IMAGE DENOISING: POINTWISE ADAPTIVE APPROACH

SPOKOINY, V.G.

Weierstrass Institute for Applied Analysis and Stochastics,
Mohrenstr. 39, 10117 Berlin

1991 *Mathematics Subject Classification.* 62G07; Secondary 62G20.

Key words and phrases. edge, design, grid, image, pointwise estimation, rate of image and edge estimation, data-driven window .

Abstract. The paper is concerned with the problem of image denoising. We consider the case of black-white type images consisting of a finite number of regions with smooth boundaries and the image value is assumed to be piecewise constant within each region. New method of image denoising is proposed which is adaptive (assumption free) to the number of regions and smoothness properties of edges. The method is based on a pointwise image recovering and it relies on an adaptive choice of a smoothing window. It is shown that the attainable quality of estimation depends on the distance from the point of estimation to the closest boundary and on the smoothness properties of this boundary. As a consequence, it turns out that the proposed method provides the optimal rate of the edge estimation.

1. Introduction

One of the main problem of image analysis is reconstruction of an image (a picture) from noisy data. It has been intensively studied last years, see e.g. the books of Pratt (1978), Grenander (1976, 1981), Rosenfeld and Kak (1982), Blake and Zisserman (1987), Korostelev and Tsybakov (1994). There are two special features related to this problem. First, the data is two-dimensional (or multidimensional). Second, the image is usually 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 edge estimation problem.

A large variety of methods has been proposed for solving the image and edge estimation problem in different contexts. The most popular methods of image estimation are based on Bayesian approach for certain parametric image modelling Haralick (1980), Geman and Geman (1984), Ripley (1988) among other. Some nonparametric methods based on penalizing and regularization technique have been developed in Titterington (1985), Shiau, Wahba and Johnson (1986), Mumford and Shah (1989), Girard (1990).

The edge detection methods mostly do not assume any underlying parametric model. The methods based on a kernel smoothing with a special choice of kernels have been discussed in Pratt (1978), Marr (1982), Lee (1983), Huang and Tseng (1988), Müller and Song (1994). Korostelev and Tsybakov (1994) developed the general asymptotic minimax theory of edge estimation. In particular, they showed that linear methods are not optimal for images with sharp edges. Imposing some smoothness restrictions on the boundary they have described the optimal rate of estimation and rate-optimal estimators for images with the structure of a boundary fragment. The proposed methods are essentially nonlinear and they involve linewise change-point analysis.

In the present paper, we propose another approach which is based on direct image estimation at each design point. We apply a simple linear estimator which is the average of observations over a window selected in a data-driven way. Then we study which accuracy of edge estimation is provided by this procedure. It has been already mentioned that linear methods are only suboptimal in edge estimating. The results of this paper show that a non-linearity which is incorporated in the linear method by an adaptive choice of an averaging window allows to get optimal quality of edge recovering.

The presented approach 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), Spokoiny (1996) where pointwise adaptive procedures were applied to estimating a function with heterogeneous smoothness properties, allowing, for instance, jumps or jumps of derivatives. The methods based on pointwise (or local) adaptation are especially fruitful in such situations since a complex object like a function with heterogeneous smoothness properties admits a simple description in a small neighborhood of each point and the procedure, being applied at this point, adapts exactly to the underlying local structure. In essence, the procedure searches for a largest local vicinity of the point of estimation in which the local structural assumption fits well to the data.

Now we apply this idea to the problem of image estimation. We focus on the case of a piecewise constant images i.e. we assume that the image consists of a finite number of regions and the image value is constant within each region. The image is observed with a noise on the regular grid in the unit square and we estimate the image value separately at each design point via a data-driven choice of the averaging window. The benefit of this approach is that it is very general in nature and it is not required to specify the number of regions, difference between values of the image function f for different regions or regularity of each edge. Moreover, this method can be applied to estimating any function which can be well approximated by a constant function in a local vicinity of each point. Further developments of this approach lead to local approximation by a linear function or, more generally, by a polynomial.

We consider the regression model

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

where $X_i \in [0,1]^d$, $i = 1, \ldots, n$, are given design points and ξ_i are individual independent random errors. Below we will suppose that ξ_i, $i = 1, \ldots, n$, are i.i.d. $\mathcal{N}(0, \sigma^2)$ with a given noise level σ.

Next we suppose that the cube $[0,1]^d$ is split into M regions A_m, $m = 1, \ldots, M$ each of them is a connected set with an edge (boundary) G_m. We suppose also that the function f is constant within each region A_m, i.e.

$$f(x) = a_m \mathbf{1}(x \in A_m) \tag{1.2}$$

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

The idea of the proposed method is quite simple and natural. We search for a maximal possible window U containing x^0 in which the function f is well approximated by a constant which is exactly the resulting estimate. Of course, the key role for such an approach plays the choice of the considered class of windows. We will discuss this problem a little bit later. Now for a moment we suppose that we are given a class $\mathcal{U}$ of windows U each of them is 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 calculate the residuals $\varepsilon_{U,i} = Y_i - \hat{f}_U$ and we test the hypothesis that these residuals $\varepsilon_{U,i}$ can be treated within the window U as a 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, Section 3 contains the results describing the quality of this procedure. In Section 4 we specify the general results to the case of the equidistant design.

2. Estimation Procedure

Let data Y_i, X_i, $i = 1, \ldots, n$ obey model (1.1). We will estimate $f(x^0)$ for a given x^0.

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

$$N_U = \#\{X_i \in U\}. \tag{2.1}$$

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

$$\hat{f}_U(x^0) = \hat{f}_U = \frac{1}{N_U} \sum_U Y_i. \tag{2.2}$$

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 is a subwindow of U, i.e. $V \subset U$. We require only that $N_V := \#\{X_i \in V\} \geq 2$ for all $V \in \mathcal{V}(U)$. One example of the choice of the families $\mathcal{U}$ and $\mathcal{V}(U)$ for the case of the equidistant design is presented in Section 4.

Below we need an upper estimate of the cardinality of $\mathcal{V}(U)$ in the form

$$\#\mathcal{V}(U) \leq n^\alpha \tag{2.3}$$

with some $\alpha > 0$.

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 $\hat{f}_V - \hat{f}_U$,

$$\sigma_{U,V}^2 = \frac{N_U - N_V}{N_U N_V} \sigma^2. \tag{2.4}$$

Define now

$$\varrho_{U,V} = \mathbf{1}\left(|T_{U,V}| > t\sqrt{\log n}\right)$$

where

$$t = \sqrt{2(\alpha + r)}.$$

The parameter r has meaning of the norm in which we measure losses of estimation. Typically r is taken equal to 2.

We say that U is rejected if $\varrho_{U,V} = 1$ at least for one $V \in \mathcal{V}(U)$ i.e. if $\varrho_U = 1$ where

$$\varrho_U = \sup_{V \in \mathcal{V}(U)} \varrho_{U,V} = \mathbf{1}\left(\sup_{V \in \mathcal{V}(U)} |T_{U,V}| > t\sqrt{\log n} \right).$$

The adaptive procedure selects among all non-rejected U from $\mathcal{U}$ such one which maximizes N_U,

$$\hat{U} = \operatorname*{argmax}_{U \in \mathcal{U}} \{ N_U : \varrho_{U,V} = 0 \text{ for all } V \in \mathcal{V}(U) \} \tag{2.5}$$

and

$$\hat{f}(x^0) = \hat{f}_{\hat{U}}(x^0) = \hat{f}_{\hat{U}}. \tag{2.6}$$

3. Main results

Below we describe some properties of the above proposed estimation procedure and state the result about the corresponding accuracy of estimation.

We begin with the following remark. An "ideal" window for estimating $f(x^0)$ coincides clearly with the region A_m containing x^0. Hence the idea of the proposed procedure is to select adaptively the largest window among the considered class $\mathcal{U}$ which is contained in A_m. A necessary property of every such procedure is to accept with a high probability each window contained in A_m. Our first result shows that the above procedure possesses this properties.

Proposition 3.1. *Let $x^0 \in A_m$ for some $m = 1, \dots, M$ and let $U \in \mathcal{U}$ be such that $U \subseteq A_m$. Then*

$$\boldsymbol{P}_f(\varrho_U = 1) \le n^{-r}.$$

Proof. Let some U with the above property be fixed and let $V \in \mathcal{V}(U)$. The function f is constant on U and hence on V and 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 $\boldsymbol{E} T_{U,V} = 0$. Recall now that the multiplicator $\sigma_{U,V}$ was defined as the standard deviation of the stochastic term of the difference $\hat{f}_V - \hat{f}_U$. Hence $\boldsymbol{E} T_{U,V}^2 = 1$. Since also $T_{U,V}$ is a linear combination of Gaussian variables ξ_i, then $T_{U,V}$ is also a Gaussian zero mean random variable with the unit variance and therefore it is standard normal. Now

$$\boldsymbol{P}_f\left(|T_{U,V}| > \sqrt{2(\alpha + r)\log n} \right) \le \exp\{-(\alpha + r)\log n\} = n^{-(\alpha + r)}.$$

This estimate and condition (2.3) allow to bound the probability of rejecting U in the following way

$$\boldsymbol{P}_f(\varrho_U = 1) \le \sum_{V \in \mathcal{V}(U)} \boldsymbol{P}_f\left(|T_{U,V}| > \sqrt{2(\alpha+r)\log n}\right) \le \#\mathcal{V}(U) n^{-(\alpha+r)} \le n^{-r}.$$

□

Remark 3.1. The above result prompts the following definition of an "ideal" window U^* from $\mathcal{U}$ containing a given point x^0 from region A_m: it is the largest (in the number of design points) window from $\mathcal{U}$ which is contained in A_m. Set

$$N^* = \max\{N_U : U \in \mathcal{U},\, U \subseteq A_m\}.$$

If there is no any window with this property, then we simply set $N^* = 0$.

The next statement can be viewed as a complement to Proposition 3.1. We consider now the case of a "bad" window containing two non-intersecting subwindows V_1 and V_2 with different values of the image function f. We intend to show that our procedure rejects with a high probability every such a window if both V_1 and V_2 contain enough design points.

Proposition 3.2. *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,*

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

If

$$|a_1 - a_2| \ge (\sigma_{U,V_1} + \sigma_{U,V_2})\, t\sqrt{\log n} + \sigma\sqrt{N_{V_1}^{-1} + N_{V_2}^{-1}}\sqrt{2r\log n} \tag{3.1}$$

with $t = \sqrt{2(\alpha+r)}$ and $\sigma_{U,V}^2 = \sigma(N_V^{-1} - N_U^{-1})$, then

$$\boldsymbol{P}_f(\varrho_U = 0) \le n^{-r}.$$

Remark 3.2. 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 (3.1) is fulfilled if

$$|a_1 - a_2| \ge \sigma\sqrt{\log n}\left(\sqrt{2(\alpha+r)} + \sqrt{2r}\right)\left(N_{V_1}^{-1/2} + N_{V_2}^{-1/2}\right). \tag{3.2}$$

Proof. By definition

$$\boldsymbol{P}_f(\varrho_U = 0) \le \boldsymbol{P}_f(\varrho_{U,V_1} = \varrho_{U,V_2} = 0).$$

Next, the event $\{\varrho_{U,V} = 0\}$ means $|T_{U,V}| \le t\sqrt{\log n}$ or equivalently

$$|\hat{f}_U - \hat{f}_V| \le t\sigma_{U,V}\sqrt{\log n}.$$

This yields in particular

$$|\hat{f}_{V_1} - \hat{f}_{V_2}| \le t(\sigma_{U,V_1} + \sigma_{U,V_2})\sqrt{\log n}.$$

Now using the fact that $V_1 \cap V_2 = \emptyset$, we get the following decomposition, cf. the proof of Proposition 3.1,

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

where $s_{1,2} = \sigma(N_{V_1}^{-1} + N_{V_2}^{-1})$ and $\zeta_{1,2}$ is a standard normal random variable. Therefore,

$$\begin{aligned} \boldsymbol{P}_f(\varrho_U = 0) &\le \boldsymbol{P}(|a_1 - a_2 + s_{1,2}\zeta_{1,2}| \le t(\sigma_{U,V_1} + \sigma_{U,V_2})\sqrt{\log n}) \\ &\le \boldsymbol{P}\left(s_{1,2}|\zeta_{1,2}| \ge |a_1 - a_2| - t(\sigma_{U,V_1} + \sigma_{U,V_2})\sqrt{\log n}\right). \end{aligned}$$

Using the condition of the proposition, we obtain

$$\boldsymbol{P}_f(\varrho_U = 0) \le \boldsymbol{P}(|\zeta_{1,2}| > \sqrt{2r \log n}) \le n^{-r}$$

and the assertion follows. □

We need one more result concerning the situation when $x^0 \in A_m$, a window U from $\mathcal{U}$ is not contained in A_m but there is a subwindow V of U which is in A_m.

Proposition 3.3. *Let $x^0 \in A_m$, $U \in \mathcal{U}$ and let V from $\mathcal{V}(U)$ be such that $V \subseteq A_m$. If $\varrho_{U,V} = 0$, then*

$$\boldsymbol{E}_f|\hat{f}_U - f(x^0)|^r \le (C\sigma^2 N_V^{-1} \log n)^{r/2}$$

with $C = (t+1)^2$.

Proof. The event $\{\varrho_{U,V} = 0\}$ means that $|\hat{f}_U - \hat{f}_V| \le t\sigma_{U,V}\sqrt{\log n}$. Therefore,

$$|\hat{f}_U - a_m| \le |\hat{f}_U - \hat{f}_V| + |\hat{f}_V - a_m| \le t\sigma_{U,V}\sqrt{\log n} + |\hat{f}_V - a_m|.$$

Next, $\sigma_{U,V} \le \sigma N_V^{-1/2}$ and $\zeta = \sigma^{-1} N_V^{1/2}(\hat{f}_V - a_m)$ is a standard Gaussian random variable, see the proof of Proposition 3.1. This gives

$$\boldsymbol{E}_f|\hat{f}_U - f(x^0)|^r \le \boldsymbol{E}\left|t\sigma\sqrt{N_V^{-1}\log n} + \sigma N_V^{-1/2}\zeta\right|^r \le \left((t+1)\sigma\sqrt{N_V^{-1}\log n}\right)^r$$

as required. □

To state the results about the quality of estimation by our adaptive procedure, we have to be more definitive with the choice of the class $\mathcal{U}$. However, we would like to keep some freedom in specification of this class which might be helpful for applications. We impose therefore some rather mild conditions on this class and then formulate the results under these conditions.

Recall that $\mathcal{U}$ is defined as a set of windows containing the point of interest x^0, and for each $U \in \mathcal{U}$, the family $\mathcal{V}(U)$ is assigned for testing U. Below we assume that these families fulfill the following conditions:

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

$(U.2)$ there is a natural number K such that for every $U, U' \in \mathcal{U}$, the intersection $U \cap U'$ contains a testing window $V \in \mathcal{V}(U)$ with

$$N_V \ge K.$$

The choice of the constant K will be discussed later.

Discussion 3.1. Condition $(U.2)$ from the above did not appear in the univariate case, see Spokoiny (1996), but it is rather essential in the multivariate situation. This can be illustrated by the following example. Let the point of interest x^0 lie near the boundary of a region, say A_1, with another region, say A_2. Then it might be possible that there is a thin near x^0 set U containing x^0 but all or almost all the remaining design points in U belong to A_2. If additionally the total number of such points in U is large, it may happen that we estimate $\hat{f}(x^0) \asymp a_2$ while really $f(x^0) = a_1$. Condition $(U.2)$ is just to prohibit U from $\mathcal{U}$ to be thin near x^0.

Conditions $(U.1)$ and $(U.2)$ rely only on the structure of the set $\mathcal{U}$ of considered windows. We need one more condition which relies also on the properties of the edge of the region A_m containing x_0.

$(U.3)$ There is a constant $\varepsilon > 0$ such that if $U \in \mathcal{U}$ and if for every V from $\mathcal{V}(U)$ with $N_V \geq K$, it holds

$$V \cap A_m \neq \emptyset,$$

then either $N_U < N^*$ or there is $V_1 \in \mathcal{V}(U)$ such that $V_1 \subseteq A_m$ and

$$N_{V_1} \geq \epsilon N_U.$$

Discussion 3.2. The reason for introducing this condition can be explained by the following considerations. Let $x^0 \in A_m$ and let U^* be the corresponding "ideal" window. Suppose that $N_{U^*} \geq K$. By Proposition 3.1, our procedure rejects U^* with a very small probability. Let now the procedure selects a window $\hat{U}$. Then, conditionally $\varrho_{U^*} = 0$, we have $N_{\hat{U}} \geq N_{U^*}$. By condition $(U.2)$ there is a window $V \in \hat{U} \cap U^*$ with $N_V \geq K$. Since $U^* \subseteq A_m$, then also $V \subseteq A_m$. If $\hat{U}$ contains another subwindow V' with $N_{V'} \geq K$ which lies outside of A_m and if K is large enough, then by Proposition 3.2 such a window U will be rejected with a high probability. Therefore, typically we have for the window $\hat{U}$ that $N_{\hat{U}} \geq N_{U^*}$ and every V' from $\mathcal{V}(\hat{U})$ has a non-trivial intersection with A_m. Exactly for this situation, we suppose in addition that it contains a subwindow $V_1 \in \mathcal{V}(U)$ which is in A_m and N_{V_1} is of order $N^* = N_{U^*}$. Then Proposition 3.3 ensures a considerable quality of estimation.

Now we are in a position to state the main result.

Theorem 3.1. *Let the image function $f(x)$ be piecewise constant due to (1.2) with $|a_m| \leq 1$, $m = 1, \ldots, n$. Let some $K > 0$ be fixed and let the conditions $(U.1)$ through $(U.3)$ be satisfied. If for all $m' \neq m$, it holds*

$$|a_m - a_{m'}| \geq 2\sigma\sqrt{\log n}\left(\sqrt{2(\alpha + r)} + \sqrt{2r}\right) K^{-1/2}, \tag{3.3}$$

then

$$\boldsymbol{E}_f|\hat{f}(x^0) - f(x^0)|^r \leq \left(C\frac{\sigma^2 \log n}{\epsilon N^*}\right)^{r/2} \tag{3.4}$$

with C from Proposition 3.3.

Discussion 3.3. Condition (3.3) gives some information about a reasonable choice of the parameter K entering in condition $(U.2)$. To be able to provide a high quality of image recognition, we have to allow small values δ for the difference $|a_m - a_{m'}|$. This results in increasing the required value of K, namely K should be at least $C(\alpha, r)\delta^{-2}\sigma^2 \log n$ where the constant $C(\alpha, r)$ depends on α, r only. In particular, to be able to proceed with images with essentially different values a_m, we need to take K of a logarithmic order. At the same time, a large value of K decreases the quality of estimation near the boundary, see the next section. Therefore, the choice of this parameter is a rather delicate problem which needs some further study.

Discussion 3.4. Suppose that for each point x^0 of the cube $[0,1]^d$, we are given adjusted families $\mathcal{U}$ and $\mathcal{V}(U)$ (depending, of course, on x^0) satisfying conditions $(U.1)$ through $(U.2)$ with some fixed integer K. Similarly to the above, we may define for every x the corresponding 'ideal' window $U^* = U^*(x)$ and the data-driven window $\hat{U} = \hat{U}(x)$. This leads to the global image estimator $\hat{f}(x) = \hat{f}_{\hat{U}(x)}$. Denote, given K and $\epsilon > 0$, by $\mathcal{D}(K,\epsilon)$ the set of point x for which $N_{U^*(x)} \geq K$ and condition $(U.3)$ holds with a prescribed ϵ. Then the result of Theorem 3.1 can be stated uniformly in x.

Theorem 3.2. *Under the condition (3.3),*

$$\boldsymbol{E}_f \sup_{x\in\mathcal{D}(K,\epsilon)} \left|\sqrt{N_{U^*(x)}}(\hat{f}(x) - f(x))\right|^r \leq (C\sigma^2\epsilon^{-1}\log n)^{r/2}.$$

Now we present the proof of Theorem 3.1. Theorem 3.2 can be proved in the same line.

3.1. Proof of Theorem 3.1

Let $x^0 \in A_m$ and let U from $\mathcal{U}$ be such that $U \subseteq A_m$ and $N_U = N^*$. Then due to Proposition 3.1 the window U will be rejected only with a very small probability, namely

$$\boldsymbol{P}_f(\varrho_U = 1) \leq n^{-r}.$$

Denote

$$\begin{aligned} a^* &= \max\{|a_m - a_{m'}|,\, m' = 1,\dots,M\}, \\ \xi^* &= \max\{|\xi_i|,\, i = 1,\dots,n\}. \end{aligned}$$

By the conditions of the theorem, $a^* \leq 2$. Next, see e.g. Petrov (1975), for each $z \geq \log n$,

$$\boldsymbol{P}(\xi^* > \sqrt{2z}) \leq ne^{-z^2}$$

that yields

$$\boldsymbol{E}(2+\xi^*)^r \mathbf{1}\left(\xi^* > \sqrt{2(r+1)\log n}\right) \leq C(r)n^{-r}$$

with a constant $C(r)$ depending on r only. Since obviously

$$|\hat{f}(x^0) - f(x^0)| \leq a^* + \xi^* \leq 2 + \xi^*,$$

then we get

$$\begin{aligned} &\boldsymbol{E}_f|\hat{f}(x^0) - f(x^0)|^r \mathbf{1}(\varrho_U = 1) \\ &\quad\leq \boldsymbol{E}(2+\xi^*)^r \mathbf{1}\left(\xi^* > \sqrt{2r\log n}\right) + \left(2 + \sqrt{2r\log n}\right)^r \boldsymbol{P}_f(\varrho_U = 1) \\ &\quad\leq C(r)n^{-r} + \left(2 + \sqrt{2(r+1)\log n}\right)^r n^{-r}. \end{aligned}$$

This bound is essentially smaller that is required in the theorem and we may therefore focus on the case when U is accepted i.e. $\varrho_U = 0$.

Let window $\hat{U}$ be selected by the procedure. Then $\varrho_{\hat{U}} = 0$ and, assuming also $\varrho_U = 0$, by definition of $\hat{U}$

$$N_{\hat{U}} \geq N_U.$$

Next, due to condition $(U.2)$, there is a subwindow V in $\hat{U}$ with at least K design points which is contained in A_m. If also $\hat{U}$ contains another subwindow V' with $N_{V'} \geq K$ which lies outside of A_m, then we observe by Proposition 3.2, see also Remark 3.2, that the probability to accept $\hat{U}$ is very small,

$$\boldsymbol{P}_f(\varrho_{\hat{U}} = 0) \leq n^{-r},$$

and arguing as above we reduce our consideration to the case when $\hat{U} \backslash A_m$ does not contain any such V'.

By condition $(U.3)$, there is $V \in \mathcal{U}$ such that $V \subseteq \hat{U} \cap A_m$ and $N_V \geq \epsilon N_{\hat{U}} \geq \epsilon N_U$. The definition of $\hat{U}$ and $(U.1)$ ensure that $\varrho_{\hat{U},U'} = 0$ and by Proposition 3.3

$$\boldsymbol{E}_f |\hat{f}_{\hat{U}} - f(x^0)|^r \leq (C\sigma^2 \epsilon^{-1} N_U^{-1} \log n)^{r/2}$$

as required.

4. The case of an equidistant design

In this section we specify our procedure and results to the case of a regular equidistant design in the unit square $[0,1]^2$.

Suppose therefore that we are given n design points $X_1, \ldots, X_n$ 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 we 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 above, we consider the problem of estimating the image value at a point x^0 by observations $Y_1, \ldots, Y_n$ described by the model equation (1.1). 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 compare it with the existing in the literature.

4.1. An example of the set of windows

Shortly our set $\mathcal{U}$ can be characterized as a set of parallelograms containing x^0 and such that two sides of each parallelogram are vertically or horizontally oriented and all its vertices are grid points.

To present a formal description, we need some more notation. Our procedure involves two external parameters K and D. The integer K enters in condition $(U.2)$ and in Theorem 3.1 and we have to ensure in our construction that the number of design points in the intersection of every two windows of the constructed family is at least K. The parameter D controls the number of considered orientations.

Denote by $\mathcal{P}_D$ the set of all pairs of integers with absolute values at most D and such that the fraction p/q is unreducible,

$$\mathcal{P}_D = \{(p,q) : |p|, |q| \le D,\ p/q \text{ is unreducible}\}.$$

To get a unique representation of elements from $\mathcal{P}_D$ by a pair (p,q), we additionally suppose that $\max\{|p|,|q|\} = \max\{p,q\}$. Obviously the number P_Q^* of elements in $\mathcal{P}_D$ is at most D^2,

$$P_D^* = \#\mathcal{P}_D \le D^2.$$

Each element (p,q) from $\mathcal{P}_D$ determines a line passing through x^0 and another grid point, namely the point with coordinates $(x_1^0 + p\delta, x_2^0 + q\delta)$. If S_D is the square with the center at x^0 and with the side length $2\delta D$, then $\mathcal{P}_D$ can be identified with the set of lines passing through x^0 and another grid point in S_D. Motivated by this reasoning, we will call each element from $\mathcal{P}_D$ *an orientation.* For each $(p,q) \in \mathcal{P}_D$, introduce two vectors $\beta_{p,q}$ and $\beta'_{p,q}$ by

$$\beta_{p,q} = (Kp\delta, Kq\delta), \qquad \beta'_{p,q} = \begin{cases} (\delta, 0) & |p| \le q, \\ (0, \delta) & p > |q|. \end{cases} \tag{4.1}$$

Note that the vector $\beta_{p,q}$ passes through $K+1$ grid points and the vector $\beta'_{p,q}$ connects two neighbor horizontal or vertical points.

Now for each orientation (p,q) from $\mathcal{P}_D$, we introduce a family $\Pi_{p,q}$ of (p,q)-*oriented* parallelograms. First, given two integers l_1 and l_2, denote by $I_{p,q}(l_1,l_2)$ the interval connecting two points $x^0 + l_1\beta_{p,q}$ and $x^0 + l_2\beta_{p,q}$. This interval has the length $|l_1 - l_2| r\delta\sqrt{p^2+q^2}$. Next, given two more integers w_1 and w_2, define a parallelogram $P_{p,q}(l_1,l_2,w_1,w_2)$ whose two sides are the shifts of the interval $I_{p,q}(l_1,l_2)$ in the direction $\beta'_{p,q}$ by w_1 (respectively w_2) grid points. In other words, this parallelogram has its vertices at

$$\begin{array}{ll} x^0 + l_1\beta + w_1\beta' & \qquad x^0 + l_1\beta + w_2\beta' \\ x^0 + l_2\beta + w_1\beta' & \qquad x^0 + l_2\beta + w_2\beta' \end{array}$$

all of them are grid points. Clearly, the other two sides of this parallelogram are either horizontal or vertical in accordance with $\beta'_{p,q}$ and they are of the length $|w_1 - w_2|\delta$. This value will be called the width of the parallelogram.

Now we define the set $\mathcal{U}$ of considered windows which we take in the form of parallelograms. Let $\mathcal{L}_+$ be a diadic geometrical grid defined by

$$\mathcal{L}_+ = \{l = 1, 2, 4, 8, \ldots : l \le \tfrac{\sqrt{n}}{K}\}.$$

Let also $\mathcal{L}_- = -\mathcal{L}_+ = \{-l : l \in \mathcal{L}_+\}$, and

$$\mathcal{L} = \mathcal{L}_+ \cup \{0\} \cup \mathcal{L}_- = \{l = 0, \pm 2^j, j = 0, 1, 2, \ldots : |l|K \le \sqrt{n}\}.$$

Below we assume that the numbers l_1 and l_2 take values in the sets $\mathcal{L}_+$ and $\mathcal{L}_-$ respectively. Similarly we define

$$\begin{aligned} \mathcal{N}_+ &= \{w = 0, 1, 2, 4, 8, \ldots : w \le \sqrt{n}\}, \\ \mathcal{N}_- &= -\mathcal{N}_+, \\ \mathcal{N} &= \mathcal{N}_+ \cup \mathcal{N}_- = \{w = 0, \pm 2^j, j = 0, 1, 2, \ldots : |w| \le \sqrt{n}\} \end{aligned}$$

and suppose that $w_1 \in \mathcal{N}_+$, $w_2 \in \mathcal{N}_-$ and $w_1 + |w_2| \geq K$. Therefore,

$$\begin{aligned} \mathcal{U} &= \{P_{p,q}(l_1, l_2, w_1, w_2) : \\ &\quad (p,q) \in \mathcal{P}_D,\, l_1, -l_2 \in \mathcal{L}_+,\, w_1, -w_2 \in \mathcal{N}_+,\, w_1 + |w_2| \geq K\}. \end{aligned} \tag{4.2}$$

Obviously, the minimal windows in this family are $P_{p,q}(1,-1,K,0)$, $P_{p,q}(1,-1,K-1,-1)$, ... $P_{p,q}(1,-1,0,-K)$. The cardinality of $\mathcal{U}$ can be roughly estimated in the following way,

$$\#\mathcal{U} \leq \#\mathcal{P}_D(\#\mathcal{L}_+)^2(\#\mathcal{N}_+)^2 \leq P_D^* \left(\log_2\left(\tfrac{\sqrt{n}}{K}\right)\log_2(\sqrt{n})\right)^2.$$

Define also another set of windows

$$\mathcal{U}' = \{P_{p,q}(l_1, l_2, w_1, w_2) : (p,q) \in \mathcal{P}_D,\, l_1 > l_2 \in \mathcal{L},\, w_1 \geq w_2 \in \mathcal{N}\}.$$

This set is larger than $\mathcal{U}$ since it is not required for windows $U \in \mathcal{U}'$ to contain x^0 and there is no constraints on the minimal width: even zero width is allowed; in this case the window is simply a shift of the interval $I_{p,q}(l_1, l_2)$.

Note that by construction, every window $U = P_{p,q}(l_1, l_2, w_1, w_2)$ from $\mathcal{U}'$ contains at least $K(l_1 - l_2)(w_1 - w_2 + 1)$ grid points.

Define now for a given U from $\mathcal{U}$ the testing set $\mathcal{V}(U)$ as the set of all parallelograms from $\mathcal{U}'$ which are contained in U,

$$\mathcal{V}(U) = \{V \in \mathcal{U}' : V \subset U\}. \tag{4.3}$$

Clearly

$$\#\mathcal{V}(U) \leq \#\mathcal{P}_D\,(\#\mathcal{L}/2)^2(\#\mathcal{N}/2)^2 \leq 4P_D^* \left(\log_2\left(\tfrac{\sqrt{n}}{K}\right)\log_2(\sqrt{n})\right)^2$$

and the constraint in (2.3) is satisfied with say $\alpha = 2$.

Condition $(U.1)$ is fulfilled by construction for the above defined set $\mathcal{U}$. Now we are checking $(U.2)$.

Lemma 4.1. *The sets $\mathcal{U}$ and $\mathcal{V}(U)$ fulfill $(U.2)$.*

Proof. Let U and U' be two different windows from $\mathcal{U}$ with orientations (p,q) and (p',q') respectively. Obviously it suffices to consider minimal windows with these orientations. Using the fact that the grid is invariant under shifts by vectors $\beta'_{p,q}$ we can reduce the problem to the case of two parallelograms of the form $P_{p,q}(1,-1,K,0)$ and $P_{p',q'}(1,-1,w_1,w_2)$ with $w_1 - w_2 = K$. Moreover, it is easy to see that the smallest intersection corresponds to the case when $w_1 = 0$, $w_2 = -K$ and p/q is close to p'/q'. But even in this situation, if, for instance, $p^2 + q^2 \leq {p'}^2 + {q'}^2$, this intersection contains an interval of the form $[x^0, x^0 \pm \beta_{p,q}]$ which passes through $K+1$ grid points. □

4.2. The accuracy of estimation for the case of a boundary fragment

Now we are going to apply Theorem 3.1 to the case of the regular design.

Assume that the point of interest x^0 belongs to a region A with the edge G. To apply Theorem 3.1 to the considered situation we have to show that there is a window U from $\mathcal{U}$ which is contained in A. First we note that our construction of the family $\mathcal{U}$ is oriented to the case when the point x^0 lies near the boundary of the region and this boundary is regular in the sense that it can be well approximated by a straight line in some vicinity of the point x^0.

Let $d(x,G)$ denote the distance from a point x to the edge G,

$$d(x,G) = \inf_{x' \in G} |x - x'|$$

where $|x-x'|$ means the usual Euclidean distance between x and x'. The fact that x^0 is near the boundary means that the distance $d(x,G)$ is small.

Next, let us fix for a moment an orientation $(p,q) \in \mathcal{P}_D$ and integers $l_1 \in \mathcal{L}_+$ and $l_2 \in \mathcal{L}_-$. Given b, denote by $I_{p,q}(l_1,l_2;b)$ the shift of the interval $I_{p,q}(l_1,l_2)$ by the vector $b\beta'_{p,q}$. In particular, $I_{p,q}(l_1,l_2;0)$ coincides with $I_{p,q}(l_1,l_2)$. All such intervals form the band (or the fragment) which is horizontally or vertically oriented in accordance with $\beta'_{p,q}$. Set

$$\Delta_{p,q}(l_1,l_2) = \inf_b \sup_{x \in I_{p,q}(l_1,l_2;b)} d(x,G). \tag{4.4}$$

A small value of $\Delta_{p,q}(l_1,l_2)$ means that the edge G can be well approximated by a straight line with an orientation (p,q) from $\mathcal{P}_D$ in a neighborhood of x^0.

Proposition 4.1. *If $(p,q) \in \mathcal{P}_D$ is such that $d(x^0,G) > 2\Delta_{p,q}(l_1,l_2)$, then the interval $I_{p,q}(l_1,l_2)$ does not intersect G.*

Proof. Let b be such that

$$d(x,G) \le \Delta_{p,q}(l_1,l_2)$$

for all $x \in I_{p,q}(l_1,l_2;b)$.

Denote for every point x by $x(b)$ the point $x + b\beta'_{p,q}$. If $x \in I_{p,q}(l_1,l_2)$, then $x(b) \in I_{p,q}(l_1,l_2;b)$. By the triangle inequality one has for all $x \in I_{p,q}(l_1,l_2)$

$$d(x,G) \ge |x - x(b)| - d(x(b),G) = b - d(x(b),G) \ge b - \Delta_{p,q}(l_1,l_2).$$

At the same time, again by the triangle inequality

$$d(x^0,G) \le |x^0 - x^0(b)| + d(x^0(b),G) = b + d(x^0(b),G)$$

which implies along with the theorem condition that $b \ge d(x^0,G) - d(x^0(b),G) > \Delta_{p,q}(l_1,l_2)$. Combining the above estimates we get $d(x,G) > 0$ and the assertion follows. □

Now we optimize the choice of the orientation (p,q) over $\mathcal{P}_D$ and define

$$\Delta_D(l_1,l_2) = \inf_{(p,q)\in\mathcal{P}_D} \Delta_{p,q}(l_1,l_2). \tag{4.5}$$

The above assertion means that the condition $d(x^0,G) > 2\Delta_D(l_1,l_2)$ ensures the existence of a window of zero width which lies in A. To guarantee the existence of a window of the width $r\delta$, we assume in addition that x^0 is an internal point of the unit square $[0,1]^2$ and that the region A has, at least locally in some vicinity of the point x_0, the structure of a smooth boundary fragment. More precisely, let two integers l_1 from $\mathcal{L}_+$ and l_2 from $\mathcal{L}_-$ be fixed and let the orientation (p,q) minimize the value $\Delta_{p,q}(l_1,l_2)$ or, equivalently, $\Delta_{p,q}(l_1,l_2) = \Delta_D(l_1,l_2)$. Suppose also without loss of generality that $|p| > |q|$ and hence, the vector $\beta'_{p,q}$ is vertical, and consider the vertical band (the fragment) which is formed by all the shifts

$I_{p,q}(l_1, l_2, b)$ of the interval $I_{p,q}(l_1, l_2)$ in the vertical direction. We suppose that the curve G can be parametrized within this fragment by the equation $x_2 = g(x_1)$ with some continuous univariate function g. If $|p| \le |q|$, then the horizontal fragment should be considered instead of the horizontal one and the other parametrization of the curve G in the form $x_1 = g_1(x_2)$ is to be used.

Now we apply Theorem 3.1 to the case of a smooth boundary fragment.

Theorem 4.1. *Let the image function $f(x)$ be of the form*

$$f(x) = \begin{cases} a, & x \in A' = \{x_2 > g(x_1)\}, \\ 0, & x \in A = \{x_2 \le g(x_1)\}, \end{cases}$$

and let the point x^0 belong to A with the distance $d(x^0, G)$ to the edge G described by the equation $x_2 = g(x_1)$. Let K satisfy the condition

$$K \ge 4\sigma^2 a^{-2} \left(\sqrt{2(\alpha + r)} + \sqrt{2r} \right)^2 \log n.$$

Also we assume that condition $(U.3)$ with some $\epsilon > 0$ is fulfilled for the region A and for the above defined set of windows $\mathcal{U}$. If there are two integers $l_1 \in \mathcal{L}_+$ and $l_2 \in \mathcal{L}_-$ such that

$$d(x^0, G) > 2\Delta_D(l_1, l_2),$$

and if $\min\{x_1^0, x_2^0\} \ge \varkappa$ with some $\varkappa > 0$, then

$$\boldsymbol{E}_f |\hat{f}(x^0)|^r \le \left(C \frac{\sigma^2 \log n}{\epsilon N^*} \right)^{r/2}$$

where $N^* \ge \varkappa n^{1/2}(l_1 + |l_2|)$.

Proof. The statement of this theorem is a direct application of Theorem 3.1. We only need to verify that $N^* \ge \varkappa \delta(l_1 + |l_2|)$. Let (p, q) be such that $\Delta_D(l_1, l_2) = \Delta_{p,q}(l_1, l_2)$. Then Proposition 4.1 ensures that the interval $I_{p,q}(l_1, l_2)$ does not intersect the boundary curve G and the structure of the image yields that every window of the form $P_{p,q}(l_1, l_2, 0, w_2)$ is in the region A. Making use of the condition $x_2^0 \ge \varkappa$, we can easily check that the number of grid points in such a window with $|w_2| \approx \varkappa \sqrt{n}$ is about $\varkappa n^{1/2}(l_1 + |l_2|)$ as required. □

4.3. Edge estimation

Now we discuss the problem of edge estimation. Note that the above procedure is assigned for estimating the image function f and there is no any edge estimation subroutine. However, in the case of an image with the structure of a boundary fragment, the procedure estimates the value $f(x^0)$ consistently and even with the rate $n^{-1/4}$ if the point x^0 is bounded away from the edge, more precisely, if $d(x^0, G) > 2\Delta_D(l_1, l_2)$ with some l_1, l_2. The minimal distance between the point x^0 and edge G which is sufficient for consistent estimation of $f(x^0)$ can be regarded as the accuracy of edge estimation. The result of Theorem 4.1 gives some information about the accuracy provided by our procedure. Now we aim to transfer this result into a more readable form and then to compare with the earlier results on edge estimation.

The problem of the edge estimation was considered in details in Korostelev and Tsybakov (1994). They have shown that the rate of edge estimation critically depends on the smoothness properties of the function g defining the edge. In particular, if g belongs to a Hölder class $\Sigma(\gamma, L)$, then the accuracy of edge estimation, being measured in the Hausdorff metric, is $(n/\log n)^{-\gamma/(\gamma+1)}$. We are going to show that our procedure provides essentially with the same rate. Note however, that Korostelev and Tsybakov (1994) stated their results under a random or jittered design, see p.92. Under the regular design, the rate of edge estimation is equal to the grid step $\delta = n^{-1/2}$, Korostelev and Tsybakov (1994, p.99). This can be explained, for example, by the fact that if the edge G is a straight horizontal line, then for any shift of this line within an interval between two neighbor grid lines, we have the same distribution on the space of observations. To get a better rate of edge estimation, it is required to 'randomize' the design in some way.

We proceed under the regular design but we estimate the value of the image at a grid point. We will see that this fact also allows us to get a better accuracy of estimation. We discuss only the situation with $\gamma \in [1,2]$. To cover the case when $\gamma > 2$, the construction of the class $\mathcal{U}$ is to be modified.

In view of the result of Theorem 4.1, we have to estimate the value of $\Delta_D(l_1, l_2)$ which characterizes the quality of approximation of the edge G by a straight line with an orientation from $\mathcal{P}_D$ in a vicinity of the point x^0. The size of this vicinity is determined by the values l_1 and l_2 and obviously it is sufficient to consider the case of the smallest vicinity with $l_1 = |l_2| = 1$.

Let the function $g(x_1)$ describe the edge G. We suppose that this function belongs to a Hölder ball $\Sigma(\gamma, L)$ with some $\gamma \in (1,2]$. This implies for all $h > 0$

$$\sup_{|t|\le h} |g(x_1^0 + t) - g(x^0) - g'(x^0)t| \le Lh^\gamma.$$

Here g' stands for the derivative of g. Denote also $c = g'(x_1^0)$.

Lemma 4.2. *For every integer* $D' \le D$ *,*

$$\Delta_D(1,-1) \le L(K\delta D')^\gamma + K\delta\rho(c; D')$$

where

$$\rho(c; D') = \inf_{(p,q)\in\mathcal{P}_{D'}} |cp - q|.$$

Proof. Let us fix some $D' \le D$ and let (p,q) from $\mathcal{P}_{D'}$ be such that $|cp - q| = \rho(c; D')$. The definition of $\mathcal{P}_{D'}$ and the assumption $|c| \le 1$ imply $|q| \le p \le D'$.

Set $b = g(x_1^0)$. We will estimate the value $d(x,G)$ for a point $x = (x_1, x_2)$ from the interval $I_{p,q}(1,-1;b)$. By definition, $|x_1 - x_1^0| \le h \equiv K\delta p$ and $x_2 = g(x_1^0) + (x_1 - x_1^0)q/p)$. Denote also by x_G the point on the boundary curve G with the same first coordinate as x, i.e. $x_G = (x_1, g(x_1))$. We obtain

$$\begin{aligned} d(x,G) &\le |x - x_G| = |g(x_1) - g(x_1^0) + (x_1 - x_1^0)q/p| \\ &\le |g(x_1) - g(x_1^0) + (x_1 - x_1^0)c| + h|c - q/p| \le Lh^\gamma + \rho(c; D')h/p. \end{aligned}$$

Now, using $h = K\delta p$ along with $p \le D'$, we get

$$d(x,G) \le L(K\delta D')^\gamma + K\delta\rho(c; D')$$

and the assertion follows. □

The next statement shows that the value $\rho(c;D')$ can be estimated by $1/D'$.

Lemma 4.3. *For every integer* D' *and for all* $c \in [-1,1]$

$$\rho(c,D') \le 1/Q.$$

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

$$\frac{q_k}{p_k} = \cfrac{1}{a_1 + \cfrac{1}{a_2 + \cdots \cfrac{1}{a_{k-1} + \cfrac{1}{a_k}}}}.$$

This approximation has the following properties, Khintchine (1949, Section 3,4):

$$\left| c - \frac{q_k}{p_k} \right| \le \frac{1}{p_k p_{k+1}}, \tag{4.6}$$

$$p_k \ge 2^{(k-1)/2}, \qquad k \ge 0. \tag{4.7}$$

Given an integer D', denote

$$k^* = \max\{k : p_k \le D'\}$$

so that $p_{k^*+1} > D'$.

By (4.6), $|cp_{k^*} - q_{k^*}| \le 1/p_{k^*+1} < 1/D'$ and the assertion follows. □

Putting together the last two results, we get for an arbitrary orientation c of the edge near x^0 the following bound

$$\Delta_D(1,-1) \le L(K\delta D')^\gamma + K\delta/D'.$$

An optimization of the choice of D' leads to the following

Proposition 4.2. *If* $D \ge (n/K^2)^{\frac{\gamma-1}{2(\gamma+1)}}$, *then*

$$\Delta_D(1,-1) \le (L+1)(n/K^2)^{-\frac{\gamma}{\gamma+1}}.$$

Proof. We simply plug $D' = (n/K^2)^{\frac{\gamma-1}{2(\gamma+1)}}$ in the above bound. □

4.4. Rate-optimality

We have seen that the accuracy of edge estimation provided by the proposed method coincides in order with the optimal rate $\psi_n = (n/\log n)^{-\gamma/(\gamma+1)}$ of edge estimation. But this coincidence is at this stage only formal since the above presented results are concerned with the estimation at grid points under the regular grid and the optimal results from Korostelev and Tsybakov (1994) are stated for a random design. Therefore, we cannot refer to them when studying the problem of optimal estimation under the regular design. The next assertion shows that the accuracy ψ_n which is attained by our procedure, cannot be essentially improved by any estimation method.

Let some grid point x^0 be fixed and let an image have the structure of a smooth boundary fragment (at least locally near the point x^0) with an edge G determined by a function $g = g(x_1)$ from the Hölder ball $\Sigma(\gamma, L)$ with $\gamma \in (1, 2]$. The function g determines the image function f_g with $f_g(x) = \mathbf{1}(x_2 \geq g(x_1))$ for $x = (x_1, x_2)$. We stand also $G = G_g$ for the corresponding edge i.e. $G = \{x : x_2 = g(x_1)\}$.

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

Theorem 4.2. *Let K, D be integers and $(p, q) \in \mathcal{P}_D$. Then there are two functions g_0 and g_1 from $\Sigma(\gamma, 1)$ such that $g_0'(x_1^0) = q/p$, $g_1'(x_1^0) = q/p$,*

$$d(x^0, G_{g_k}) \geq \min\{\delta/p, (\delta K p)^\gamma\}, \qquad k = 0, 1, \tag{4.8}$$

and for any estimator $\tilde{f}$

$$\max\left\{E_{g_0}|\tilde{f}(x^0) - f_{g_0}(x^0)|, E_{g_1}|\tilde{f}(x^0) - f_{g_1}(x^0)|\right\} \geq c > 0, \tag{4.9}$$

where c is a positive constant depending on K only.

We defer the proof of the theorem to the Appendix.

Discussion 4.1. The maximal value of the right hand-side of (4.8) is attained for $p = p^* = \delta^{-(\gamma-1)/(\gamma+1)} K^{-\gamma/(\gamma+1)}$ and this value is of order $n^{-\gamma/(\gamma+1)}$. Therefore, the rate of estimation of $f(x^0)$ cannot be better than $n^{-\gamma/(\gamma+1)}$ and our procedure is at least near rate-optimal.

4.5. Computational discussion

At the conclusion we shortly discuss computational efforts corresponding to the above procedure. The procedure is specified to estimating the value $f(x^0)$ at a given grid point x^0 but now we assume that this procedure is carried out for each grid points of the unit square. To reduce the required number of operations, it is reasonable to make a preprocessing for each orientation $(p, q) \in \mathcal{P}_D$. This means that we reorder all the grid points from this square due to their projections on the vectors $\beta_{p,q}$ and $\beta'_{p,q}$. Suppose for simplicity that $|p| > |q|$ and therefore $\beta'_{p,q}$ is the horizontal vector. For each grid point $x = (x_1, x_2)$ set

$$\begin{aligned} z_{p,q}(x) &= p x_1 + q x_2 \\ z'_{p,q}(x) &= (x, \beta'_{p,q}) = x_1 \end{aligned}$$

and define two values $N_{p,q}(x)$ and $F_{p,q}(x)$ such that $N_{p,q}(x)$ is the number of points x' of the grid with smaller coordinates $z_{p,q}$ and $z'_{p,q}$,

$$N_{p,q}(x) = \sum_{i=1}^{n} \mathbf{1}\left(z_{p,q}(X_i) \leq z_{p,q}(x),\, z'_{p,q}(X_i) \leq z'_{p,q}(x)\right),$$

and $F_{p,q}(x)$ is the sum of the observations Y_i over the same index set,

$$F_{p,q}(x) = \sum_{i=1}^{n} Y_i \mathbf{1}\left(z_{p,q}(X_i) \le z_{p,q}(x),\, z'_{p,q}(X_i) \le z'_{p,q}(x)\right).$$

The expressions for (p,q) with $|p| \le |q|$ differ only in the place $(x, \beta'_{p,q}) = x_2$.

It is not difficult to see that the calculation of the arrays $N_{p,q}(X_i)$ and $F_{p,q}(X_i)$ for $i = 1, \dots, n$ requires of order n operations. Therefore, the first step (the initialization) of the algorithm requires const. nP_D^* operations where $P_D^* = \#\mathcal{P}_D$ is the total number of considered orientations from $\mathcal{P}_D$.

Having done this preprocessing, the number N_U of design points in each parallelogram $U = P_{p,q}(l_1, l_2, w_1, w_2)$ and the sum S_U of observations Y_i over this parallelograms can be calculated by a finite number of operations:

$$\begin{aligned}
N_U &= N_{p,q}(x^0 + l_1\beta + w_1\beta') - N_{p,q}(x^0 + l_1\beta + w_2\beta') \\
&\quad - N_{p,q}(x^0 + l_2\beta + w_1\beta') + N_{p,q}(x^0 + l_2\beta + w_2\beta'), \\
S_U &= F_{p,q}(x^0 + l_1\beta + w_1\beta') - F_{p,q}(x^0 + l_1\beta + w_2\beta') \\
&\quad - F_{p,q}(x^0 + l_2\beta + w_1\beta') + F_{p,q}(x^0 + l_2\beta + w_2\beta').
\end{aligned}$$

Therefore, the total number of operations for estimating the whole image can be very roughly bounded by

$$\text{const. } n\#(\mathcal{U})\#(\mathcal{V}(U)) \le \text{const. } n(P_D^*)^2 \left(\log_2\left(\frac{\sqrt{n}}{K}\right)\log_2(\sqrt{n})\right)^4.$$

We see that the computational difficulty of the algorithm is of a smaller order than $nD^4(\log n)^8$ which is still feasible for realization by modern computers.

Appendix. Proof of Theorem 4.2

Different methods for obtaining the lower bounds in edge estimation are presented in Korostelev and Tsybakov (1994). We cannot directly apply these methods since they are developed for a random design and we operate with the regular design. But we follow the same route and thus we present here 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 $(p,q) \in \mathcal{P}_D$. Set

$$h = \min\{pK\delta, (\delta/p)^{1/\gamma}\},$$

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 by $[-1,1]$;

(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

$$\begin{aligned} g_0(x_1) &= x_1 q/p - \phi_h(0)/2 \\ g_1(x_1) &= x_1 q/p + \phi_h(0)/2 - \phi_h(x_1 - x_1^0). \end{aligned}$$

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

$$\begin{aligned} A_k &= \{x = (x_1, x_2) : x_2 \le g_k(x_1)\}, \\ G_k &= \{x = (x_1, x_2) : x_2 = g_k(x_1)\}, \qquad k = 0, 1. \end{aligned}$$

Set also

$$B = A_1 \backslash A_0 = \{x = (x_1, x_2) : g_0(x_1) < x_2 \le g_1(x_1)\}.$$

Below we make use of the following technical assertion.

Lemma 4.4. *The following assertions hold*

(*i*) $g_0, g_1 \in \Sigma(\gamma, 1)$ *and* $g_0'(x_1^0) = g_1'(x_1^0) = q/p$ *;*

(*ii*) $d(x^0, G_k) \ge \varkappa h^\gamma$, $k = 0, 1$, *for some* $\varkappa > 0$ *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 us fix on the line $x_2 = x_1 q/p$ the open interval corresponding to $x_1 \in (x_1^0 - h, x_1^0 + h)$. Since $h \le pK\delta$, then this interval passes at most through $2K - 1$ design points. We intend to show that there is no other design points in B that implies the assertion in view of property (c) of ϕ.

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

$$|q'\delta - x_1 q/p| > |\phi_h(x_1 - x_1^0) - \phi_h(0)/2|.$$

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

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

and hence $|q'\delta - x_1 q/p| \ge \delta/p$. 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 $h^\gamma \le \delta/p$ and (iii) follows. □

Denote $f_k(x) = \mathbf{1}(x \notin A_k) = \mathbf{1}(x_2 > g_k(x_1))$ for $x = (x_1, x_2)$. Note that $f_0(x^0) = 0$ and $f_1(x^0) = 1$ and we have to show that for any estimator $\tilde{f}(x^0)$

$$\max\left\{E_0|\tilde{f}(x^0)|, E_1|\tilde{f}(x^0) - 1|\right\} \ge c > 0, \tag{4.10}$$

where E_k stands for E_{g_k}, $k = 0, 1$, and c is some fixed positive constant.

Let $Z = dP_1/dP_0$. 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

$$\max\{E_0|\tilde{f}(x^0)|, E_1|\tilde{f}(x^0) - 1|\} \ge E_0 \mathbf{1}(Z \ge 1) = P_0(Z \ge 1).$$

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 ζ is a standard normal random variable and Lemma 4.4, (ii) and (iii), implies that

$$\begin{aligned} P_0(Z > 1) &= P_0\left(\exp\left\{\sigma^{-1}\sqrt{N}\zeta - \sigma^{-2}N/2\right\} > 1\right) \\ &= P_0\left(\zeta > \sigma^{-1}\sqrt{N}/2\right) \\ &\le P_0\left(\zeta > \sigma^{-1}\sqrt{K/2}\right) \\ &= 1 - \Phi\left(\sigma^{-1}\sqrt{K/2}\right) > 0 \end{aligned}$$

where Φ is the Laplace function and the required assertion follows.

References

[1] Blake, A. and Zisserman, A. (1987). *Visual reconstruction.* MIT Press, Cambridge, MA.

[2] Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images. *IEEE Trans. on PAMI* **6** 721–741.

[3] Girard, D. (1990). From template matching to optimal approximation by piecewise smooth curves. In: *Curves and Surfaces in Computer Vision and Graphics*, Proc. Conf. Santa Clara, Calif., 1990, 174–182.

[4] Grenander, U. (1976, 1978, 1981). Lectures in Pattern Theory, vol. 1–3. Springer, New York. (Applied Math. Sciences, vol. 18, 24, 33).

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

[6] Huang, J.S. and Tseng, D.H. (1988). Statistical Theory of edge detection. *Graphics and Image Processing* **43** 337–346.

[7] Khintchine (1949) *Continued fractions.* ITTL, Moscow.

[8] Korostelev, A. and Tsybakov, A. (1993) *Minimax Theory of Image Reconstruction.* Springer Verlag, New York–Heidelberg–Berlin.

[9] Lee, J.S. (1983). Digital image smoothing and the sigma-filter. *Computer Vision, Graphics and Image Processing* **24** 255–269.

[10] Lepskii, O.V. (1990). One problem of adaptive estimation in Gaussian white noise. *Theory Probab. Appl.* **35**, no. 3, 459–470.

[11] Lepskii, O.V. (1992). Asymptotic minimax adaptive estimation. 2. Statistical model without optimal adaptation. Adaptive estimators. *Theory Probab. Appl.* **37**, no. 3, 468–481.

[12] Lepski, O., Mammen, E. and Spokoiny, V. (1997). Ideal spatial adaptation to inhomogeneous smoothness: an approach based on kernel estimates with variable bandwidth selection. *Annals of Statistics*, to appear.

[13] Lepski, O. and Spokoiny, V. (1997). Optimal pointwise adaptive methods in nonparametric estimation. *Annals of Statistics*, to appear.

[14] Marr, D. (1982). *Vision.* Freeman and Co., San Francisco.

[15] Müller, H.G. and Song, K.S. (1994). Maximum estimation of multidimensional boundaries. *J. Multivariate Anal.* **50**, no.2, 265–281.

[16] Mumford, D. and Shah, J. (1989). Optimal approximation by piecewise smooth functions and associated variational problem. *Comm. Pure Appl. Math.* **42** 577–685.

[17] Petrov, V.V. (1975) *Sums of Independent Random Variables.* Springer, New York.

[18] Pratt, W.K. (1978). *Digital Image Processing.* J.Wiley, New York.

[19] Ripley, B.D. (1988). *Statistical Inference for Spatial Processes.* Cambridge Univ. Press.

[20] Rosenfeld, A. and Kak, A.C. (1982). *Digital Picture Processing.* Academic Press, London.

[21] Shiau, J.J. Wahba, G. and Johnson, D.R. (1986). Partial spline models for the inclusion of tropopause and frontal boundary information in otherwise two- or three-dimensional objective analysis. *J. Atmospheric and Oceanic Technology* **3** 714–725.

[22] Spokoiny, V. (1996). Estimation of a function with discontinuities via local polynomial fit with an adaptive window choice. Preprint **291**, Weierstrass Institute, Berlin. (submitted to *Ann. Statist.*)

[23] Titterington, D.M. (1985). Common structure of smoothing technique in statistics. *Intern. Stat. Review* **53** 141–170.

Weierstrass Institute for Applied Analysis and Stochastics, Mohrenstr. 39, 10117 Berlin, Germany., E-mail: spokoiny@wias-berlin.de