligned}$$

Thus,  $f \in H(\beta, C)$  implies, for all  $\eta$ ,

$$(f_\eta(x) - f(x))^2 \leq c_9 \eta^{2\beta-d},$$

where  $c_9$  is a positive constant. Let  $\eta \leq \rho(\log n/n)^{1/(2\beta+d-1)}$ , where  $\rho > 0$  is small enough; then, as  $n \rightarrow \infty$ ,

$$(f_\eta(x) - f(x))^2 \leq c_9 \rho^{2\beta-d} \left( \frac{\log n}{n} \right)^{2\beta-d/(2\beta+d-1)} \leq \frac{\rho^{1-2d} c^* c_2}{4(2\beta+d-1)} \left( \frac{\log n}{n} \right)^{2\beta-d/(2\beta+d-1)}.$$

Thus, for all  $\eta \leq \rho(\log n/n)^{1/(2\beta+d-1)}$ , where  $\rho > 0$  is small enough, we have, as  $n \rightarrow \infty$ ,

$$|f_\eta(x) - f(x)| \leq \frac{1}{2} v \left( \rho \left( \frac{\log n}{n} \right)^{1/(2\beta+d-1)} \right) \lambda \left( \rho \left( \frac{\log n}{n} \right)^{1/(2\beta+d-1)} \right).$$

Therefore, by the definition of  $\delta_n(x, f)$  in (15), we have

$$\delta_n(x, f) \geq \rho \left( \frac{\log n}{n} \right)^{1/(2\beta+d-1)}.$$

Finally, we obtain using Theorem 2, as  $n \rightarrow \infty$ ,

$$E_f[(f^*(x) - f(x))^2] \leq c_{10} \left( \frac{\log n}{n} \right)^{2\beta-d/(2\beta+d-1)},$$

where  $c_{10}$  is a positive constant. □

## Acknowledgements

I would like to thank O. Lepski, P. Massart, V. Spokoiny and A. Tsybakov for helpful discussions, and the three referees for comments leading to improvements in the paper.

## References

- Brown, L.D. and Low, M.G. (1996) A constrained risk inequality with applications to nonparametric functional estimation. *Ann. Statist.*, **24**, 2524–2535.
- Donoho, D.L. (1995) Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition. *Appl. Comput. Harmon. Anal.*, **2**, 101–126.
- Donoho, D.L. and Johnstone, I.M. (1994) Ideal spatial adaptation by wavelet shrinkage. *Biometrika*, **81**, 425–455.
- Donoho, D.L. and Low, M.G. (1992) Renormalization exponents and optimal pointwise rates of convergence. *Ann. Statist.*, **20**, 944–970.
- Donoho, D.L., Johnstone, I.M., Kerkyacharian, G. and Picard, D. (1995) Wavelet shrinkage: Asymptopia? *J. Statist. Soc. Ser. B*, **57**, 301–369.
- Efroimovich, S.Yu. and Pinsker, M.S. (1984) Learning algorithm for nonparametric filtering. *Automat. Remote Control*, **45**, 1434–1440.
- Efromovich, S. (1997) Robust and efficient recovery of a signal passed through a filter and then contaminated by non-Gaussian noise. *IEEE Trans. Inform. Theory*, **43**, 1184–1191.
- Goldenshluger, A. (1999) On pointwise adaptive nonparametric deconvolution. *Bernoulli*, **5**, 907–925.
- Golubev, G.K. and Nussbaum, M. (1992) Adaptive spline estimates for nonparametric regression models. *Theory Probab. Appl.*, **37**, 521–529.
- Herman, G.T. (1980) *Image Reconstruction from Projections*. New York: Academic Press.
- Johnstone, I.M. (1999) Wavelet shrinkage for correlated data and inverse problems: adaptivity results. *Statist. Sinica*, **9**, 51–83.
- Johnstone, I.M. and Silverman, B.W. (1990) Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.*, **18**, 251–280.
- Korostelev, A.P. and Tsybakov, A.B. (1991) Optimal rates of convergence of estimators in a probabilistic setup of tomography problem. *Problems Inform. Transmission*, **27**, 73–81.
- Korostelev, A.P. and Tsybakov, A.B. (1993) *Minimax Theory of Image Reconstruction*, Lecture Notes in Statist. 82. New York: Springer-Verlag.
- 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.
- Lepski, O.V. and Spokoiny, V.G. (1995) Local adaptation to inhomogeneous smoothness: Resolution level. *Math. Methods Statist.*, **4**, 239–258.
- Lepski, O.V. and Spokoiny, V.G. (1997) Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.*, **25**, 2512–2546.
- Lepskii, O.V. (1990) On a problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.*, **35**, 454–466.
- Natterer, F. (1980) A Sobolev space analysis of picture recognition. *SIAM J. Appl. Math.*, **39**, 402–411.
- Natterer, F. (1986). *The Mathematics of Computerized Tomography*. Chichester: Wiley.
- O’Sullivan, F. (1986) A statistical perspective on ill-posed inverse problems. *Statist. Sci.*, **1**, 502–527.
- Tsybakov, A.B. (1998) Pointwise and sup-norm sharp adaptive estimation of functions on the Sobolev classes. *Ann. Statist.*, **26**, 2420–2469.

Received December 1997 and revised May 2000