The Annals of Statistics
2003, Vol. 31, No. 1, 30–57

IMAGE DENOISING: POINTWISE ADAPTIVE APPROACH

BY JÖRG POLZEHL AND VLADIMIR SPOKOINY

Weierstrass Institute

A new method of pointwise adaptation has been proposed and studied in Spokoiny [(1998) *Ann. Statist.* **26** 1356–1378] in the context of estimation of piecewise smooth univariate functions. The present paper extends that method to estimation of bivariate grey-scale images composed of large homogeneous regions with smooth edges and observed with noise on a gridded design. The proposed estimator $\hat{f}(x)$ at a point x is simply the average of observations over a window $\widehat{U}(x)$ selected in a data-driven way. The theoretical properties of the procedure are studied for the case of piecewise constant images. We present a nonasymptotic bound for the accuracy of estimation at a specific grid point x as a function of the number of pixels n, of the distance from the point of estimation to the closest boundary and of smoothness properties and orientation of this boundary. It is also shown that the proposed method provides a near-optimal rate of estimation near edges and inside homogeneous regions. We briefly discuss algorithmic aspects and the complexity of the procedure. The numerical examples demonstrate a reasonable performance of the method and they are in agreement with the theoretical issues. An example from satellite (SAR) imaging illustrates the applicability of the method.

1. Introduction. One typical problem of image analysis is the reconstruction of an image from noisy data. It has been intensively studied within the last years; see, for example, the books of Pratt (1978), Grenander (1976, 1978, 1981), Rosenfeld and Kak (1982), Blake and Zisserman (1987) and Korostelev and Tsybakov (1993). There are two special features related to this problem. First, the data is two-dimensional (or multidimensional). Second, images are often composed of several regions with rather sharp edges. Within each region the image preserves a certain degree of uniformity while on the boundaries between the regions it has considerable changes. This leads to the edge estimation problem.

A large variety of methods have been proposed for solving the image and edge estimation problem in different contexts. The most popular methods of image estimation are based on the Bayesian or Markov random field approach; see Haralick (1980), Geman and Geman (1984), Ripley (1988) and Winkler (1995), among others. Nonparametric methods based on penalization and regularization have been developed in Titterington (1985), Mumford and Shah (1989) and Girard (1990).

Edge detection methods mostly do not assume any underlying parametric model. Methods based on kernel smoothing with a special choice of kernels have

Received October 1999; revised December 2001.
AMS 2000 subject classifications. Primary 62G07; secondary 62H35.
Key words and phrases. Image and edge estimation, pointwise adaptation, averaging window.

been discussed in Pratt (1978), Marr (1982), Lee (1983), Huang and Tseng (1988) and Müller and Song (1994). There are a number of proposals for nonparametric smoothing of images which allow for preserving the sharp edge structure. We mention: modal regression [see, e.g., Scott (1992)] the nonlinear Gaussian filter [see Godtliebsen, Spjøtvoll and Marron (1997)] the M-smoother of Chu, Glad, Godtliebsen and Marron (1998), the adaptive weights smoother from Polzehl and Spokoiny (2000) and different proposals based on wavelets [see, e.g., Nason and Silverman (1994), Engel (1994) or Donoho (1999) and references therein].

Tsybakov (1989) proposed a two-step procedure with the first step a preliminary image classification while the second step performs the usual kernel smoothing over the classified regions; for some extensions in this direction see also Qiu (1998). The method from Tsybakov (1989) leads to the near-optimal rate $(n^{-1}\log n)^{1/2}$ of nonparametric estimation of piecewise constant images composed from several connected regions with piecewise-Lipschitz boundaries, $n^{-1/2}$ being the grid step. Unfortunately, this method leads only to a suboptimal quality of edge estimation for the case of images with smooth edges. It also requires some prior information about the image structure, such as the number of regions in the image and image contrasts. This motivates the further study of the problem of optimal edge estimation.

A general asymptotic minimax theory of edge estimation has been developed in Korostelev and Tsybakov (1993), mostly for images composed of two homogeneous regions with a prespecified edge orientation (boundary fragment). In particular, they showed that linear methods are not optimal for images with sharp edges. Imposing some smoothness restrictions on the boundary, they found the minimax rate $n^{-\gamma/(\gamma+1)}$ of edge estimation, γ being the degree of edge smoothness, and constructed rate-optimal estimators for images with the structure of a boundary fragment. The proposed methods are essentially nonlinear and they involve a local change-point analysis as a building block. For the most interesting case when $\gamma > 1$, both the methods and results apply only under a random or "jittered" design. Barron, Birgé and Massart (1999) extended their results to edges from Sobolev-type classes, applying a general theory of adaptive estimation on sieves. Hall and Raimondo (1998) and Donoho (1999) studied the quality of edge estimation under gridded design. They showed an essential difference in studying the edge estimation problem under a random and gridded design and established the global rate of edge estimation for the boundary fragment case.

In the present paper, we also restrict ourselves to the deterministic equispaced design, focusing on the problem of image estimation at design points. Our results are stated for the case of piecewise constant images with smooth edges, and they show that the estimation quality depends strongly on the distance from the point of estimation x to the closest edge and on edge smoothness and orientation. Our method is based on direct image estimation without any preliminary edge recovering. We apply a simple linear estimator which is the average of observations over a window selected in a data-driven way. In spite of the fact that linear methods are only suboptimal in edge estimation, the results of this paper show that a nonlinearity which is incorporated in the linear method by an adaptive choice of an averaging window allows getting a near-optimal quality of image and edge recovery.

The approach presented can be viewed as one more application of the idea of pointwise adaptive estimation; see Lepski (1990, 1992), Lepski, Mammen and Spokoiny (1997), Lepski and Spokoiny (1997) and Spokoiny (1998). The first three papers mentioned consider the problem of an adaptive choice of one estimator from a family of estimators which can be ordered by their variances. A typical example is given by kernel estimators with a fixed kernel and different bandwidths. Spokoiny (1998) discussed an adaptive choice of an asymmetric averaging window for local polynomial estimation including different one-sided windows. The corresponding family is only partially ordered (i.e., there could be many estimators with the same or similar variance) and the original idea from Lepski (1990) does not apply. The main idea of the approach in Spokoiny (1998) can be expressed very easily: the procedure searches for a largest local vicinity of the point of estimation where the simple structural assumption fits well to the data.

We now apply this idea to the problem of image estimation. We focus on the case of piecewise constant images, that is, we assume that the image is composed of a finite number of regions and the image value is constant within each region. The number of regions, the difference between values of the image function f for different regions and the regularity of edges are unknown and may be different for different parts of the image. Therefore, our structural assumption is not very restrictive and allows reasonable fitting of a large class of images. Moreover, our method can be extended to estimation of any function that can be well approximated by a constant function in a local vicinity of each point.

The pointwise approach has the obvious advantage of being able to express the quality of estimation in one specific grid point depending on local image characteristics such as the distance to the closest edge, orientation and smoothness of this edge, image contrast and noise level. For the case of an image fragment, the method leads immediately to rate near-optimal estimation in any global (integral) norm as studied in Hall and Raimondo (1998) or Donoho (1999). But the inverse is not generally true: a global rate optimality does not guarantee a good local quality of estimation. A well-known example is given by the so-called Gibbs effect, which is encountered in many methods that are asymptotically optimal in a global norm. An estimator has strong fluctuations near points of discontinuities. We therefore focus on pointwise estimation quality and refer to Lepski, Mammen and Spokoiny (1997) for relations between local and global accuracy of estimation in the univariate regression.

In what follows we consider the regression model

$$Y_i = f(X_i) + \xi_i, \qquad i = 1, \dots, n, \tag{1.1}$$

where $X_i \in [0,1]^d$, $i = 1, \dots, n$, are given design points and ξ_i are individual independent random errors. Below we will suppose that ξ_i, $i = 1, \dots, n$, are i.i.d. $\mathcal{N}(0, \sigma^2)$ with a given noise level σ. The procedure and the general "oracle" result from Section 3 apply for an arbitrary fixed design. In Section 4 and our numeric examples we assume the design points need to form an equidistant grid in the square $[0,1]^2$. Next we suppose that the cube $[0,1]^d$ is split into M regions A_m, $m = 1, \dots, M$, each of them a connected set with an edge (boundary) G_m. The function f is assumed constant within each region A_m,

$$f(x) = \sum_{m=1}^{M} a_m \mathbf{1}(x \in A_m), \tag{1.2}$$

where $a_1, \dots, a_M$ are unknown constants. The problem is to estimate the image function $f(x)$ or, equivalently, to estimate the values $a_1, \dots, a_M$ and to decide for each point X_i what the corresponding region is.

The idea of the proposed method is quite simple. We search for a window U, containing x^0, of maximal size, in which the function f is well approximated by a constant. Further, this constant is taken as the resulting estimate. Of course, the choice of the class of candidate windows plays the key role for such an approach. We will discuss this problem later. We suppose for a moment that we are given a class $\mathcal{U}$ of windows U, each of them being a subset of the unit cube $[0,1]^d$ containing the point of interest x^0. By N_U we denote the number of design points in U. The assumption that f is constant in U leads to the obvious estimator $\hat{f}_U$ of $f(x^0)$ which is the mean of observations Y_i over U.

To characterize the quality of the window U we compute the residuals $\varepsilon_{U,i} = Y_i - \hat{f}_U$ and test the hypothesis that these residuals $\varepsilon_{U,i}$ can, within the window U, be treated as pure noise. Finally, the procedure selects the maximal (in number of points N_U) window for which this hypothesis is not rejected.

The paper is organized as follows. In the next section we present the procedure, and Section 3 contains the results on the quality of this procedure. In Section 4 we specify the general results to the case of $d = 2$ and discuss the problem of edge estimation in this context. Section 5 contains simulated examples and applications. Some possible extensions of the method and theory are listed in Section 6. The proofs are mostly deferred to Section 7.

2. Estimation procedure. Let data Y_i, X_i, $i = 1, \dots, n$, obey model (1.1). We will estimate $f(x^0)$ for a given x^0. Typically x^0 is a design point; that is, the image is recovered at the same point where it is observed.

Given a family of windows $\mathcal{U}$ and $U \in \mathcal{U}$, set N_U for the number of points X_i in U,

$$N_U = \#\{X_i \in U\}.$$

We will suppose that $N_U \geq 2$ for each $U \in \mathcal{U}$. We assign to each $U \in \mathcal{U}$ the estimator $\hat{f}_U$ of $f(x^0)$ by

$$\hat{f}_U = \frac{1}{N_U} \sum_U Y_i.$$

Here the sum over U means the sum over design points in U.

Our adaptation method is based on the analysis of the residuals $\varepsilon_{U,i} = Y_i - \hat{f}_U$. We introduce another family $\mathcal{V}(U)$ of windows V, each of them a subwindow of U; that is, $V \subset U$. One example for the choice of the families $\mathcal{U}$ and $\mathcal{V}(U)$ in the two-dimensional case is presented in Section 4. By C_U we denote the cardinality of $\mathcal{V}(U)$, $C_U = \#\mathcal{V}(U)$.

For each $V \in \mathcal{V}(U)$ set

$$T_{U,V} = \frac{1}{\sigma_{U,V} N_V} \sum_V \varepsilon_{U,i} = \frac{1}{\sigma_{U,V} N_V} \sum_V (Y_i - \hat{f}_U) = \frac{\hat{f}_V - \hat{f}_U}{\sigma_{U,V}},$$

where $\sum_V$ means summation over the index set $\{i : X_i \in V\}$ and $\sigma_{U,V}$ is the standard deviation of the difference $\hat{f}_V - \hat{f}_U$,

$$\sigma_{U,V}^2 = \sigma^2 \frac{N_U - N_V}{N_U N_V} = \sigma^2 (N_V^{-1} - N_U^{-1}).$$

Define now

$$\rho_{U,V} = \mathbf{1}(|T_{U,V}| > t_U), \tag{2.1}$$

where t_U is a threshold which may depend on U that determines the probability of a wrong classification.

We say that U is rejected if $\rho_{U,V} = 1$ for at least one $V \in \mathcal{V}(U)$, that is, if $\rho_U = 1$ with

$$\rho_U = \sup_{V \in \mathcal{V}(U)} \rho_{U,V} = \mathbf{1}\Big(\sup_{V \in \mathcal{V}(U)} |T_{U,V}| > t_U \Big).$$

The adaptive procedure selects among all nonrejected U's one which maximizes N_U,

$$\widehat{U} = \operatorname*{arg\,max}_{U \in \mathcal{U}} \{N_U : \rho_{U,V} = 0 \text{ for all } V \in \mathcal{V}(U)\}.$$

If there is more than one nonrejected set U attaining the maximum, then any of them can be taken. Finally we set

$$\hat{f}(x^0) = \hat{f}_{\widehat{U}}(x^0) = \hat{f}_{\widehat{U}}.$$

The algorithm involves a multiple testing procedure, and the choice of thresholds t_U is important. Particular examples of this choice will be given in

Section 4. For further exposition we keep this choice free. Our theoretical results only require the following two conditions to be fulfilled: for given $\alpha > 0$,

$$\sum_{V \in \mathcal{V}(U)} e^{-t_U^2/2} = C_U e^{-t_U^2/2} \le \alpha \qquad \forall U \in \mathcal{U}, \tag{2.2}$$

$$\sum_{U \in \mathcal{U}} e^{-t_U^2/2} \le \alpha. \tag{2.3}$$

3. One "oracle" result. Below we describe some properties of the proposed estimation procedure and state the result about the corresponding accuracy of estimation.

Let x^0 be a given point. Our target is the image value $f(x^0)$. In the sequel we assume that x^0 is from region A that coincides with some A_m for $m \le M$.

Let also a family $\mathcal{U}$ of windows containing x^0 and for each $U \in \mathcal{U}$ a family of test subwindows $\mathcal{V}(U)$ be fixed. Our result is stated using the notion of an "ideal" (or "oracle") window U^* from $\mathcal{U}$. Namely, let U^* stand for a "good" window from $\mathcal{U}$ in the sense that U^* is contained in A and it is reasonably large (in the number of design points).

If this window U^* were known from an "oracle" then one would apply the corresponding estimate $\hat{f}_{U^*}$ for recovering $f(x^0)$. Our first result claims that the accuracy of the adaptive estimator is essentially as good as the accuracy of the "oracle" estimator.

Similar results can be found in Lepski (1990) and Lepski and Spokoiny (1997) but only for the case when the considered family of estimators can be ordered by their standard deviation. For instance, Lepski and Spokoiny (1997) considered a family of kernel estimators $\tilde{f}_h(x)$ with a fixed kernel and variable bandwidth h. This family is naturally ordered by the corresponding variance $\sigma_h^2(x) = \operatorname{Var} \tilde{f}_h(x)$ (which typically decreases as bandwidth increases). Here the considered family of estimators is only partially ordered and there may exist many estimators $\hat{f}_U$ (corresponding to different windows U) with the same variance $v_U = \operatorname{Var} \hat{f}_U = \sigma^2 N_U^{-1}$.

To our knowledge, the first paper treating a pointwise adaptive estimation for a partially ordered family of estimators is Spokoiny (1998) where a univariate regression problem was considered but left- and right-sided kernels were admissible. The multivariate situation is even more complicated and requires a more careful definition of the considered set of windows $\mathcal{U}$ and $\mathcal{V}(U)$, $U \in \mathcal{U}$.

Namely, we require that the sets $\mathcal{U}$ and $\mathcal{V}(U)$ fulfill the following conditions:

(U.1) Every set U from $\mathcal{U}$ contains x^0.

(U.2) For any $U^* \in \mathcal{U}$, there is an integer number $K = K(U^*)$ such that for every $U \in \mathcal{U}$ with $N_U > N_{U^*}$, the intersection $U \cap U^*$ contains a testing window $V \in \mathcal{V}(U)$ with $N_V \ge K$.

Conditions (U.1) and (U.2) rely only on the sets $\mathcal{U}$ and $\mathcal{V}$. To state the result we need one more condition which also relies on the region A the point of estimation x^0 belongs to. Namely, it concerns windows U from $\mathcal{U}$ that are not "good" in the sense that they have a nontrivial part outside of A. We first introduce a subclass $\mathcal{U}' = \mathcal{U}'(U^*, A)$ of "bad" windows U such that $U \backslash A$ contains a "massive" testing window V from $\mathcal{V}(U)$,

$$\mathcal{U}'(U^*, A) = \{U \in \mathcal{U} : N_U > N_{U^*} \text{ and } \exists\, V \in \mathcal{V}(U) \\ \text{with } V \subset U \backslash A \text{ and } N_V \geq K\} \tag{3.1}$$

with K from condition (U.2). We will see (Lemma 7.2 below) that the procedure can select a "bad" window only with a small probability. However, selection of any "nonbad" window U is possible. Our last condition relies exactly on such windows, $U \notin \mathcal{U}'$, with $N_U > N_{U^*}$ and it requires that the intersection $U \cap A$ is "massive."

(U.3) For any $U^* \in \mathcal{U}$ and for every $U \in \mathcal{U}$ with $N_U > N_{U^*}$ and $U \notin \mathcal{U}'(U^*, A)$, there is a $V \in \mathcal{V}(U)$ such that $V \subseteq A$ and $N_V \geq \nu N_{U^*}$ with some fixed number $\nu > 0$.

Now we are in a position to state the main result. For an "ideal" window U^*, define

$$t^* = \max_{U \notin \mathcal{U}'(U^*, A)} t_U \tag{3.2}$$

with $\mathcal{U}'(U^*, A)$ from (3.1) and t_U from the definition of the procedure; see (2.1).

THEOREM 3.1. *Let the image function $f(x)$ be piecewise constant; see* (1.2). *Let U^* be an "ideal" window and conditions* (U.1) *through* (U.3) *be satisfied for this U^* with some $K > 0$ and $\nu > 0$. Let also the thresholds t_U fulfill* (2.2) *and* (2.3). *If $x^0 \in A_m$ and*

$$|a_m - a_{m'}| \geq 6\sigma t^* K^{-1/2} \qquad \forall\, m' \neq m, \tag{3.3}$$

then

$$\mathbf{P}_f\big(|\hat{f}(x^0) - f(x^0)| > 2\sigma(\nu N_{U^*})^{-1/2} t^*\big) \leq 3\alpha.$$

DISCUSSION. Here we briefly comment on the result of Theorem 3.1. Note first that the "oracle" estimator $\hat{f}_{U^*}$ has the accuracy of order $N_{U^*}^{-1/2}$. Due to the above result, the adaptive estimator $\hat{f}(x^0)$ has the same accuracy up to some fixed factor t^* provided that conditions (U.1), (U.2), (U.3) and (3.3) are fulfilled. The value t^* is typically of order $\sqrt{\log n}$; see examples in Section 4.1. This factor can be viewed as a payment for pointwise adaptation and it necessarily appears even in a simple one-dimensional situation [Lepski (1990) and Lepski and Spokoiny (1997)].

If the region A is large (i.e., A is comparable in size to the whole square) and if x^0 is an internal point of A, then typically there are large windows U with of order n points inside, that is, $N_{U^*} \asymp n$. Therefore, inside each "large" region, the proposed procedure estimates the image value with the rate $n^{-1/2}$ up to a log-factor. If x^0 lies near the boundary of the region A, then the size of U^* depends on the distance of x^0 to the boundary of A and on the smoothness properties of this boundary. The same is valid for the quality of estimation. More detailed discussion can be found in Section 4.3.

4. Two-dimensional images with gridded design. In this section we specify our procedure and results to the two-dimensional case with the regular equidistant design in the unit square $[0, 1]^2$. We also discuss the problem of edge estimation.

Suppose we are given n design points $X_1, \dots, X_n$ with $X_i = (X_{i,1}, X_{i,2}) \in [0, 1]^2$. Without loss of generality we may assume that $\sqrt{n}$ is an integer and denote $\delta = n^{-1/2}$. Now each design (or grid) point X_i can be represented in the form $X_i = (k_1\delta, k_2\delta)$ with nonnegative integers k_1, k_2.

As previously, we consider the problem of estimating the image value at a point x^0 by observations $Y_1, \dots, Y_n$ described by the model equation (1.1). In this section we restrict ourselves to estimation on the grid; that is, we suppose additionally that x^0 is a grid point.

We begin by describing one possible choice of the set of windows $\mathcal{U}$. Then we specify the result of Theorem 3.1 to this case and consider the problem of edge estimation. Finally, we discuss the accuracy of estimation near an edge as a function of the noise level, image contrast and edge orientation and state the asymptotic optimality of the proposed method.

4.1. *An example of a set of windows.* Our construction involves two external integer parameters, D and s, which control the complexity of the algorithm.

For a fixed number $d \le D$, let Q_d be the axis-parallel square with centre at x^0 and with side length $2d\delta$, δ being $n^{-1/2}$. Obviously Q_d contains exactly $N_d = (2d + 1)^2$ design points. First we describe the set $\mathcal{U}_d$ of all windows U associated with this square. In words, this set contains the whole square Q_d and all its parts defined by linear splits with different orientations. Each orientation is determined by a pair of integers p, q such that the fraction p/q is unreducible. Define

$$\mathcal{R}_s = \{(p, q) : |p| \le s, |q| \le s, p/q \text{ is unreducible}, (p, q) \ne (0, 0)\}$$

and put $r_s = \#\mathcal{R}_s$. Note that (p, q) and $(-p, -q)$ are two different orientations. In particular, $r_1 = 8$, $r_2 = 16$, $r_3 = 32$, $r_4 = 80$, $r_5 = 112$. It is obvious that $r_s \le 4s(s + 1)$.

For every $(p, q) \in \mathcal{R}_s$, we define a subset $U_{d,(p,q)}$ of Q_q by the linear split with the orientation (p, q),

$$U_{d,(p,q)} = \{x = (x_1, x_2) \in Q_q : p(x_1 - x_1^0) - q(x_2 - x_2^0) \ge -\rho_{d,(p,q)}\}, \tag{4.1}$$

where the constant $\rho_{d,(p,q)}$ is introduced to ensure condition (U.2). We define the value $\rho_{d,(p,q)}$ by the condition

$$N_U \ge N_{Q_d}/2 + K_d = d(2d+1) + K_d \tag{4.2}$$

with $U = U_{d,(p,q)}$ and $K_d = 1 + [\log(2d+1)]$. In the typical situation with $d/K_d \ge \max\{|p|, |q|\}$, there exist at least $2K_d + 1$ grid points within Q_d lying on the line $p(x_1 - x_1^0) - q(x_2 - x_2^0) = 0$ and thus $\rho_{d,(p,q)} = 0$. In that case, the set $U_{d,(p,q)}$ is defined by the linear split with the orientation (p, q) passing through x^0. A general bound for $\rho_{d,(p,q)}$ is given in Lemma 7.6 below.

Let $\mathcal{U}_d$ be the set of all such windows $U_{d,(p,q)}$ plus the whole square Q_d,

$$\mathcal{U}_d = Q_d + \{U_{d,(p,q)} : (p,q) \in \mathcal{R}_s\}.$$

We identify here every two windows which contain the same collection of grid points. It is obvious that every $\mathcal{U}_d$ contains at most $r_d + 1$ different windows, r_d being $\#\mathcal{R}_d$. Moreover, it is easy to check that $\#\mathcal{U}_d = r_d + 1$ if $s > 2d$. The set of all candidate windows U is composed of all $\mathcal{U}_d$ for $d \le D$,

$$\mathcal{U} = \mathcal{U}_0 \cup \mathcal{U}_1 \cup \cdots \cup \mathcal{U}_D.$$

This construction can be viewed as a local version of the wedgelets proposed by Donoho (1999).

Let also $\mathcal{V}_d$ be the set of intersections of two windows from $\mathcal{U}_d$,

$$\mathcal{V}_d = \{U \cap U' : U, U' \in \mathcal{U}_d\}.$$

Now, for U from $\mathcal{U}_d$ we define a family of testing windows $\mathcal{V}(U)$ by taking all windows V from $\mathcal{V}_{d'}$ with $d' \le d$ and $V \subset U$,

$$\mathcal{V}(U) = \bigcup_{d'=0}^{d} \{V \in \mathcal{V}_{d'} : V \subset U\}.$$

For the above defined set $\mathcal{U}$, condition (U.1) is fulfilled by construction. Let U^* be a window from $\mathcal{U}_d$. One can easily check that condition (U.2) for this U^* and $K = K_d$ is fulfilled as well.

It is obvious that $\#\mathcal{V}_d \le r_d(r_d+1)/2$ and hence, the total number C_U of windows in $\mathcal{V}(U)$ for $U \in \mathcal{U}_d$ is bounded by $d r_d(r_d+1)/2$.

We finally define

$$t_U = t_d = \sqrt{2\lambda + 2\mu \log(d+1)}, \qquad U \in \mathcal{U}_d, \tag{4.3}$$

with some fixed positive constants λ, μ. The bound $r_d \le 4d(d+1)$ easily yields

$$C_U e^{-t_U^2/2} \le d r_d (r_d+1) e^{-\lambda}(d+1)^{-\mu} \le e^{-\lambda} \qquad \forall U \in \mathcal{U}_d,$$

$$\sum_{U \in \mathcal{U}} e^{-t_U^2/2} \le \sum_{d=1}^{D} (r_d+1) e^{-t_d^2/2} \le e^{-\lambda},$$

if, for example, $\mu \ge 5$, so that the thresholds (4.3) with properly selected λ, μ fulfill conditions (2.2) and (2.3).

Further we discuss the properties of the estimate $\hat{f}(x^0)$ corresponding to the previously described sets $\mathcal{U}$ and $\mathcal{V}(U)$, $U \in \mathcal{U}$ and the thresholds t_U from (4.3).

4.2. *Accuracy of estimation inside a homogeneous region.* We begin with the very simple situation when the point of interest x^0 lies inside a homogeneous region $A = A_m$ for some $m \le M$. We will see that in such a case the value $f(x^0)$ is estimated at the rate $n^{-1/2}$ up to a logarithmic factor.

THEOREM 4.1. *Let the point x^0 belong to a homogeneous region $A = A_m$ together with the square $\{x : \|x - x^0\|_\infty \le \varepsilon\}$ with some $\varepsilon > \delta$. Let also the image contrast $b = \max\{|a_m - a_{m'}|, m' \ne m\}$ satisfy the condition*

$$b \ge 4t^*\sigma/\sqrt{K^*} \tag{4.4}$$

with t^* from (3.2) *and $K^* = \max_{d \le D} K_d = K_D$. If $D \ge n^{1/2}\varepsilon$, then*

$$\mathbf{P}_f\big(|\hat{f}(x^0) - f(x^0)| \ge 2\sigma t^*(n^{1/2}\varepsilon)^{-1}\big) \le 3e^{-\lambda}.$$

This result is a straightforward corollary of Theorem 3.1. It suffices to note that the window U^* coinciding with the square Q_D belongs to the family $\mathcal{U}$ and it is contained in A. Hence, $N_{U^*} \ge 4D^2$. The condition (U.3) is fulfilled in this situation with $\nu = 0.5$.

4.3. *Accuracy of estimation near an edge.* Now we apply Theorem 3.1 to the case when the point of interest x^0 lies near an edge of the corresponding region. We first illustrate the importance of a careful estimation near an edge by the following example.

EXAMPLE 4.1. Let A be a circle inside the unit square with radius $r > 0$. We do not suppose that the center of this circle is at a grid point. The radius r may be also arbitrary. We set $\rho = C/n$ with some constant $C > 1$ and consider a band of width ρ near the edge of A. Note that this width is essentially smaller than the grid step $\delta = n^{-1/2}$, if C is not too large. The Lebesgue measure of this band is about $2\pi r\rho$, so, for the uniform random design, the mean number of design points inside this band would be about $2\pi r\rho n = 2\pi rC$. It can be shown by using the arguments from the theory of continuous fractions, see Hall and Raimondo (1998) or Lemma 7.5 below, that under the equidistant design, we have essentially the same (in order) number of design points inside this band. This is illustrated in the left of Figure 1. On the other side, it is well known that the quality of estimation near an edge is especially important by visualization. Even single errors in image segmentation are visible and they lead to a significant deterioration of the image. Therefore, a desirable property of the estimation procedure would be to recover precisely the image value for all points separated from the edge with a distance of a smaller order than $n^{-1/2}$.

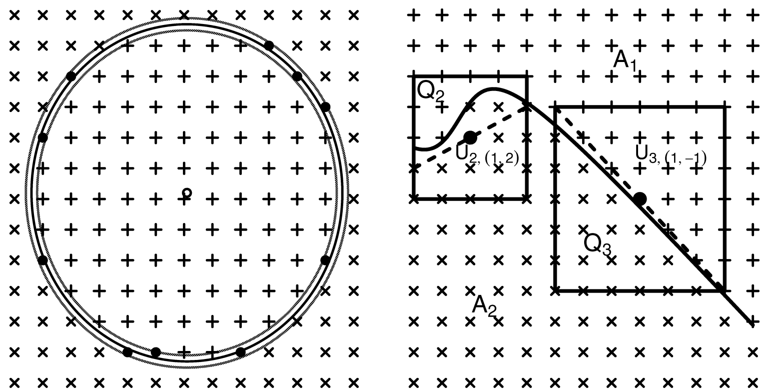

FIG. 1. *Left*: *Band of width* $\rho = 0.4\delta$ *around a dislocated circle of radius* 5.5δ. *The band contains* 10 *points* ($2\pi r\rho n \approx 13.8$). *Right*: *Sets* Q_d *and* $U_{d,(p,q)}$ *for two points* x^0.

Let x^0 belong to a region A_m and lie near the edge G with another region $A_{m'}$. We assume also that this edge is regular in the sense that it can be well approximated by a straight line in some small vicinity of the point x^0.

Without loss of generality we may assume that the edge G can be parametrized in a neighborhood of the point x^0 by the equation $x_2 = g(x_1)$ with some differentiable function g and that $|g'(x_1^0)| \le 1$. [Otherwise another parametrization of the form $x_1 = g(x_2)$ is to be used.] Now the image function f can, at least in a neighborhood of the point x^0, be represented in the form

$$f(x) = \begin{cases} a_{m'}, & x \in A_{m'} = \{x_2 > g(x_1)\}, \\ a_m, & x \in A_m = \{x_2 \le g(x_1)\}. \end{cases} \tag{4.5}$$

The distance from x^0 to the edge G of A_m can be characterized by the value $g(x_1^0) - x_2^0$.

In the next result we suppose that the edge function g is smooth in the sense that it belongs to the Hölder class $\Sigma(\gamma, P)$ with some parameters $\gamma \in (1, 2]$ and $P > 0$. This means that g satisfies the condition

$$|g'(s) - g'(t)| \le P|s - t|^{\gamma} \qquad \forall s, t.$$

This setup is essentially as in Korostelev and Tsybakov (1993) with the only difference in the assumption of a gridded design and of $\gamma \le 2$. Korostelev and Tsybakov (1993) proposed a polynomial approximation of the edge within a vertical strip of the width of order $h^* = (nP)^{-1/(\gamma+1)}$ around the point of estimation which leads to the optimal estimation rate $n^{-\gamma/(\gamma+1)}P^{1/(\gamma+1)}$. It follows from the next theorem that the estimate $\hat{f}(x^0)$ delivers essentially the same quality of estimation. Moreover, our method for this special setup can also be reduced to local approximation of the boundary by a linear function.

We first describe the "ideal" window for this situation. Define d^* as the minimal value d satisfying $d \ge n^{(\gamma-1)/2(\gamma+1)}(4K^*/P)^{1/(\gamma+1)}$. For ease of exposition we assume an equality in this relation for $d = d^*$, that is,

$$d^* = n^{(\gamma-1)/2(\gamma+1)}(4K^*/P)^{1/(\gamma+1)}. \tag{4.6}$$

The square Q_{d^*} has the side length $d^*\delta$ which is of order h^*. Let $z = g'(x_1^0)$ be the slope of the edge near the point of estimation, and let (p, q) be a pair from $\mathcal{R}_{d^*}$ minimizing the value $|z - p/q|$ over $\mathcal{R}_{d^*}$. The next result shows that if the distance $g(x_1^0) - x_2^0$ is sufficiently large (of order $n^{-\gamma/(\gamma+1)}$) then the smoothness of the edge ensures that the window $U^* = U_{d^*,(p,q)}$ belongs to the region A.

THEOREM 4.2. *Let the image function $f(x)$ be of the form* (4.5) *in a neighborhood $Q_\varepsilon(x^0)$ of the point x^0 with some positive $\varepsilon > 0$ and let a grid point x^0 belong to A. The function g describing the edge G is supposed to be in the Hölder class $\Sigma(\gamma, P)$ with $\gamma \in (1, 2]$. Let the image contrast b satisfy condition* (4.4). *Let also the parameters D, s of the adaptive procedure satisfy $D \ge d^*$ and $s > 2d^*$ with d^* from* (4.6). *If $d^* \le \varepsilon n^{1/2}$ and if the distance $g(x_1^0) - x_2^0$ satisfies*

$$g(x_1^0) - x_2^0 \ge 2P^{1/(\gamma+1)}(4K^*/n)^{\gamma/(\gamma+1)},$$

then

$$\mathbf{P}_f\big(|\hat{f}(x^0) - f(x^0)| \ge 2\sigma t^*/d^*\big) \le 3e^{-\lambda}.$$

DISCUSSION. The statement of Theorem 4.2 is essentially nonasymptotic, that is, it applies for a fixed image f and a fixed n. However, this result delivers some clear information about dependence of estimation accuracy on n. For our construction, it obviously holds that $t^* \le \max_U t_U \le C\sqrt{\log n}$ and condition (4.4) on K^* implies $K^* \ge (4t^*\sigma/b)^2 = C\sigma^2 b^{-2}\log n$ for some fixed constant C. Therefore, if we define the class of images which satisfy the conditions of Theorem 4.2 [i.e., the class of images with the structure of a boundary fragment in a vicinity of the point x^0 and with the edge function g from $\Sigma(\gamma, P)$] then the proposed procedure provides a reasonable quality of estimation uniformly over this class for all points x^0 which are separated away from the edge with the distance of order $\psi_\gamma(n) = (n/\log n)^{-\gamma/(\gamma+1)}$.

Points lying on the edge or very close to the edge can be misclassified by the procedure. The width of the "band of insensitivity" around the edge, that is, the minimal distance between the point x^0 and the edge G which is sufficient for estimation of $f(x^0)$ with a small estimation error can be regarded as the accuracy of edge estimation. Korostelev and Tsybakov (1993) described the quality of recovering the edge function g which belongs to a Hölder class $\Sigma(\gamma, 1)$, and showed that the optimal rate of edge estimation, being measured in the Hausdorff metric, is $(n/\log n)^{-\gamma/(\gamma+1)}$ which formally coincides in order with the accuracy delivered by our procedure. Note meanwhile, that Korostelev and Tsybakov (1993) stated their results for $\gamma > 1$ only under a random or "jittered" design; see page 92 there. Under the regular (gridded) design, the rate of edge estimation is equal to the grid step $\delta = n^{-1/2}$ [Korostelev and Tsybakov (1993), page 99]. We proceed under the equispaced (gridded) design focusing on estimating the value of the image at a grid point. The gridded structure of the design is essential for our results and the proofs are based on number-theoretical arguments similar to those in Hall and Raimondo (1998).

The result of Theorem 4.2 delivers some additional information about dependence of the quality of estimation near an edge on the noise level σ, the image contrast b and the orientation of the edge G described by the value $z = g'(x_1^0)$.

4.3.1. *Accuracy versus noise level and image contrast.* It follows from Theorem 4.2 that the image value $f(x^0)$ can be recovered with a sufficient precision if the point of interest x^0 is separated from the edge G with the distance of order $(n^{-1}K^* P)^{-\gamma/(\gamma+1)}$. This expression depends on the noise level σ only through K^* which must satisfy $K^* \geq C\sigma^2 b^{-2} \log n$ with some constant C; see (4.4). We see that when the noise level increases, the quality of edge recognition decreases by the factor $\sigma^{2\gamma/(\gamma+1)}$.

All this remains valid for dependence of the quality of estimation on the value of image contrast $b = \max\{|a_m - a_{m'}|,\ m' \neq m\}$. The only difference is that this dependence is with another sign; when the contrast increases the quality increases as well, and vice versa. Both these issues are in accordance with the one-dimensional case [Spokoiny (1998)] and with similar results for a random design [Mammen and Tsybakov (1995)].

4.3.2. *Accuracy versus edge orientation.* We now discuss a problem which appears only for the regular design. The issue is dependence of the quality of edge estimation on the edge orientation. This orientation is characterized by the edge slope $z = g'(x_1^0)$. By inspecting the proof of Theorem 4.2 one can see that the quality of estimation depends critically on the quality of approximation of z by rational numbers with bounded denominators. It follows from the result that the worst case edge orientation leads just to the above indicated quality of estimation near an edge. At the same time, if z is a rational number, $z = p/q$, with a bounded q, or if z is very close to such a rational number, then a stronger result can be stated. Similar assertions can be found in Hall and Raimondo (1998) and Donoho (1999).

THEOREM 4.3. *Let the image function $f(x)$ be of the form* (4.5) *in a neighborhood $\{x : \|x - x^0\|_\infty \leq \varepsilon\}$ of the point x^0 with some positive $\varepsilon > 0$ and let the grid point x^0 belong to A. The function g describing the edge G is supposed to be in the Hölder class $\Sigma(\gamma, P)$ with $\gamma \in (1, 2]$ and additionally $z := g'(x_1^0)$ satisfies*

$$|z - p/q| \leq \frac{1}{q(q+1)}$$

for some integer numbers $p \le q$. *Let condition* (4.4) *hold true. If* $D \ge qK_q$ *and also* $\varepsilon n^{1/2} \ge qK_q$, *and if the distance* $g(x_1^0) - x_2^0$ *satisfies*

$$g(x_1^0) - x_2^0 \ge 2P(qK_q n^{-1/2})^\gamma, \tag{4.7}$$

then

$$\mathbf{P}_f\big(|\hat{f}(x^0) - f(x^0)| \ge 2\sigma t^*(qK_q)^{-1}\big) \le 3e^{-\lambda}.$$

As a corollary of this result we conclude that for an edge with a rational (e.g., with horizontal or vertical) orientation, the band of insensitivity is of order $(n^{-1}\log n)^{\gamma/2}$ which approaches n^{-1} for $\gamma = 2$.

4.3.3. *Rate optimality.* The next natural question is about the optimality of the previous results on estimation near an edge. The next assertion shows that the accuracy $\psi_\gamma(n) = (n/\log n)^{-\gamma/(\gamma+1)}$ cannot be essentially improved uniformly over the class of all boundary fragments.

From Theorem 4.3 we know that some improvement in the accuracy of estimation near an edge is still possible for images with a special edge orientation. We will see that the accuracy delivered by our procedure is at least near optimal in this situation too.

Let some grid point x^0 be fixed and let the underlying image have the structure of a smooth boundary fragment with an edge G determined by a function $g = g(x_1)$ from the Hölder ball $\Sigma(\gamma, 1)$ with $\gamma \in (1, 2]$. The function g determines the image function f_g with $f_g(x) = \mathbf{1}(x_2 \ge g(x_1))$ for $x = (x_1, x_2)$. We use $G = G_g$ for the corresponding edge, that is, $G = \{x : x_2 = g(x_1)\}$.

We are interested in the minimal distance between the point x^0 and the edge G which allows for a sufficiently precise estimation of $f(x^0)$ if the image function f is of the form f_g with g from $\Sigma(\gamma, 1)$.

THEOREM 4.4. *Let* K, D *be integers and let* $z = p/q$ *be an unreducible rational number with* $0 \le p \le q$. *Let then* ψ_n *stand for*

$$\psi_n = \min\{n^{-1/2}q^{-1}, (qKn^{-1/2})^\gamma\}.$$

Then there exists a constant $\kappa > 0$ *depending only on* γ *and two functions* g_0 *and* g_1 *from* $\Sigma(\gamma, 1)$ *such that* $g_0'(x_1^0) = z$, $g_1'(x_1^0) = z$,

$$g_0(x_1^0) \ge x_2^0 + \kappa\psi_n, \qquad g_1(x_1^0) \le x_2^0 - \kappa\psi_n, \tag{4.8}$$

and such that for any estimator $\tilde{f}$,

$$\mathbf{P}_{g_0}\big(|\tilde{f}(x^0) - f_{g_0}(x^0)| > 1/2\big) + \mathbf{P}_{g_1}\big(|\tilde{f}(x^0) - f_{g_1}(x^0)| > 1/2\big) \ge c, \tag{4.9}$$

where c *is some positive number depending only on* K.

REMARK 4.1. If we apply this theorem with a small q, then we get the lower bound for the result of Theorem 4.3. Maximizing ψ_n with respect to q leads to the choice $q \approx K^{-\gamma/(\gamma+1)}n^{(\gamma-1)/(2\gamma+2)}$ and to the lower bound $\psi_n \approx (K/n)^{\gamma/(\gamma+1)}$ coinciding in order with the upper bound from Theorem 4.2.

5. Implementation, simulation results and applications. This section discusses some numerical properties of the presented method.

5.1. *Algorithmic complexity.* Here we give some hints about the implementation of the proposed method and present an upper bound for its numerical complexity.

Since the method is pointwise, the whole procedure is to be repeated at every design point. Hence, the whole complexity is roughly n times the complexity of the basic one-point procedure. The basic procedure requires one to compute the mean of observations Y_i over every window $U \in \mathcal{U}_d$ and $V \in \mathcal{V}_d$, $d \le D$. To reduce the numerical effort, it is reasonable to perform the following preprocessing step. First, for each orientation (p, q) from $\mathcal{R}_s$, define $\phi = \phi_{p,q} = (-p, q)^\top$ and

$$e_\phi = \begin{cases} (\operatorname{sign} q, 0)^\top, & \text{for } |p| \le |q|, \\ (0, -\operatorname{sign} p)^\top, & \text{otherwise.} \end{cases}$$

Next, compute a family of two-dimensional "distribution" functions $F_\phi(t, \rho)$ and $N_\phi(t, \rho)$ for $\phi = \phi_{p,q}$, $(p, q) \in \mathcal{R}_s$, with

$$N_\phi(t, \rho) = \sum_{i=1}^{n} \mathbf{1}(e_\phi^\top X_i \le t, \phi^\top X_i \le \rho),$$

$$F_\phi(t, \rho) = \sum_{i=1}^{n} \mathbf{1}(e_\phi^\top X_i \le t, \phi^\top X_i \le \rho) Y_i.$$

It is easy to see that every such function can be computed by $O(n)$ operations, that is, the preprocessing step requires $O(r_s n)$ operations. As soon as all these functions have been computed, for every central point x^0, the mean of observations over every $U = U_{d,(p,q)} \in \mathcal{U}_d$ can be found by a finite number of arithmetic operations: with $\phi = \phi_{p,q}$,

$$N_U = N_\phi(e_\phi^\top x^0 + d\delta, \phi^\top x^0) - N_\phi\big(e_\phi^\top x^0 - (d+1)\delta, \phi^\top x^0\big) - S_{e_\phi},$$

where $S_{e_\phi} = N_{-e_\phi}(e_{-e_\phi}^\top x^0 - (d+1)\delta, e_\phi^\top x^0 + d\delta) - N_{-e_\phi}(e_{-e_\phi}^\top x^0 - (d+1)\delta, e_\phi^\top x^0 - (d+1)\delta)$. The formulas for the sums $\sum_U Y_i$ can be obtained in the same manner by replacing $N_{\cdot,\cdot}^{(\ell)}$ with $F_{\cdot,\cdot}^{(\ell)}$.

Next, for every fixed $U \in \mathcal{U}_d$, one can compute the mean of observations over every subwindow $V = U \cap U'$ with $U' \in \mathcal{U}_d$. Again, each subwindow requires only a few arithmetic operations. Since we have at most $D(r_s + 1)$ windows in $\mathcal{U}$ and $Dr_s(r_s+1)/2$ testing windows in all the $\mathcal{V}_d$'s, the complexity of the procedure at one point is $O(Dr_s^2)$ leading to the complexity $O(nDr_s^2)$ for the whole method.

5.2. *Parameter specification.* The construction of the sets $\mathcal{U}$ and $\mathcal{V}(U)$ depends upon two parameters, s and D. These parameters are introduced only to control the complexity of the method and should be sufficiently large while still providing computational feasibility. Our experience indicates that the procedure delivers a reasonable quality of restoration with $s \geq 3$ and $D \geq 8$.

The adaptive procedure (more precisely, the test of homogeneity for a particular $V \subset U$) relies on the thresholds t_U [see (2.1)] which are defined in (4.3) using two further parameters, λ, μ. These parameters are similar to usual wavelet thresholds. In particular, the choice of large λ or μ would lead to oversmoothing of the image and a small value of $\lambda + \mu$ results in keeping too much noise in the restored image. Our theoretical results only require that λ, μ fulfill the conditions (2.2) and (2.3). For a practical data-driven choice some cross-validation technique can be applied. Our experience with different artificial images indicates that the rule $\lambda + \mu \log(D + 1) \approx 2.5$ leads to a reasonable quality of estimation in almost all cases with μ going to zero as the signal-to-noise ratio decreases.

5.3. *Simulated results.* The simulation results presented in this section are based on an implementation of the above proposal. The artificial image shown on the left side of Figure 2 possesses different image contrasts and varying smoothness of edges. Using this image of 40×80 pixels we conducted a small simulation study of size $n_{\text{sim}} = 250$. Distorted images are generated by adding i.i.d. standard white noise. This provides us with signal-to-noise ratios of 1, 2, 3 and 4 along the different edges. Smoothness of the boundaries varies along the boundary of the ellipse in the center of the image.

We use the parameters $s = 3$ and $D = 8$. This means that we consider 32 possible orientations and the largest considered windows for each point contain $17 \times 17 = 289$ design points. The thresholds t_U are taken due to (4.3) with $\lambda = 2.5$ and $\mu = 0$. The central image in Figure 2 displays the pointwise mean absolute error (MAE) estimated from the 250 simulations. The right image illustrates the mean number of points in the sets $\widehat{U}(x)$ for each pixel x.

The simulation results are clearly in agreement with the theory provided in Sections 3 and 4. Absolute errors decrease and the number of points in the selected window increases with distance to the boundary. The quality of edge estimation improves with increasing contrast. The absolute error of the estimate essentially depends on the size of the ideal window U^* and therefore on the smoothness properties of the boundary. Note that large errors only occur in boundary points inside the ellipse, in points where the boundary is not smooth (corners) and at boundaries where the image contrast is small. Errors for the first two types of points are due to high variability because of insufficient points within the ideal window U^* while the small image contrast leads to poor edge estimation, that is, biased estimates, in the third case.

We have conducted the same simulation study with different settings of λ and μ, finding very similar behavior. Keeping $\lambda + \mu \log(D)$ constant we observe slightly

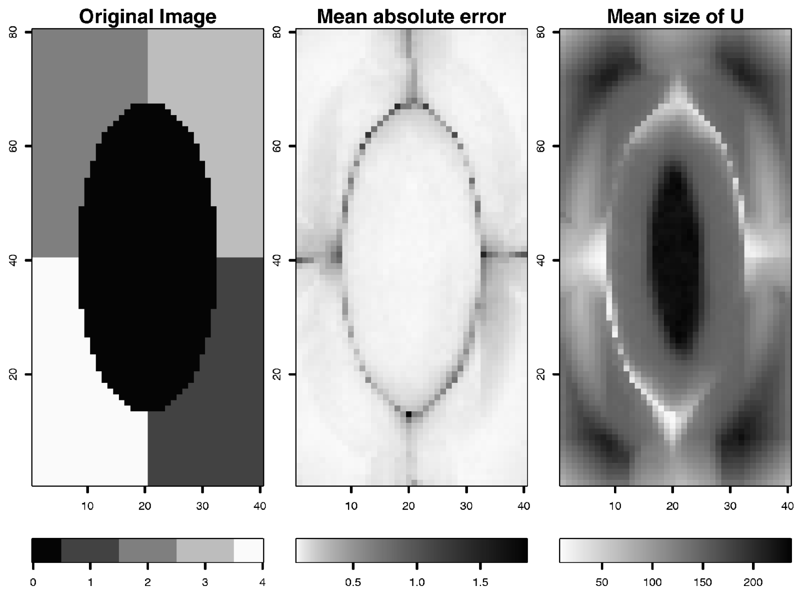

FIG. 2. *Simulation results: artificial image* (*left*), *pointwise mean absolute error* (*center*) *and mean size of selected sets* $\widehat{U}$ (*right*). *Obtained from* 250 *simulations with* $D = 8$, $s = 3$, $\lambda = 2.5$ *and* $\mu = 0$.

better results near the boundaries for small λ and very marginal losses within large regions.

We briefly illustrate the feasibility of the approach using a real life example. Figure 3 (left) shows a log-transformed C-band, HH-polarization, synthetic aperture radar (SAR) image of 250×250 pixels recorded by Dr. E. Attema

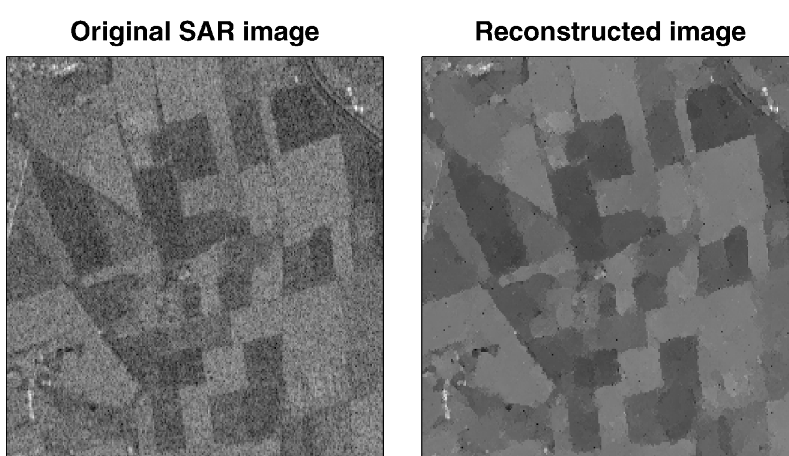

FIG. 3. *Synthetic aperture radar* (*SAR*) *image* (*left*) *and estimate with* $D = 10$, $s = 5$, $\lambda = 2.5$ *and* $\mu = 0$ (*right*).

at the European Space Research and Technology Centre in Noordwijk, Netherlands. This example is also used in Glasbey and Horgan (1995). The image shows an area near Thetford forest, England. The data can be obtained from ftp://peipa.essex.ac.uk/ipa/pix/books/glasbey-horgan/.

The reconstruction generated with parameters $D = 10$, $s = 5$, $\lambda = 2.5$ and $\mu = 0$ is given on the right side of Figure 3. The variance estimate $\hat{\sigma}^2 = 230$ has been calculated from a residual image.

All structures visible in the original image, with the exception of thin objects in the upper right, are maintained in the reconstruction. At the same time, efficient noise reduction is obtained for large homogeneous regions. This clearly shows the adaptiveness of our procedure.

6. Conclusions and outlook. The present paper offers a new approach to image denoising based on the idea of pointwise adaptation. The proposed procedure allows us to estimate a wide class of images. It is fully adaptive; that is, no prior information about image structure is required to be specified. The results claim near-optimal accuracy of the method in the sense of estimation near an edge and inside a large homogeneous region. Reconstruction results for artificial and real images are in agreement with the theory and illustrate the applicability of the method.

Below we list some possibilities to extend the method.

Piecewise smooth images and relations with two-dimensional smoothing. The method we discussed is oriented toward (nearly) piecewise constant images. This assumption can be restrictive for some applications. Similarly to Spokoiny (1998), one can handle piecewise smooth images using a local linear (or polynomial) approximation and show that the resulting procedure provides both a spatially adaptive estimation of the image function and rate optimal edge recovering.

"*Thin*" *objects and nonsmooth edges.* Another important structural assumption for our method and results is that the image is composed of large regions with smooth edges. It can be seen from our simulations and applications to real data that the method leads to oversmoothing shape corners on edges and "thin" objects like lines. Note, however, that the general approach is very flexible and the procedure (more precisely, the set of windows) can be adjusted to detecting any specific structure in the image like "thin" objects or breakpoints of boundary curves. The related construction should be, of course, more involved.

Non-Gaussian errors. Our model equation (1.1) assumes Gaussian errors ε_i. Since grey level images are usually coded by integers within a specified range, for example, between 1 and 256, this assumption can be satisfied only approximately. However, the assumption of Gaussian errors is reasonable in many real applications, and it is usually confirmed by any model check. Our theoretical results can easily be extended to sub-Gaussian errors satisfying some moment conditions.

Edge estimation and image classification. In many practical applications, image denoising is used only as a preliminary step for further image analysis, for example, classification or pattern recognition. This would require edge estimation and image classification. Our approach can naturally be used for those purposes, simply basing image classifications on the estimated image function. We consider this issue to be an important topic for further research.

Relations to adaptive weights smoothing. Polzehl and Spokoiny (2000) propose another method for image denoising called adaptive weights smoothing (AWS). The underlying idea of AWS can be viewed as a generalization of the pointwise adaptive method. For every point X_i the corresponding local model is described by means of weights ω_{ij} assigned to every point X_j. The pointwise adaptive choice of a window U corresponds to the special case with $\omega_{ij} = \mathbf{1}(X_j \in U)$. The value $f(X_i)$ is estimated from the local model by weighted least squares

$$\hat{f}(X_i) = \arg\inf_{\theta} \sum_j \omega_{ij} |Y_j - \theta|^2 = \frac{\sum_j \omega_{ij} Y_j}{\sum_j \omega_{ij}}. \tag{6.1}$$

The weights ω_{ij} are computed iteratively using the idea of structural adaptation. The AWS procedure provides excellent numerical results. The theoretical properties of AWS are, because of its iterative nature, very difficult to study. However one can expect that both methods possess similar optimality properties.

7. Proofs. In this section we present the proofs of Theorems 3.1–4.4.

7.1. *Proof of Theorem* 3.1. We begin with some preliminary results. The idea of the proposed procedure is to select adaptively the largest window among the considered class $\mathcal{U}$ which is contained in A. A necessary property of every such procedure is to accept each window contained in A with a high probability. Our first result shows that the previously described procedure possesses this property.

LEMMA 7.1. *Let $x^0 \in A$ and let $U \in \mathcal{U}$ be such that $U \subseteq A$. If* (2.2) *holds then*

$$\mathbf{P}_f(\rho_U = 1) \le \alpha.$$

PROOF. Let some U with the property $U \subseteq A$ be fixed and let $V \in \mathcal{V}(U)$. The function f is constant on U and hence on V. Using the model equation (1.1) we obtain

$$T_{U,V} = \sigma_{U,V}^{-1} \left(\frac{1}{N_V} \sum_V \xi_i - \frac{1}{N_U} \sum_U \xi_i \right).$$

Obviously we have $\mathbf{E}_f T_{U,V} = 0$. Recall now that the factor $\sigma_{U,V}$ was defined as the standard deviation of the stochastic term of the difference $\hat{f}_V - \hat{f}_U$. Hence $\mathbf{E}_f T_{U,V}^2 = 1$. Since $T_{U,V}$ is a linear combination of Gaussian variables ξ_i, $T_{U,V}$ itself is Gaussian with zero mean and the unit variance, that is, standard normal.

Therefore, $\mathbf{P}_f(|T_{U,V}| > t_U) \le \exp\{-t_U^2/2\}$. This and condition (2.2) allow bounding the probability of rejecting U in the following way:

$$\mathbf{P}_f(\rho_U = 1) \le \sum_{V \in \mathcal{V}(U)} \mathbf{P}_f(|T_{U,V}| > t_U) \le C_U e^{-t_U^2/2} \le \alpha. \qquad \square$$

The next statement can be viewed as a complement to Lemma 7.1. We now consider the case of a "bad" window containing two nonintersecting subwindows V_1 and V_2 with different values of the image function f. The result says that such a window will be rejected with a high probability.

LEMMA 7.2. *Let $U \in \mathcal{U}$ and let $V_1, V_2 \in \mathcal{V}(U)$ be such that the function f is constant within each V_j,*

$$f(x) = a_j, \qquad x \in V_j, j = 1, 2.$$

Denote $s_{V_1,V_2} = \sigma\sqrt{N_{V_1}^{-1} + N_{V_2}^{-1}}$. If

$$|a_1 - a_2| \ge (\sigma_{U,V_1} + \sigma_{U,V_2} + s_{V_1,V_2}) t_U \tag{7.1}$$

with $\sigma_{U,V}^2 = \sigma^2(N_V^{-1} - N_U^{-1})$, then

$$\mathbf{P}_f(\rho_U = 0) \le e^{-t_U^2/2}.$$

REMARK 7.1. In view of the trivial inequalities $\sigma_{U,V} \le \sigma N_V^{-1/2}$ and $\sqrt{N_{V_1}^{-1} + N_{V_2}^{-1}} \le N_{V_1}^{-1/2} + N_{V_2}^{-1/2}$, condition (7.1) holds if

$$|a_1 - a_2| \ge 2\sigma t_U \big(N_{V_1}^{-1/2} + N_{V_2}^{-1/2}\big). \tag{7.2}$$

PROOF OF LEMMA 7.2. By definition

$$\mathbf{P}_f(\rho_U = 0) \le \mathbf{P}_f(\rho_{U,V_1} = \rho_{U,V_2} = 0).$$

Next, the event $\{\rho_{U,V} = 0\}$ means $|T_{U,V}| \le t_U$ or equivalently $|\hat{f}_U - \hat{f}_V| \le \sigma_{U,V} t_U$. This yields

$$|\hat{f}_{V_1} - \hat{f}_{V_2}| \le (\sigma_{U,V_1} + \sigma_{U,V_2}) t_U.$$

Now using the fact that $V_1 \cap V_2 = \varnothing$, we get the following decomposition [cf. the proof of Lemma 7.1]:

$$\hat{f}_{V_1} - \hat{f}_{V_2} = a_1 - a_2 + N_{V_1}^{-1} \sum_{V_1} \xi_i - N_{V_2}^{-1} \sum_{V_2} \xi_i = a_1 - a_2 + s_{1,2}\zeta_{1,2},$$

where $\zeta_{1,2}$ is a standard normal random variable. Therefore,

$$\begin{aligned}\mathbf{P}_f(\rho_U = 0) &\le \mathbf{P}\bigl(|a_1 - a_2 + s_{1,2}\zeta_{1,2}| \le (\sigma_{U,V_1} + \sigma_{U,V_2})t_U\bigr)\\ &\le \mathbf{P}\bigl(s_{1,2}|\zeta_{1,2}| \ge |a_1 - a_2| - (\sigma_{U,V_1} + \sigma_{U,V_2})t_U\bigr).\end{aligned}$$

Using the condition of the lemma, we obtain

$$\mathbf{P}_f(\rho_U = 0) \le \mathbf{P}(|\zeta_{1,2}| > t_U) \le e^{-t_U^2/2}$$

as required. □

We need one more technical result for the situation when a window U from $\mathcal{U}$ is not entirely contained in A but there is its subwindow V which is in A.

LEMMA 7.3. *Let $x^0 \in A$, $U \in \mathcal{U}$ and let V from $\mathcal{V}(U)$ be such that $V \subseteq A$. If $\rho_{U,V} = 0$, then the difference $|\hat{f}_U - f(x^0)|$ can be estimated in the following way: for any $z \ge 1$,*

$$\mathbf{P}\bigl(|\hat{f}_U - f(x^0)| > \sigma N_V^{-1/2}(z + t_U), \rho_{U,V} = 0\bigr) \le \exp\{-z^2/2\}.$$

PROOF. Let a be the image grey level within A. The event $\{\rho_{U,V} = 0\}$ means that $|\hat{f}_U - \hat{f}_V| \le \sigma_{U,V} t_U$. Therefore,

$$|\hat{f}_U - a| \le |\hat{f}_U - \hat{f}_V| + |\hat{f}_V - a| \le \sigma_{U,V} t_U + |\hat{f}_V - a|.$$

Next, $\sigma_{U,V} \le \sigma N_V^{-1/2}$ and, since $V \subset A$, $\zeta = \sigma^{-1} N_V^{1/2}(\hat{f}_V - a)$ is a standard Gaussian random variable; see the proof of Lemma 7.1. This gives

$$\mathbf{P}_f\bigl(|\hat{f}_U - f(x^0)| > \sigma N_V^{-1/2}(z + t_U), \rho_{U,V} = 0\bigr) \le \mathbf{P}(|\zeta| > z) \le \exp\{-z^2/2\}$$

as required. □

Now we turn directly to the proof of Theorem 3.1. First of all, since U^* is contained in A, due to Lemma 7.1 the window U^* will be rejected only with a very small probability, namely $\mathbf{P}_f(\rho_{U^*} = 1) \le \alpha$. This obviously implies

$$\mathbf{P}_f\bigl(\bigl|\hat{f}(x^0) - f(x^0)\bigr| > z, \rho_{U^*} = 1\bigr) \le \mathbf{P}_f(\rho_{U^*} = 1) \le \alpha$$

and it suffices to consider only the situation when U^* is accepted, that is, $\rho_{U^*} = 0$.

Let the window $\widehat{U}$ be selected by the procedure. Then $\rho_{\widehat{U}} = 0$ and, since $\rho_{U^*} = 0$, by definition of $\widehat{U}$,

$$N_{\widehat{U}} \ge N_{U^*}.$$

Next, due to condition (U.2), there is a subwindow V in $\widehat{U} \cap U^*$ with at least K design points which is contained in A.

Let $\mathcal{U}' = \mathcal{U}'(U^*, A)$ be the class of all "bad" windows; see (3.1). By Lemma 7.2 (see also Remark 7.1) the probability of accepting a "bad" window $\widehat{U}$ is very small. More precisely, Lemma 7.2 and condition (2.3) imply

$$\mathbf{P}_f(\widehat{U} \in \mathcal{U}') \le \sum_{U \in \mathcal{U}'} \mathbf{P}_f(\rho_U = 0) \le \sum_{U \in \mathcal{U}'} e^{-t_U^2/2} \le \alpha,$$

and arguing as above we reduce our consideration to the case when $\widehat{U}$ is not "bad," that is, $\widehat{U} \in \mathcal{U}'' = \mathcal{U} \backslash \mathcal{U}'$. By condition (U.3), for each $U \in \mathcal{U}''$ with $N_U \ge N_{U^*}$, there is a $V \in \mathcal{V}(U)$ such that $V \subseteq \widehat{U} \cap A$ and $N_V \ge \nu N_{U^*}$. We denote this V by $V(U)$. The definition of $\widehat{U}$ ensures that $\rho_{\widehat{U},V} = 0$ and we conclude, using Lemma 7.3 with $z = t^* = \max_{U \notin \mathcal{U}'} t_U$ and (2.3),

$$\begin{aligned}
&\mathbf{P}_f\bigl(\bigl|\hat{f}_{\widehat{U}} - f(x^0)\bigr| > 2\sigma(\nu N_{U^*})^{-1/2} t^*\bigr) \\
&\quad \le 2\alpha + \sum_{U \in \mathcal{U}''} \mathbf{P}_f\bigl(\bigl|\hat{f}_U - f(x^0)\bigr| > \sigma N_{V(U)}^{-1/2}(t_U + t^*), \rho_{U,V(U)} = 0\bigr) \\
&\quad \le 2\alpha + \sum_{U \in \mathcal{U}''} e^{-t_U^2/2} \le 3\alpha
\end{aligned}$$

and the assertion follows.

7.2. *Proof of Theorem* 4.2. The statement of this theorem is a direct application of Theorem 3.1. The main problem is to verify that there is a window U^* from $\mathcal{U}_{d^*}$ which is contained in A. Then automatically, $N_{U^*} \ge N_{d^*}/2 \ge 2(d^*)^2$ and the assertion follows from Theorem 3.1.

Let $z = g'(x_1^0)$. We know that $|z| \le 1$. To be more definite we suppose that $0 \le z \le 1$. The case of a negative z can be considered in the same way. We denote also

$$\Delta = (4K^*/n)^{\gamma/(\gamma+1)} P^{1/(\gamma+1)}$$

so that $\Delta d^* = 4K^*\delta$; see (4.6).

LEMMA 7.4. *For all x_1 with $|x_1 - x_1^0| \le \delta d^*$,*

$$x_2^0 + \Delta + z(x_1 - x_1^0) \le g(x_1).$$

PROOF. The smoothness condition $g \in \Sigma(\gamma, P)$ implies for all $h > 0$,

$$\sup_{|t| \le h} |g(x_1^0 + t) - g(x_1^0) - zt| \le Ph^\gamma.$$

Therefore,

$$g(x_1^0 + t) \ge g(x_1^0) + zt - Ph^\gamma \ge x_2^0 + 2\Delta + zt - Ph^\gamma$$

for all $|t| \le h$. Now we apply $h = \delta d^*$ and the assertion follows because

$$P h^\gamma = P(\delta d^*)^\gamma \le P^{1/(\gamma+1)}(4K^*\delta^2)^{\gamma/\gamma+1} = \Delta. \qquad \square$$

To define the "ideal" window U^*, we utilize the following number-theoretical result.

LEMMA 7.5. *Let z be any number with $0 \le z \le 1$. Then for every integer number d there is a rational number p/q with $0 \le p < q \le d$ such that*

$$|zq - p| \le (d+1)^{-1}.$$

PROOF. Suppose without loss of generality that z is an irrational number from the interval [0, 1]. Denote by $(p_k/q_k)_{k\ge1}$ the sequence of rational numbers which gives the best rational approximation of z; see Khintchine (1949). It can be defined as a sequence of continued fractions: We begin with $r_0 = z^{-1}$ and define inductively $n_k = \lfloor r_{k-1} \rfloor$, $r_k = (r_{k-1} - n_k)^{-1}$ for $k = 1, 2, \ldots$; then p_k/q_k can be described as the following continued fraction:

$$\frac{p_k}{q_k} = \frac{1}{n_1 + \frac{1}{n_2 + \cdots \frac{1}{n_{k-1} + \frac{1}{n_k}}}}.$$

This approximation has the following properties [Khintchine (1949), Sections 3,4]:

$$\left| z - \frac{p_k}{q_k} \right| \le \frac{1}{q_k q_{k+1}}. \tag{7.3}$$

Given a number d, denote

$$k^* = \max\{k : q_k \le d\}$$

so that $q_{k^*+1} \ge d + 1$. By (7.3), $|zq_{k^*} - p_{k^*}| \le 1/q_{k^*+1} \le 1/(d+1)$ and the assertion follows. $\square$

An application of this lemma with $z = g'(x_1^0)$ and $d = d^*$ leads to a pair $(p, q) \in \mathcal{R}_{d^*}$ with $|zq - p| \le 1/(d^* + 1)$. We define $U^* = U_{d^*,(p,q)}$; see (4.1). The result of the theorem will follow from Theorem 3.1 if we check that $U^* \subset A$, that is, U^* lies below the curve G. To this end, we have to bound the quantity $\rho_{d,(p,q)}$.

LEMMA 7.6. *For every d and each $(p, q) \in \mathcal{R}_s$ with $|p| \le |q| \le d$,*

$$|\rho_{d,(p,q)}| \le 4K_d \delta q/d.$$

Moreover, if $d/|q| \ge K_d$, then $\rho_{d,(p,q)} = 0$.

PROOF. It suffices to consider the case with $0 \le p \le q \le d$ and to show that there exist at least K_d different grid points of the form $x = (x_1^0 + n_1\delta, x_2^0 + n_2\delta)$ with $0 < n_1 \le d$, $0 < n_2 \le d$ such that

$$-2qK_d/d \le pn_1 - qn_2 \le 0.$$

Define $k_0 = [d/q]$. There exist at least $2k_0 + 1$ grid points on the line $L_{p,q} = \{x : p(x_1 - x_1^0) - q(x_2 - x_2^0) = 0\}$ within the square Q_d and therefore, the total number of grid points on and under the line $L_{p,q}$ is at least $k_0 + (2d+1)^2/2$. If $k_0 \ge K_d$, then condition (4.1) is fulfilled with $\rho_{d,(p,q)} = 0$. If $k_0 < K_d$, then set $(p_0, q_0) = (p, q)$ and define successfully for $k = 1, 2, \dots,$

$$z_k = p_{k-1}/q_{k-1} + 1/(q_{k-1}d),$$

and (p_k, q_k) as a pair from $\mathcal{R}_d$ with $|z_k q_k - p_k| \le 1/(d+1)$ due to Lemma 7.5. The equality $(p_{k-1}, q_{k-1}) = (p_k, q_k)$ is impossible since $|z_k q_{k-1} - p_{k-1}| = 1/d > 1/(d+1)$. Therefore, $p_k/q_k > p_{k-1}/q_{k-1}$, but

$$p_k/q_k - p_{k-1}/q_{k-1} = 1/(q_{k-1}d) + p_k/q_k - z_k < 1/(dq_{k-1}) + 1/(dq_k).$$

We also set $m_k = [d/q_k]$ and continue as previously until $m_0 + m_1 + \dots + m_k \ge K_d$.

The grid point $X^{(k,\ell)} = (x_1^0 + \ell q_k\delta, x_2^0 + \ell p_k\delta)$ with $\ell = 1, \dots, m_k$ lies inside Q_d and it holds that $p\ell q_k - q\ell p_k < 0$ and

$$\begin{aligned}
|p\ell q_k - q\ell p_k| &\le q m_k q_k |p/q - p_k/q_k| \le qd \sum_{i=1}^{k} |p_{i-1}/q_{i-1} - p_i/q_i| \\
&\le \frac{q}{d} \sum_{i=1}^{k} (d/q_{i-1} + d/q_i) \le \frac{q}{d} \sum_{i=1}^{k} (m_{i-1} + m_i + 2) \le 4qK_d/d.
\end{aligned}$$

Therefore, there exist at least $m_0 + \dots + m_d \ge K_d$ grid points inside Q_d satisfying the required conditions and the assertion follows. $\square$

Let $z = g'(x_1^0)$. Consider the split of the square Q_{d^*} by the line passing through the point x^0 with the slope z. Since $s \ge 2d^* + 1$, there exists a pair $(p, q) \in \mathcal{R}_s$ defining the same split (containing the same collection of the grid points inside Q_{d^*}). Therefore, without loss of generality one may suppose that $z = p/q$.

Lemmas 7.4 and 7.6 imply for $x = (x_1, x_2) \in U_{d^*,(p,q)}$ in view of $4K_d\delta \le 4K^*\delta = \Delta d^*$,

$$\begin{aligned}
x_2 &\le x_2^0 + (x_1 - x_1^0)p/q + \rho_{d^*,(p,q)}/q \le x_2^0 + (x_1 - x_1^0)z + 4\delta K_d/d^* \\
&\le x_2^0 + (x_1 - x_1^0)z + \Delta \le g(x_1),
\end{aligned}$$

that is, $x \in A$ and hence, $U^* \subset A$.

Since the set A has the structure of a boundary fragment within the square $|x - x^0| \le \varepsilon$, condition (U.3) for the above U^* is clearly satisfied with $\nu = 0.5$. Since also $N_{U^*} \ge 2(d^*)^2$, an application of Theorem 3.1 leads exactly to the desired statement.

7.3. *Proof of Theorem* 4.3. The proof of this result can be derived along the same line as the proof of Theorem 4.2 and is even simpler. Indeed, we may take the line L passing through x^0 with the angle $z = p/q$. Then this line passes also through the design points $X^{(k)} = (x_1^0 + kq, x_2^0 + kp)$ for all integer k. The interval between the points $X^{(-K)}$ and $X^{(K)}$ on this line contains at least $2K + 1$ design points and therefore the window $U_{d,L}$ with $d = Kq$ is in $\mathcal{U}$. The condition (4.7) provides that this window is also in A and we end similarly to Theorem 4.2.

7.4. *Proof of Theorem* 4.4. Different methods for obtaining the lower bound results in edge estimation are presented in Korostelev and Tsybakov (1993). We cannot apply these methods directly since they are developed for a random design and we operate with the regular design. However, we follow the same route and we therefore present only a sketch of the proof, concentrating on the points specific for our situation.

Let some γ from the interval $(1, 2]$ and some integers K, D be fixed. Let also $z = p/q$ be a unreducible rational number with $p \le q \le D$. Set

$$h = \min\{qK\delta, (\delta/q)^{1/\gamma}\},$$

where $\delta = n^{-1/2}$.

Let now ϕ be a smooth function satisfying the conditions:

(a) ϕ is symmetric and nonnegative;
(b) $\phi(0) = \sup_t \phi(t)$ and $0 < \phi(0) \le 1$;
(c) ϕ is compactly supported on $[-1, 1]$;
(d) ϕ belongs to the Hölder ball $\Sigma(\gamma, 1)$.

Denote

$$\phi_h(t) = h^\gamma \phi(t/h).$$

Then (d) ensures that $\phi_h \in \Sigma(\gamma, 1)$ for all $h > 0$. Next, set

$$g_1(x_1) = (x_1 - x_1^0)p/q - \phi_h(0)/2,$$
$$g_2(x_2) = (x_1 - x_1^0)p/q + \phi_h(0)/2 - \phi_h(x_1 - x_1^0).$$

Each function g_k determines the boundary fragment A_k with the edge G_k,

$$A_k = \{x = (x_1, x_2) : x_2 \le g_k(x_1)\},$$
$$G_k = \{x = (x_1, x_2) : x_2 = g_k(x_1)\}, \qquad k = 1, 2.$$

Set also

$$B = A_2 \backslash A_1 = \{x = (x_1, x_2) : g_1(x_1) < x_2 \le g_2(x_1)\}.$$

Below we make use of the following technical assertion.

LEMMA 7.7. *The following assertions hold*:

(i) $g_1, g_2 \in \Sigma(\gamma, 1)$ *and* $g_1'(x_1^0) = g_2'(x_1^0) = q/p$;
(ii) $|g(x_1^0) - x_2^0| \ge \kappa h^\gamma$ *for some* $\kappa > 0$ *depending on* ϕ *only*;
(iii) *the number* N *of design points in the set* B *is at most* $2K - 1$,

$$N = \#\{X_i \in B\} \le 2K - 1.$$

PROOF. Assertions (i) and (ii) are obvious. We comment on (iii).

Let L be the line passing through x^0 with the angle z; that is, L is described by the equation $x_2 - x_2^0 = z(x_1 - x_1^0)$. We fix two points $x^- = (x_1^0 - Kq\delta, x_2^0 - Kp\delta)$ and $x^+ = (x_1^0 + Kq\delta, x_2^0 + Kp\delta)$ on this line. Since $h \le qK\delta$, the interval passes exactly through $2K - 1$ design points. We intend to show that there are no other design points in B, which implies the assertion in view of property (c) of ϕ.

Let $x = (x_1, x_2)$ be a design point with coordinates $(x_1^0 + q'\delta, x_2^0 + p'\delta)$ such that $p'/q' \ne p/q$. To verify that $x \notin B$, it suffices to check that

$$|p'\delta - q'\delta p/q| > |\phi_h(q'\delta) - \phi_h(0)/2|.$$

Since $p'/q' \ne p/q$, then

$$|p' - q'p/q| = q^{-1}|p'q - q'p| \ge q^{-1}$$

and hence $|p'\delta - q'\delta p/q| \ge \delta/q$. In view of (b), we have $\phi_h(x_1 - x_1^0) \le \phi_h(0) \le h^\gamma$ and by definition of h we have $h^\gamma \le \delta/q$ and (iii) follows. $\square$

Denote $f_k(x) = \mathbf{1}(x \notin A_k) = \mathbf{1}(x_2 > g_k(x_1))$ for $x = (x_1, x_2)$, $k = 1, 2$. Note that $f_1(x^0) = 0$ and $f_2(x^0) = 1$. Now for any estimator $\tilde f(x^0)$,

$$\begin{aligned} R &:= \mathbf{P}_1(|\tilde f(x^0)| > 1/2) + \mathbf{P}_2(|\tilde f(x^0) - 1| > 1/2), \\ &= \boldsymbol{E}_1\{\mathbf{1}(|\tilde f(x^0)| > 1/2) + Z\mathbf{1}(|\tilde f(x^0) - 1| > 1/2)\}, \end{aligned} \tag{7.4}$$

where $\boldsymbol{E}_k$ stands for $\boldsymbol{E}_{g_k}$, $k = 1, 2$, and $Z = d\mathbf{P}_2/d\mathbf{P}_1$. It is easy to show that the optimal decision $\tilde f(x^0)$ for the latter two-point problem is of the form $\tilde f(x^0) = \mathbf{1}(Z \ge 1)$ and hence

$$R \ge \boldsymbol{E}_1\mathbf{1}(Z \ge 1) = \mathbf{P}_1(Z \ge 1).$$

Next, making use of the model equation (1.1) we get the following representation of the likelihood Z:

$$Z = \exp\left\{\sigma^{-2}\sum_B \xi_i - \frac{N\sigma^{-2}}{2}\right\},$$

where the sum over B means the sum over design points X_i falling in B and the random errors ξ_i are normal $\mathcal{N}(0, \sigma^2)$. If we set

$$\zeta = \frac{1}{\sigma\sqrt{N}}\sum_B \xi_i,$$

then Lemma 7.7(ii) and (iii) implies that ζ is under $\mathbf{P}_1$ a standard normal random variable and

$$\begin{aligned}\mathbf{P}_1(Z > 1) &= \mathbf{P}_1(\exp\{\sigma^{-1}\sqrt{N}\zeta - \sigma^{-2}N/2\} > 1)\\ &= \mathbf{P}_1(\zeta > \sigma^{-1}\sqrt{N}/2)\\ &\le \mathbf{P}_1(\zeta > \sigma^{-1}\sqrt{K/2})\\ &= 1 - \Phi(\sigma^{-1}\sqrt{K/2}) > 0,\end{aligned}$$

where Φ is the standard normal cdf and the required assertion follows.

Acknowledgments. The authors thank A. Juditski, O. Lepski, A. Tsybakov and Yu. Golubev for important remarks and discussion.

REFERENCES

BARRON, A., BIRGÉ, L. and MASSART, P. (1999). Risk bounds for model selection via penalization. *Probab. Theory Related Fields* **113** 301–413.

BLAKE, A. and ZISSERMAN, A. (1987). *Visual Reconstruction*. MIT Press.

CHU, C. K., GLAD, I. K., GODTLIEBSEN, F. and MARRON, J. S. (1998). Edge-preserving smoothers for image processing (with discussion). *J. Amer. Statist. Assoc.* **93** 526–556.

DONOHO, D. L. (1999). Wedgelets: Nearly minimax estimation of edges. *Ann. Statist.* **27** 859–897.

ENGEL, J. (1994). A simple wavelet approach to nonparametric regression from recursive partitioning schemes. *J. Multivariate Anal.* **49** 242–254.

GEMAN, S. and GEMAN, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Trans. Pattern Analysis Machine Intelligence* **6** 721–741.

GIRARD, D. (1990). From template matching to optimal approximation by piecewise smooth curves. In *Curves and Surfaces in Computer Vision and Graphics* (L. A. Ferrari and R. S. de Figueiredo, eds.) 174–182. SPIE, Bellingham, WA.

GLASBEY, C. A. and HORGAN, G. W. (1995). *Image Analysis for the Biological Sciences*. Wiley, New York.

GODTLIEBSEN, F., SPJØTVOLL, E. and MARRON, J. S. (1997). A nonlinear Gaussian filter applied to images with discontinuities. *J. Nonparametr. Statist.* **8** 21–43.

GRENANDER, U. (1976, 1978, 1981). *Lectures in Pattern Theory* **1**, **2**, **3**. Springer, New York.

HALL, P. and RAIMONDO, M. (1998). On global performance of approximations to smooth curves using gridded data. *Ann. Statist.* **26** 2206–2217.

HARALICK, R. M. (1980). Edge and region analysis for digital image data. *Comput. Graphics Image Processing* **12** 60–73.

HUANG, J. S. and TSENG, D. H. (1988). Statistical theory of edge detection. *Comput. Vision Graphics Image Processing* **43** 337–346.

KHINTCHINE, A. I. (1949). *Continued Fractions*. ITTL, Moscow.

KOROSTELEV, A. and TSYBAKOV, A. (1993). *Minimax Theory of Image Reconstruction*. Springer, New York.

LEE, J. S. (1983). Digital image smoothing and the sigma filter. *Comput. Vision Graphics Image Processing* **24** 255–269.

LEPSKI, O. V. (1990). A problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35** 454–466.

LEPSKI, O. V. (1992). Asymptotically minimax adaptive estimation. II. Statistical model without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.* **37** 433–448.

LEPSKI, O., MAMMEN, E. and SPOKOINY, V. (1997). Optimal spatial adaptation to inhomogeneous smoothness: An approach based on kernel estimates with variable bandwidth selectors. *Ann. Statist.* **25** 929–947.

LEPSKI, O. and SPOKOINY, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Ann. Statist.* **25** 2512–2546.

MAMMEN, E. and TSYBAKOV, A. (1995). Asymptotical minimax recovery of sets with smooth boundaries. *Ann. Statist.* **23** 502–524.

MARR, D. (1982). *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. Freeman, San Francisco.

MÜLLER, H.-G. and SONG, K. S. (1994). Maximin estimation of multidimensional boundaries. *J. Multivariate Anal.* **50** 265–281.

MUMFORD, D. and SHAH, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* **42** 577–685.

NASON, G. P. and SILVERMAN, B. W. (1994). The discrete wavelet transform in *S*. *J. Comput. Graph. Statist.* **3** 163–191.

POLZEHL, J. and SPOKOINY, V. (2000). Adaptive weights smoothing with applications to image restoration. *J. R. Stat. Soc. Ser. B Stat. Methodol.* **62** 335–354.

PRATT, W. K. (1978). *Digital Image Processing*. Wiley, New York.

QIU, P. (1998). Discontinuous regression surfaces fitting. *Ann. Statist.* **26** 2218–2245.

RIPLEY, B. D. (1988). *Statistical Inference for Spatial Processes*. Cambridge Univ. Press.

ROSENFELD, A. and KAK, A. C. (1982). *Digital Picture Processing*. Academic Press, London.

SCOTT, D. W. (1992). *Multivariate Density Estimation*. Wiley, New York.

SPOKOINY, V. (1998). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. *Ann. Statist.* **26** 1356–1378.

TITTERINGTON, D. M. (1985). Common structure of smoothing techniques in statistics. *Internat. Statist. Rev.* **53** 141–170.

TSYBAKOV, A. (1989). Optimal orders of accuracy of the estimation of nonsmooth images. *Problems Inform. Transmission* **25** 180–191.

WINKLER, G. (1995). *Image Analysis, Random Fields and Dynamic Monte Carlo Methods*. Springer, Berlin.

WEIERSTRASS INSTITUTE
MOHRENSTRASSE 39
D-10117 BERLIN
GERMANY
E-MAIL: polzehl@wias-berlin.de