MINIMAX LINEWISE ALGORITHM FOR IMAGE RECONSTRUCTION

by

A.P. Korostelev* **and A.B. Tsybakov****

November 1992

Abstract

We study the problem of estimating the edges in noisy images by linewise procedures. We show that the straightforward estimation method (naïve linewise procedure) does not attain the asymptotically minimax rate of accuracy as the number of observations tends to ∞. We propose the modified linewise procedure which has the asymptotically minimax rate.

AMS 1980 Subject Classification: 62 G 05

Key words and phrases: image estimation, edge estimation, minimax rate of convergence, linewise algorithm, nonparametric regression, change-point problem.

* Institute for System Analysis, Academy of Sciences of Russia Prospekt 60-let Oktyabrya 9, Moscow 117312, Russia.

** CORE, Université Catholique de Louvain, Louvain-la-Neuve, Belgium and Institute for Problems of Information Transmission, Russian Academy of Sciences, Russia.

1. Introduction

There exists an extensive literature on image analysis (see e.g. the books of Pratt (1978), Rosenfeld and Kak (1982), Marr (1982), Blake and Zisserman (1987)). One of the important problems in image analysis is reconstruction of pictures from noisy data, or image estimation. There are two special features related to this problem. First, the data arrays are 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. Therefore, it is important to find the boundaries (edges) of such regions. This leads to edge detection or edge estimation problems.

A large variety of methods has been proposed for solving the image and edge estimation problems in different contexts. An overview of them can be found in the above mentioned books. The most popular methods are based on certain parametric image models with large number of parameters and on Bayesian approach to estimation (Haralick (1980), Geman and Geman (1984), Ripley (1988) among others). The methods of edge detection, however, often have the local character and do not assume any underlying parametric model (Marr (1982), Nagao and Matsuyama (1979), Huang and Tseng (1988)). Another nonparametric approach to the construction of image estimators is related to penalizing and regularization techniques (Titterington (1985), Torre and Poggio (1986), Shiau, Wahba and Johnson (1986), Mumford and Shah (1989), Girard (1990)).

In Tsybakov (1989), Korostelev (1991), Korostelev and Tsybakov (1992) the minimax approach to image and edge estimation was developed which makes possible to compare different image and edge estimators on the common scale. This approach is similar to the minimax theory of nonparametric regression and density function estimation (Ibragimov and Khasminskii (1979), Bretagnolle and Huber (1979), Stone (1980, 1982)).

It has been shown in Korostelev and Tsybakov (1992) that the best possible rates of convergence (minimax rates) in various image estimation problems are attained by some estimators which are, however, complicated from the computational viewpoint. On the other hand, there exist many simple data processing

algorithms in image analysis which do apply in practice. We study here one of them: the linewise algorithm, and show that it does not have the minimax rate of convergence in its naïve form. Next we propose a modified linewise algorithm which attains the best possible (minimax) rate. The study of linewise procedures reveals in the most obvious form the link between image estimation, change-point problem and nonparametric regression.

2. Edge Estimation Problem

Consider a simple two-dimensional image model. Assume that G is a set that belongs to the unit square $K = [0,1] \times [0,1]$. Let n_1 be an integer.

Let the sample of $n = n_1^2$ observations is given,

$$Y_{ij} = I(X_{ij} \in G) + \xi_{ij}, \qquad i,j = 1,\dots,n_1, \tag{2.1}$$

where $X_{ij} \in K$ are the design points, ξ_{ij} are i.i.d. $(0,\sigma^2)$-Gaussian random errors and $I(.)$ denotes the indicator function.

We assume that G is of the form

$$G = \{x = (x_1, x_2) \in K : \quad 0 \le x_1 \le 1, \quad 0 \le x_2 \le g(x_1)\}$$

where $g \in \Sigma(\gamma, L, h), \gamma > 1, L > 0, 0 < h < 1/2$, and $\Sigma(\gamma, L, h)$ denotes the class of functions $g(x_1), h < g(x_1) < 1-h$, defined on the interval $[0,1]$ and having the Hölder kth derivative:

$$\mid g^{(k)}(x_1) - g^{(k)}(x_1') \mid \le L \mid x_1 - x_1' \mid^{\gamma-k}, x_1, x_1' \in [0,1],$$

where k is the maximal integer such that $k < \gamma$. Denote by $\mathcal{G}(\gamma, L, h)$ the class of all domains G satisfying these conditions.

Define the design points

$$X_{ij} = (x_{ij}^{(1)}, x_{ij}^{(2)}) = (i/n_1, j/n_1 - \eta_i), \quad i,j = 1,\dots,n_1, \tag{2.2}$$

where η_i are independent random variables uniformly distributed in $[0, 1/n_1]$. The design (2.2) is composed of random shifts of the regular grid; these shifts are independent for each line $x_1 = i/n_1, \quad i = 1,\dots,n_1$.We assume everywhere that η_i's are independent of $\xi_{ij}, i,j = 1,\dots,n_1$.

A set $G \in \mathcal{G}(\gamma, L, h)$ is called *image*. An estimator $\hat{G}_n$ of image G is assumed to be a closed subset of K measurable with respect to observations $(X_{ij}, Y_{ij}), \quad i, j = 1, \ldots, n_1$. For an arbitrary estimator $\hat{G}_n$ introduce the maximal risk

$$r_n\left(\hat{G}_n, \psi_n\right) = \sup_{G \in \mathcal{G}(\gamma, L, h)} E_G\left[\psi_n^{-1} d(G, \hat{G}_n)\right]$$

where E_G is the expectation with respect to the distribution of observations $(X_{ij}, Y_{ij}), \quad i, j = 1, \ldots, n_1, \psi_n$ is a normalizing factor, and $d(G, \hat{G}_n)$ is the Lebesgue measure of symetric difference $G \Delta \hat{G}_n$,

$$d(G, \hat{G}_n) = \text{ mes } (G \Delta \hat{G}_n).$$

Note that the boundary of $\hat{G}_n$ is not necessarily one-valued function of x_1.

The problem of estimation of G (or, equivalently, of the function g) from observations (2.1.) is called *edge estimation problem.*

The positive sequence ψ_n is called *minimax rate of convergence* and the estimator G_n^* is called *optimal estimator* if

$$\lim_{n \to \infty} \sup r_n(G_n^*, \psi_n) \leq c_1, \tag{2.3}$$

and for any estimator $\hat{G}_n$

$$\lim_{n \to \infty} \inf r_n(\hat{G}_n, \psi_n) \geq c_2 \tag{2.4}$$

where c_1, c_2 are positive constants.

As proved by Korostelev and Tsybakov (1992) for slightly different designs the minimax rate of convergence in this problem is equal to $n^{-\gamma/(\gamma+1)}$. In Korostelev and Tsybakov (1992) the optimal estimator G_n^* satisfying (2.3) with $\psi_n = n^{-\gamma/(\gamma+1)}$ was proposed based on combination of piecewise-polynomial approximation with maximum likelihood estimator (MLE).

Though edge estimation problem considered here is not formally covered by the result of Korostelev and Tsybakov (1992), nevertheless, it is a routine technique to extend it to the design (2.2). The aim of this paper is to construct another optimal estimator, based on solutions of change-point problems for each line $x_1 = i/n_1, \quad i = 1, \ldots, n_1$.

3. Naïve Linewise Processing

Note that for each line $x_1 = i/n_1$ we have an independent change-point problem with the change-point parameter $\theta_i = g(i/n_1), \quad i = 1, \ldots, n_1$ (the notation θ_i should be provided with index n which is omitted for the sake of brevity). It is necessary to emphasize that we study the posterior change-point problem as in Ibragimov and Khasminskii (1979), Sec.7.2, but not as a problem of sequential analysis (cf. Siegmund (1985)).

Denote by θ^*_{in} the MLE for θ_i obtained from observations

$$((X_{i1}, Y_{i1}), \ldots, (X_{in_1}, Y_{in_1})), \quad i = 1, \ldots, n_1.$$

The MLE's θ^*_{in} are independent for different i. We describe now some asymptotic properties of the MLE in assumption that n tends to infinity.

For the Gaussian errors in (2.1) the MLE θ^*_{in} coincides with the least squares estimator (LSE) which is the solution of the following minimization problem:

$$\min_{\theta} \sum_{j=1}^{n_1} \left[Y_{ij} - I\left(X_{ij} = \left(x_{ij}^{(1)}, x_{ij}^{(2)} \right) : x_{ij}^{(2)} \le \theta \right) \right]^2$$

or equivalently,

$$\max_{\theta} J_i(\theta), \tag{3.1}$$

$$J_i(\theta) = \sum_{j=1}^{n_1} I\left(X_{ij} : x_{ij}^{(2)} \le \theta \right) Y_{ij} - (1/2) \text{ card } \left\{ j : x_{ij}^{(2)} \le \theta \right\}.$$

Note that $J_i(\theta)$ are right-continuous piecewise constant functions having jumps at points $x_{ij}^{(2)}$. Due to the Gaussian distribution of errors in (2.1) there exists for each i a unique $\hat{j} = \hat{j}(i)$ such that any solution of (3.1) belongs to the interval $\left[x_{i,\hat{j}-1}^{(2)}, x_{i,\hat{j}}^{(2)} \right)$. We assume that the maximization in (3.1) is over $\theta \in [x_{i1}^{(2)}, x_{in}^{(2)})$, and thus possible boundary effects are neglected. It is convenient to define the MLE as the mean value of the extreme solutions; this can be written as

$$(i/n_1, \theta^*_{in}) = \left(X_{i,\hat{j}-1} + X_{i,\hat{j}} \right)/2, 1 < \hat{j} \le n_1.$$

In the following we assume that $\sigma^2 = 1$ without loss of generality. Define a two-sided random walk Z_k:

$$Z_0 = 0, \quad Z_k = W_k - (1/2) \mid k \mid, \quad k = \pm 1, \pm 2, \ldots,$$

where

$$W_k = \begin{cases} \Sigma_{i=1}^{k} \overline{\xi}_i \text{ if } k > 0, \\ \Sigma_{i=k}^{-1} \overline{\xi}_i \text{ if } k < 0, \end{cases}$$

$\overline{\xi}_i$ are independent standard $(0,1)$-Gaussian random variables. Let $\hat{k}$ be the maximizer for Z_k, i.e. $Z_{\hat{k}} > Z_k$ for $k \neq \hat{k}$. Note that the paths $Z_k \to -\infty$ with probability 1 as $|k| \to \infty$. Hence, the point $\hat{k}$ is unique and finite almost surely. Introduce the probabilities

$$\pi_k = P(\hat{k} = k), \quad k = 0, \pm 1, \ldots,$$

where P is the probability associated with Z_k.

3.1. Lemma. *The probabilities π_k are positive, symmetric, i.e. $\pi_k = \pi_{-k}$, and $\pi_k \leq \exp(-\lambda \mid k \mid)$ with some $\lambda > 0$ and $|k|$ large enough.*

Proof. Using the inequality $1 - \Phi(x) \leq (2/\pi)^{1/2} \exp(-x^2/2), \quad x > 1$, which holds for standard Gaussian distribution function $\Phi(x)$, one gets

$$\pi_k \leq P(\hat{k} \geq k) \leq P\left[\sup_{k_1 \geq k} Z_{k_1} \geq 0\right] \leq \sum_{k_1 \geq k} P\left[k_1^{-1/2} W_{k_1} \geq k_1^{1/2}/2\right] =$$

$$= \sum_{k_1 \geq k} \left(1 - \Phi(k_1^{1/2}/2)\right) \leq 2\pi^{-1/2} \sum_{k_1 \geq k} \exp(-k_1/8) \leq \exp(-\lambda k)$$

for some positivie λ and k large enough.The same argument is used for $k < 0$. ∎

Define the symmetric probability density

$$\pi(x) = \pi_k \text{ if } \; k - 0.5 < x < k + 0.5, \quad k = 0, \pm 1, \ldots,$$

3.2. Lemma. *There exists a positive constant λ_1 such that for n_1 large the following inequalities hold uniformly in $G \in \mathcal{G}(\gamma, L, h)$*

$$\sup_{-\infty < x < \infty} |P_G(n_1(\theta_{in}^* - \theta_i) \leq x) - \int_{-\infty}^{x} \pi(y)dy| \leq \exp(-\lambda_1 n_1),$$

and

$$|E_G(n_1(\theta^*_{in} - \theta_i))| \le \exp(-\lambda_1 n_1/2).$$

Proof. Let an integer $j_0 = j_0(\theta_i)$ be such that $x^{(2)}_{i,j_0-1} \le \theta_i < x^{(2)}_{ij_0}$, and let $\overline{\theta}_i$ be the midpoint of this interval, i.e.

$$\overline{\theta}_i = (x^{(2)}_{i,j_0-1} + x^{(2)}_{ij_0})/2.$$

We have

$$n_1(\theta^*_{in} - \theta_i) = n_1(\theta^*_{in} - \overline{\theta}_i) + n_1(\overline{\theta}_i - \theta_i),$$

where $n_1(\overline{\theta}_i - \theta_i)$ is uniformly distributed in the interval $(-0.5, 0.5)$. Note that the random integer j_0 depends only on the design $\mathcal{X}_i = (x^{(2)}_{i1}, \dots, x^{(2)}_{in_1})$ and does not depend on the random variables ξ_{ij}.

Set $h_0 = h/2$, $k_n = [n_1 h_0]$, and define the random event

$$\mathcal{A} = \{\mathcal{X}_i : n_1 h_0 \le j_0 \le n_1(1 - h_0) + 2\}.$$

Let us show that for $n > 3/h_0$ and for any $\mathcal{X}_i \in \mathcal{A}$ we have

$$|P_G(\hat{j} = k + j_0 | \mathcal{X}_i) - \pi_k| \le 2\exp(-\lambda(n_1 h_0 - 3))(a.s.), |k| \le k_n. \tag{3.2}$$

In fact,

$$P_G(\hat{j} = k + j_0 | \mathcal{X}_i) \le \pi_k + P(\hat{k} + j_0 \notin \{2, \dots, n_1\}).$$

By Lemma 3.1

$$P(\hat{k} + j_0 \notin \{2, \dots, n_1\}) \le \sum_{|k| \ge n_1 h_0 - 2} \pi_k \le 2 \sum_{k \ge n_1 h_0 - 2} \exp(-\lambda k) \le 2\exp(-\lambda(n_1 h_0 - 3)).$$

This gives the estimation for the probability in (3.2) from above. The estimation from below is quite similar.

For any integer k, $|k| \le k_n$, and for any $x \in (-1/2, 1/2)$, we have
$P_G\left(n_1(\theta^*_{in} - \theta_i) \in (k - 1/2, k + x)\right) =$

$$= P_G\left(n_1(\theta^*_{in} - \overline{\theta}_i) = k, n_1(\overline{\theta}_i - \theta_i) \in (-1/2, x)\right) =$$
$$= P_G\left(\hat{j} - j_0 = k, n_1(\overline{\theta}_i - \theta_i) \in (-1/2, x)\right).$$

Applying (3.2), one has

$$|P_G\left(n(\theta^*_{in}-\theta_i)\in(k-1/2,k+x)\right)-\int_{k-1/2}^{k+x}\pi(y)dy|\le 2\exp(-\lambda(n_1h_0-3)).$$

Hence for any $x, x\le k_n+1/2$,

$$|P_G(n_1(\theta^*_{in}-\theta_i)\le x)-\int_{-\infty}^{x}\pi(y)dy|\le$$

$$\le\sum_{k=-k_n}^{[x-1/2]}|P_G(n_1(\theta^*_{in}-\theta_i)\in(k-1/2,k+1/2))-\int_{k-1/2}^{k+1/2}\pi(y)dy|+$$

$$+|P_G(n_1(\theta^*_{in}-\theta_i)\in(x-1/2]+1/2,x))-\int_{[x-1/2]+1/2}^{x}\pi(y)dy|+$$

$$+P_G(n_1(\theta^*_{in}-\theta_i)\le -k_n-1/2)+\int_{-\infty}^{-k_n-1/2}\pi(y)dy\le$$

$$\le 2(2k_n+1)\exp(-\lambda(n_1h_0-3))+P_G(\hat{j}-j_0<-k_n)+\sum_{k\le -k_n}\pi_k.$$

Due to Lemma 3.1 and (3.2) the last two summands decrease exponentially as $n_1\to\infty$. this proves the first inequality of Lemma 3.2.

Applying this inequality and using the symmetry of $\pi(x)$ and the obvious property $|n_1(\theta^*_{in}-\overline{\theta}_i))|<n_1$ we have that uniformly in $\theta_i\in(h,1-h)$ the inequality holds

$$|E_G(n_1(\theta^*_{in}-\theta_i))|\le\int_{-\infty}^{\infty}x\pi(x)dx+n_1\exp(-\lambda_1n_1)$$
$$=n_1\exp(-\lambda_1n_1)\le\exp(-\lambda_1n_1/2)$$

for n_1 large. ∎

3.3. Corollary. *According to Lemma 3.2 for the MLE's θ^*_{in} the stochastic expansion holds*

$$\theta^*_{in}=\theta_i+b_n(\theta_i)+n_1^{-1}\varepsilon_{in},\quad i=1,\dots,n_1, \tag{3.3}$$

where the random errors ε_{in} are zero-mean, independent for different i, and have variances bounded uniformly in i and $n\ge 1$; for each value of i $b_n(\theta_i)$ denotes a bias term satisfying for n large

$$|b_n(\theta_i)|\le\exp(-\lambda_2n_1) \tag{3.4}$$

with some positive λ_2 *uniformly in* $\theta_i \in (h, 1-h)$.

We mean by naïve linewise processing the following two-step procedure. The first step consists of calculation θ^*_{in} for all the lines $x_1 = i/n_1$. At the second step some smoothing algorithm is applied to θ^*_{in}'s which are regarded now as new observations in the nonparametric regression

$$\theta^*_{in} \approx g(i/n_1) + \varepsilon_{in}/n_1, \quad i = 1, \dots, n_1 \tag{3.5}$$

Surely, the bias terms $b_n(\theta_i)$ in (3.3) are negligible as follows from (3.4). If we have estimated the regression function g from observations (3.5) in L_1-norm, this would lead to the same accuracy for the edge estimation in d-metric. It is well-known (see Ibragimov and Khasminskii (1979), Stone (1980, 1982), Härdle (1990)) that the minimax rate of convergence in L_1-norm in nonparametric regression problem (3.5) equals $\psi_n = (\sigma_\varepsilon^2/n_1)^{\gamma/(2\gamma+1)}$ where $\sigma_\varepsilon^2 = n_1^{-2} = n^{-1}$ is the variance of one observation in (3.5). Hence the best rate of convergence available from observations (3.5) is equal to

$$\psi_n = n^{-3\gamma/(4\gamma+2)} \tag{3.6}$$

Two conclusions follow from this. First, the rate of convergence obtained by the naïve linewise processing is not the best possible. Indeed, inequality $\gamma/(\gamma+1) > 3\gamma/(4\gamma+2)$ holds for $\gamma > 1$. Second, the rate of convergence achieved in (3.6) cannot be improved uniformly in $g \in \Sigma(\gamma, L, h)$ if we use only the MLE θ^*_{in} at the second step of the procedure.

Now we concentrate on the next problem: whether it is possible to modify this linewise processing in such a way that the resulting procedure would achieve the minimax rate of convergence? The answer is positive. Certainly, for such a procedure we must have not only the values θ^*_{in} but also some additional information for each line $x_1 = i/n_1, \quad i = 1, \dots, n_1$.

4. Modified Linewise Procedure

The idea of modification is rather simple. Suppose that at each point $(i/n_1, \theta^*_{in})$ we are able to get a new observation Y^*_i which is independent both of design (2.2) and of observations (2.1). Then we can try the LSE obtained from the new observations (θ^*_{in}, Y^*_i), $\quad i = 1, \dots, n_1$. All we have to do is to "extract" an observation from the data which would have the same properties as Y^*_i and which would not spoil nice properties of the new "design points" θ^*_{in}.

In addition suppose that $n_1 = n^{1/2}$ is even. Split each sample $((X_{i1}, Y_{i1}), \dots, (X_{i,n_1}, Y_{i,n_1}))$, $\quad i = 1, \dots, n_1$, into odd and even subsamples. Denote by θ_{in}^{odd} and $\theta_{in}^{\text{even}}$ the MLE's obtained respectively from the odd and even subsamples, i.e. from

$$((X_{i1}, Y_{i1}), (X_{i3}, Y_{i3}), \dots, (X_{i,n_1-1}, Y_{i,n_1-1}))$$

and

$$((X_{i2}, Y_{i2}), (X_{i4}, Y_{i4}), \dots, (X_{i,n_1}, Y_{i,n_1}))\,.$$

By definition the MLE θ_{in}^{odd} is such that

$$(i/n_1, \theta_{in}^{\text{odd}}) = (X_{i,\hat{j}-1} + X_{i,\hat{j}+1})/2 = X_{i,\hat{j}}, \tag{4.1}$$

for some $\hat{j} = \hat{j}(i), 1 < \hat{j}(i) < n_1, \hat{j}(i)$ even (cf. Section 3).

This is the only point where the technical convenience of the design (2.2) appears: the MLE strictly coincides with one of design points. Identity (4.1) holds for $\theta_{in}^{\text{even}}$ as well but now $\hat{j} = \hat{j}(i)$ is an odd integer.

Let i be fixed, and let the even integer $j_0 = j_0(i)$ be such that

$$(i/n_1, \theta_i) \in [X_{i,j_0-1}, X_{i,j_0+1}).$$

Define $\overline{\theta}_i^{\,\text{odd}}$ as the second coordinate of the midpoint of this interval, i.e.

$$(i/n_1, \overline{\theta}_i^{\,\text{odd}}) = (X_{i,j_0-1} + X_{i,j_0+1})/2 = X_{i,j_0}, \quad i = 1, \dots, n_1.$$

The definition of $\overline{\theta}_i^{\,\text{even}}$ is quite similar, with j_0 odd.

Now we want to find the distribution functions of the random variables $n_1(\theta_{in}^{\text{odd}} - \theta_i)$ and $n_1(\theta_{in}^{\text{even}} - \theta_i)$.

Unlike the case of Lemma 3.2 these random variables are biased:

$$\liminf_{n\to\infty} \sup_{G\in\mathcal{G}(\gamma,L,h)} E_G(n_1(\theta_{in}^{\text{odd}} - \theta_i)) > 0.$$

The reason is quite obvious: the distribution of θ_{in}^{odd} doesn't change if θ_i, varies within the interval $((j_0-1)/n_1, j_0/n_1)$, j_0 even, since the corresponding points $(i/n_1, \theta_i)$ are covered by the same design interval $[X_{i,j_0-1}, X_{i,j_0+1})$ under any "shift value" $\eta_i, 0 < \eta_i < n_1^{-1}$. The same is true for $\theta_{in}^{\text{even}}$: the random shift coveres only one half of each design interval. As in Lemma 3.2, we first find the distribution of the difference $n_1(\overline{\theta}_i^{\text{odd}} - \theta_i)$. Unfortunately, this difference is not uniformly distributed as was the case in Lemma 3.2. Routine considerations show that the random variable $n_1(\overline{\theta}_i^{\text{odd}} - \theta_i)$ is uniformly distributed in

$$\begin{cases} (-1,-\tau_i)\cup(1-\tau_i,1) & \text{if } [n_1\theta_i] \text{ is even ,} \\ (-\tau_i, 1-\tau_i) & \text{if } [n_1\theta_i] \text{ is odd ,} \end{cases} \tag{4.2}$$

where $\tau_i = \tau_i(\theta_i) = n_1\theta_i - [n_1\theta_i], 0 \le \tau_i < 1$. Similarly, the random variable $n_1(\overline{\theta}_i^{\text{even}} - \theta_i)$ is uniformly distributed in

$$\begin{cases} (-1,-\tau_i)\cup(1-\tau_i,1) & \text{if } [n_1\theta_i] \text{ is odd ,} \\ (-\tau_i, 1-\tau_i) & \text{if } [n_1\theta_i] \text{ is even .} \end{cases} \tag{4.3}$$

Hence by analogy to Lemma 3.2 one may expect that the distribution functions of $n_1(\theta_{in}^{\text{odd}} - \theta_i), n_1(\theta_{in}^{\text{even}} - \theta_i)$ are approximated by the distributions having piecewise constant densities. But unlike the probability density in Lemma 3.2, they have a saw-tooth shape with periodical zero gaps. The period of zero gaps is equal to 1, and the gaps for $n_1(\theta_{in}^{\text{odd}} - \theta_i)$ alternate with those for $n_1(\theta_{in}^{\text{even}} - \theta_i)$. Thus it is reasonable to mix the estimators θ_{in}^{odd} and $\theta_{in}^{\text{even}}$ randomly, i.e. to define

$$\theta_{in}^* = \begin{cases} \theta_{in}^{\text{odd}} & \text{with probability } 1/2 \\ \theta_{in}^{\text{even}} & \text{with probability } 1/2 \end{cases} \tag{4.4}$$

where some independent stochastic procedure is applied to make randomization in (4.4). We use in (4.4) the same notation θ_{in}^* as that in Section 3 since only this randomized estimator is considered below.

4.1. Lemma. *The random variable $n_1(\theta_{in}^* - \theta_i)/2$ satisfies the statement of Lemma 3.2.*

Proof is the same as in Lemma 3.2: the discrete distributions

$$P_G\left(n_1(\theta_{in}^{\text{odd}} - \overline{\theta}_i^{\text{odd}})/2 = k|\mathcal{X}_i\right) \approx \pi_k$$

and

$$P_G\left(n_1(\theta_{in}^{\text{even}} - \overline{\theta}_i^{\text{even}})/2 = k|\mathcal{X}_i\right) \approx \pi_k$$

must be convolved with (4.2) and (4.3) and then mixed randomly.

Let $j^* = j^*(i)$ be such that $(i/n_1, \theta^*_{in}) = X_{i,j^*}, i = 1, \dots, n_1$ (here j^* is randomized and is defined by θ^*_{in}). Define $Y_i^* = Y_{i,j^*}$ where Y_{i,j^*} are observations (2.2) at the design points X_{i,j^*}:

$$Y_i^* = I(X_{i,j^*} \in G) + \zeta_i, \quad i = 1, \dots, n_1, \tag{4.5}$$

and $\zeta_i = \xi_{i,j^*}$ are i.i.d. $(0,1)$-Gaussian random variables. Note that ζ_i is independent of θ^*_{in}.

At the first step of the modified procedure the pairs (Y_i^*, θ^*_{in}) are calculated. It turns out that the total necessary information, additional to θ^*_{in}, is concentrated in Y_i^*: the knowledge of pairs (Y_i^*, θ^*_{in}) is sufficient to define the minimax linewise procedure.

The second step of the modified linewise procedure is determined by the LSE for piecewise-polynomial approximation. Let $\delta_n = n^{-1/(\gamma+1)}$, and $M = M_n = 1/\delta_n$ (assume without loss of generality that M is an integer). Define points $u_\ell = \ell\delta_n$ and slices $K_\ell = [u_{\ell-1}, u_\ell) \times R^1$, and introduce the domains $G_\ell = G \cap K_\ell, \ell = 1, \dots, M$.

In each slice K_ℓ define a parametric family of domains

$$B_\ell(\alpha) = \left\{(x_1, x_2) \in K_\ell : 0 \le x_2 \le \alpha_0 + \alpha_1(x_1 - u_{\ell-1}) + \ldots + \alpha_k(x_1 - u_{\ell-1})^k\right\}$$

where $\alpha = (\alpha_0, \dots, \alpha_k)$ is a vector of real parameters. For some values of α the curve

$$g_\ell(x_1; \alpha) = \alpha_0 + \alpha_1(x_1 - u_{\ell-1}) + \ldots + \alpha_k(x_1 - u_{\ell-1})^k, u_{\ell-1} \le x_1 < u_\ell,$$

may leave the unit square K. Let

$$A^{(K)} = \left\{\alpha : \quad 0 \le \alpha_0 + \alpha_1(x_1 - u_{\ell-1}) + \ldots + \alpha_k(x_1 - u_{\ell-1})^k \le 1, u_{\ell-1} \le x_1 < u_\ell\right\}.$$

For any $\alpha \in A^{(K)}$ the domain $B_\ell(\alpha)$ is correctly defined and located within K.

Consider the MLE $\hat{\alpha}_\ell$ of α based on observations in slice K_ℓ. To avoid technical problems we consider $\alpha \in A_n^{(K)}$ where $A_n^{(K)}$ is the discrete set

$$A_n^{(K)} = \left\{\alpha = (m_0\delta_n^\gamma, m_1\delta_n^{\gamma-1}, \ldots, m_k\delta_n^{\gamma-k}) : m_j \text{ integers } \right\} \cap A^{(K)}.$$

The MLE $\hat{\alpha}_\ell = (\hat{\alpha}_{0\ell}, \ldots, \hat{\alpha}_{k\ell})$ is defined as a solution of the extremal problem

$$J_\ell(\hat{\alpha}_1) = \max_{\alpha \in A_n^{(K)}} J_\ell(\alpha) \tag{4.6}$$

where

$$J_\ell(\alpha) = \sum_i I\left(X_{i,j^*} \in B_\ell(\alpha)\right) Y_i^* - (1/2) \text{ card } \left\{i : X_{i,j^*} \in B_\ell(\alpha)\right\}.$$

We define the edge estimator

$$\hat{g}_\ell(x_1) = \hat{\alpha}_{0\ell} + \hat{\alpha}_{1\ell}(x_1 - u_{\ell-1}) + \ldots + \hat{\alpha}_{k\ell}(x_1 - u_{\ell-1})^k$$

for $u_{\ell-1} \le x_1 < u_\ell, \quad \ell = 1, \ldots, M$. Finally, the estimator G_n^* of G is the closure of the set

$$\{(x_1, x_2) : 0 \le x_2 \le \hat{g}_\ell(x_1), 0 \le x_1 \le 1\} \tag{4.7}$$

The following theorem states that G_n^* has the best possible rate of convergence.

4.2 Theorem. *Let θ_{in}^* be the randomized estimators (4.4) of the change-point parameters $\theta_i = g(i/n_1)$ based on observations (2.1). Then the domain estimator G_n^* defined as the closure of (4.7) has the minimax rate of convergence $\psi_n = n^{-\gamma/(\gamma+1)}$ in d-metric.*

5. Proofs

We use some results of our paper (Korostelev and Tsybakov (1992)) which we refer to in this section as KT. Assume without loss of generality that $n_1\delta_n = n^C$ is an integer tending to infinity as $n \to \infty; C = (\gamma - 1)/(2(\gamma + 1))$ is positive for $\gamma > 1$. Introduce the next notation for the part of line $x_1 = i/n_1$ which belongs to the domain $B_\ell(\alpha)\Delta G_\ell$:

$$V_i = V_i(\alpha) = \{(x_1, x_2) : x_1 = i/n_1\} \cap [B_\ell(\alpha)\Delta G_\ell]\,.$$

By definition V_i are connected intervals. Let $|V_i|$ be the length of the interval V_i, and let

$$I_\ell = \{i : u_{\ell-1} \le i/n_1 < u_\ell\}$$

be the set of indices i such that $V_i \subset K_\ell$. Note that

$$\text{card } I_\ell = n_1\delta_n = n^C \to \infty \text{ as } n \to \infty.$$

We associate with each $\alpha \in A_n^{(K)}$ the set of integers $m = (m_0, \dots, m_k)$ in the natural way (see the definition of $A_n^{(K)}$) and mark this correspondence by $\alpha = \alpha(m)$. Introduce the norm $\| m \|$ as maximum of $|m_j|, 0 \le j \le k$. Let the set of integers $m_\ell^{(0)}$ correspond to the Taylor approximation of edge function $g(x_1)$ around $x_1 = u_{\ell-1}$, and $\hat{m}_\ell$ correspond to the estimator $\hat{\alpha}_\ell$ in (4.6): $\hat{\alpha}_\ell = \alpha(\hat{m}_\ell)$.

The problem (4.6) is equivalent to

$$\max_{\alpha \in A_n^{(K)}} Z_\ell(\alpha), Z_\ell(\alpha) = J_\ell(\alpha) - J_\ell^{(G)} \tag{5.1}$$

where

$$J_\ell^{(G)} = \sum_{i \in I_\ell} I\left(X_{i,j^*} \in G\right) Y_i^* - (1/2) \text{ card } \{i : X_{i,j^*} \in G\}$$

is independent of α.

5.1. Lemma. *For each $\ell, \ell = 1, \dots, M$, the random function $Z_\ell(\alpha)$ can be presented in the form*

$$Z_\ell(\alpha) = \Delta_\ell(\alpha) - (1/2) \sum_{i \in I_\ell} P_G\left(X_{i,j^*} \in V_i(\alpha)\right)$$

where the random variables $\Delta_\ell(\alpha)$ are independent for different ℓ, $E_G[\Delta_\ell(\alpha)] = 0$, and for all z small enough

$$E_G[\exp(z\Delta_\ell(\alpha))] \le \exp\left\{\lambda_0 z^2 \sum_{i\in I_\ell} P_G(X_{i,j^*} \in V_i(\alpha))\right\} \tag{5.2}$$

with a positive λ_0 which does not depend on n, ℓ, α and G.

Proof. Note that (4.5) entails the representation

$$Z_\ell(\alpha) = \sum_i I(X_{i,j^*} \in B_\ell(\alpha) \setminus G_\ell)\zeta_i - \sum_i I\left(X_{i,j^*} \in G_\ell \setminus B_\ell(\alpha)\right)\zeta_i - (1/2)\Lambda_\ell =$$

$$= \Delta_\ell(\alpha) - (1/2)\Lambda_\ell$$

where

$$\Delta_\ell(\alpha) = \sum_i I(X_{i,,j^*} \in B_\ell(\alpha) \setminus G_\ell)\zeta_i - \sum_i I\left(X_{i,j^*} \in G_\ell \setminus B_\ell(\alpha)\right)\zeta_i$$

and $\Lambda_\ell =$ card $\{i : X_{i,j^*} \in B_\ell(\alpha)\Delta G_\ell\}$ with the mean value

$$E_G(\Lambda_\ell) = \sum_{i\in I_\ell} P_G\left(X_{i,j^*} \in V_i(\alpha)\right).$$

The rest of proof is similar to that in KT, Lemma 4.2.2. ∎

5.2. Lemma. For n large enough and $\alpha \in A_n^{(K)}$

$$(1/2)d(B_\ell(\alpha), G_\ell) \le (1/n_1)\sum_{i\in I_\ell} |V_i(\alpha)| \le 2d(B_\ell(\alpha), G_\ell)$$

where $n_1 = n^{1/2}$.

Proof. Note that $(1/n_1)\sum_{i\in I_\ell} |V_i(\alpha)|$ is the Riemann sum for

$$\int_{u_{\ell-1}}^{u_\ell} |\alpha_0 + \alpha_1(x_1 - u_{\ell-1}) + \ldots + \alpha_k(x_1 - u_{\ell-1})^k - g(x_1)|dx_1 = d(B_\ell(\alpha), G_\ell).$$

The latter quantity is approximated by the Riemann sum with accuracy

$$O(d(B_\ell(\alpha), G_\ell)/ \text{ card } I_\ell) \le (1/2)d(B_\ell(\alpha), G_\ell)$$

for all sufficiently large n. ∎

5.3. Lemma. *If* $\| m - m_\ell^{(0)} \| \leq n^C/(\gamma+1)$ *and* $\alpha = \alpha(m)$, *then for all* ℓ, $\ell = 1, \ldots, M$, *and* n *large enough the following inequalities hold:*

$$\max_{i \in I_\ell} |V_i(\alpha)| \leq 1/n_1; \tag{5.3}$$

(5.4)
$$n_1(C_1|V_i(\alpha)| - \exp(-\overline{\lambda}n_1)) \leq P_G(X_{i,j^*} \in V_i(\alpha)) \leq n_1(C_2|V_i(\alpha)| + \exp(-\overline{\lambda}n_1))$$

$$(1/2)C_1 d(B_\ell(\alpha), G_\ell) - \exp(-\overline{\lambda}n_1) \leq \sum_{i \in I_\ell} P_G(X_{i,j^*} \in V_i(\alpha)) \tag{5.5}$$

$$\leq 2C_2 d(B_\ell(\alpha), G_\ell) + \exp(-\overline{\lambda}n_1),$$

with some positive constants C_1, C_2, *and* $\overline{\lambda}$.

Proof. The boundaries of $B_\ell(\alpha)$ and G_ℓ are given respectively by the edge function $g_\ell(x_1; \alpha) = \delta_n^\gamma(m_0 + m_1 y + \ldots + m_k y^k)$ where $y = (x_1 - u_{\ell-1})/\delta_n, \quad 0 \leq y \leq 1$, and by $g(x_1)$. Comparing this whith the Taylor expansion for $g(x_1)$ at $x_1 = u_{\ell-1}$, we get

$$\max_{i \in I_\ell} |V_i(\alpha)| \leq \delta_n^\gamma L/k! + \delta_n^\gamma (k+1) \| m - m_\ell^{(0)} \| \leq \delta_n^\gamma L/k! + \delta_n^\gamma n^C (k+1)/(\gamma+1) \leq$$

$$\leq \delta_n^\gamma n^C = n^{-1/2} = 1/n_1,$$

for n large, which proves (5.3).

Note that

$$P_G(X_{i,j^*} \in V_i(\alpha)) = P_G[n_1(\theta_{in}^* - \theta_i)/2 \in (n_1/2)(V_i(\alpha) - \theta_i)]$$

where $(V_i(\alpha) - \theta_i)$ is the shift of the interval $V_i(\alpha)$. By definition, one of the ends of $(V_i(\alpha) - \theta_i)$ is the origin. Hence, Lemma 4.1 applies which states that the distribution function of $(n_1/2)(\theta_{in}^* - \theta_i)$ is approximated by that having density $\pi(x)$. Taking into account (5.3) one gets (5.4). The inequalities (5.5) are the direct consequence of Lemma 5.2 and (5.4). ∎

5.4 Lemma. *If* $\| m - m_l^{(0)} \| > n^C/(\gamma+1)$ *and* $\alpha = \alpha(m) \in A_n^{(K)}$, *then*

$$\sum_{i \in I_\ell} P_G \left(X_{i,j^*} \in V_i(\alpha(m)) \right) \geq C_3 n^C \tag{5.6}$$

for n *large enough with some constant* $C_3 > 0$.

Proof. Assume for definiteness that $g(i/n_1) \leq g_\ell(i/n_1; \alpha(m))$. Then

$$P_G(X_{i,j^*} \in V_i(\alpha)) = P_G(g(i/n_1) \leq \theta_{in}^* \leq g_\ell(i/n_1; \alpha(m)))$$

Using the definition of these functions, we have

$$g_\ell(x_1; \alpha(m)) - g(x_1) = \left[\sum_{j=0}^{k} \left[\alpha_j(m) - \frac{g^{(j)}(u_{l-1})}{j!} \right] (x_1 - u_{l-1})^j \right] +$$

$$+ \left[\sum_{j=0}^{k} \frac{g^{(j)}(u_{l-1})}{j!} (x_1 - u_{l-1})^j - g(x_1) \right], \quad u_{l-1} \leq x_1 < u_l.$$

Since the last summand presents the Taylor approximation, we have

$$g_\ell(x_1; \alpha(m)) - g(x_1) \geq \delta_n^\gamma [\sum_{j=0}^{k} (m_j - m^{(0,j)}) y^j - L/k!] \geq$$

$$\geq \delta_n^\gamma \| m - m_\ell^{(0)} \| \left[\sum_{j=0}^{k} (m_j - m_\ell^{(0,j)}) \| m - m_\ell^{(0)} \|^{-1} y^j - L/k! \| m - m_\ell^{(0)} \|^{-1} \right]$$

where $y = (x_1 - u_{l-1})/\delta_n$. Since $\| m - m_\ell^{(0)} \| > n^C/(\gamma+1)$ the last term is negligible as n tends to infinity. There exists an interval U of length $|U| > 0$ and a positive C_4 such that

$$\sum_{j=0}^{k} (m_j - m^{(0,j)}) \| m - m_\ell^{(0)} \|^{-1} y^j \geq C_4 \text{ for } y \in U.$$

Indeed, the set of polynomials on the left-hand side is compact and separated from zero. It means that among $\{V_i(\alpha(m)), i \in I_l\}$ in (5.6) there exist at least $|U| n_1 \delta_n = |U| n^C$ intervals with lenghts exceeding

$$(C_4/2)\delta_n^\gamma n^C/(\gamma+1) \geq C_4/2 \; \delta_n^{\gamma+1} n_1/(\gamma+1) = C_4/2 \; n_1^{-1}/(\gamma+1).$$

For n large each of the corresponding probabilities $P_G(X_{i,j^*} \in V_i(\alpha(m)))$ is bounded from below by a positive constant in accordance with (5.4). This proves (5.6). ∎

Proof of Theorem 4.2. It was shown in KT, theorem 4.1.1, that Lemmas 5.1, 5.2, and inequality (5.5) entail inequality for any $R > 0$:

(5.7)

$$P_G\left[\max_{R\le\|m-m_\ell^{(0)}\|\le n^C/(\gamma+1)}[Z_\ell(\alpha(m)) - Z_\ell(\alpha(m_\ell^{(0)}))] > 0\right] \le C_5\exp(-C_6R).$$

with some positive constants C_5, C_6. For each m such that $\| m-m_\ell^{(0)} \| > n^C/(\gamma+1)$ Lemma 5.4 guarantees that the corresponding probability $P_G(\hat{\alpha}_\ell = \alpha(m))$ decreases exponentially as $n \to \infty$. In fact, for n large enough we have

$$P_G(\hat{\alpha}_\ell = \alpha(m)) \le P_G(Z_\ell(\alpha(m)) - Z_\ell(\alpha(m_\ell^{(0)})) > 0) \le$$

$$\le P_G\left[\Delta_\ell(\alpha(m)) - \Delta_\ell(\alpha(m_\ell^{(0)})) > (1/4)\sum_{i\in I_\ell} P_G(X_{i,j^*} \in V_i(\alpha))\right] \le$$

$$\le C_7\exp\left[\lambda_0 z^2\sum_{i\in I_\ell} P_G(X_{i,j^*} \in V_i(\alpha)) - (z/4)\sum_{i\in I_\ell} P_G(X_{i,j^*} \in V_i(\alpha))\right]$$

where $C_7 > 0$ is a constant and λ_0 is the constant from (5.2). If we take $z = (1/8\lambda_0)$ and apply Lemma 5.4, then

$$P_G(\hat{\alpha}_\ell = \alpha(m)) \le C_7\exp\left\{-C_8\sum_{i\in I_\ell} P_G\left(X_{i,j^*} \in V_i(\alpha)\right)\right\} \le C_7\exp(-C_8n^C/(\gamma+1))$$

where $C_8 = (64\lambda_0)^{-1}$. Hence

$$P_G\left(\| \hat{m}_\ell - m_\ell^{(0)} \| > n^C/(\gamma+1)\right) \le \tag{5.8}$$

$$\le C_7 \text{ card } (A_n^{(K)})\exp(-C_8n^C/(\gamma+1)) \le \exp(-C_9n^C)$$

for $C_9 = C_8/2$ and n large enough. Note that $\| \hat{m}_\ell - m_\ell^{(0)} \|$ are independent random variables, $\ell = 1, \dots, M$, and in accordance with (5.7) and (5.8) their distribution have exponentially decreasing tails.

It is shown in KT, Lemma 4.2.1, that

$$d(B_\ell(\alpha(m)), G_\ell) \le \delta_n^{\gamma+1}(\lambda_3 \| m - m_\ell^{(0)} \| + \lambda_4)$$

with some positive constants λ_3, λ_4. Hence

$$P_G(\delta_n^{-\lambda} d(G, G_n^*) \geq q) \leq P_G \left[\delta_n \sum_{\ell=1}^{M} (\lambda_3 \parallel \hat{m} - m_\ell^{(0)} \parallel + \lambda_4) \geq q \right] \leq \tag{5.9}$$

$$\leq P_G \left[M^{-1} \sum_{\ell=1}^{M} \parallel \hat{m} - m_\ell^{(0)} \parallel \geq (q - \lambda_4)/\lambda_3 \right].$$

The last probability decreases exponentially in q as $q \to \infty$. This proves the theorem. ∎

REFERENCES

Blake,A. and Zisserman,A. : (1987) *Visual Reconstruction.* MIT Press, Cambridge, MA.

Bretagnolle,J. and Huber,C. (1979) Estimation des densités: risque minimax, *Z. für Wahrscheinlichkeitstheorie und verw. Geb* 47, 119-137.

Geman,S. and Geman,D. (1984) Stochastic relaxation, Gibbs distribution, and Bayesian restoration of images. *IEEE Trans. on PAMI*, 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*, Proc. Conf. Santa Clara, California, 13-15 february, 1990, 174-182.

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

Härdle,W.(1990) *Applied Nonparametric Regression*, Cambridge Univ. Pres.

Huang,J.S. and Tseng, D.H. (1988) Statistical theory of edge detection. *Comput. Graphics and Image Processing* 43, 337-346.

Ibragimov,I.A. and Khasminskii,R.Z. (1979) *Statistical Estimation: Asymptotic Theory*, Springer-Verlag, N.Y.

Korostelev,A.P. (1991) Minimax reconstruction of two-dimensional images, *Theory of Probability and its Applications* 36, 153-159.

Korostelev,A.P. and Tsybakov,A.B. (1992) *Asymptotically minmax image reconstruction problems.* In: "Topics in Nonparametric Estimation", Khasminskii R.Z., ed., AMS, Providence, RI.

Marr,D.(1982) *Vision*, W.H. Freeman and Co., San Francisco.

Mumford,D. and Shah,J.(1989) Optimal approximation by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* 42, 577-685.

Nagao,M. and Matsujama,T. (1979) Edge preserving smoothing. *Graphics and Image Processing* 9, 394-407.

Pratt,W.K.(1978) *Digital Image Processing*, J. Wiley, N.Y.

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.

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

Stone,C.J. (1980) Optimal convergence for nonparametric estimators, *Annals of Statistics* 8, 1348-1360.

Stone,C.J. (1982) Optimal global rates of convergence for nonparametric estimators, *Annals of Statistics* 10, 1040-1053.

Titterington,D.M. (1985) Common structure of smoothing techniques in statistics. *Internat. Statist. Review* 53, 141-170.

Torre,V. and Poggio,T. (1986) On edge detection. *IEEE Trans. on PAMI* 8, 147-163.

Tsybakov,A.B. (1989) Optimal estimation accuracy of nonsmooth images, *Problems of Information Transmission* 25, 180-191.