Journal of Multivariate Analysis 112 (2012) 194–206

Contents lists available at SciVerse ScienceDirect

Journal of Multivariate Analysis

journal homepage: www.elsevier.com/locate/jmva

Poisson intensity estimation for the Spektor–Lord–Willis problem using a wavelet shrinkage approach

Bogdan Ćmiel

Faculty of Applied Mathematics, AGH University of Science and Technology, Al. Mickiewicza 30, 30-059 Cracow, Poland

ARTICLE INFO

Article history:
Received 13 July 2010
Available online 1 July 2012

AMS subject classifications:
62G05
45Q05

Keywords:
Spektor–Lord–Willis problem
Inverse problem
Rate of convergence
Minimax risk
Adaptive estimator
Empirical risk minimization

ABSTRACT

In this paper, we focus on nonparametric estimation in the stereological problem of unfolding sphere size distribution from linear sections. Using a Wavelet–Vaguelette Decomposition (WVD), we construct a rate minimax estimator of the intensity function of a Poisson process that describes the problem. This paper builds upon recent results by the same author concerning the model with a minimal detection radius and shows that this restriction is not necessary to obtain the minimax risk. The proposed adaptive estimator achieves the optimal rate of convergence over Besov balls to within logarithmic factors. Additionally, a construction of a new method of selection of a smoothing parameter by empirical risk minimization is discussed in detail. This paper also demonstrates finite sample behavior of this estimator in a numerical experiment, by using a discrete version of the wavelet algorithm.

1. Introduction

Let us consider a population of spheres randomly distributed in some opaque medium. We assume, that the centers of the spheres form a homogeneous Poisson process on $\mathbb{R}^3$, and the radii are random with a distribution Q on $[0; 1]$, independent of the centers and absolutely continuous with respect to the Lebesgue measure with probability density ρ. Since we cannot observe the spheres directly, we take a linear section through the medium and observe the line segments that are intersections of the line and the spheres. Denote the expected number of sphere centers per unit volume by c. Let n be the "size of the experiment" and $f := c\,\rho$. We thus observe a Poisson process G^n on $[0; 1]$ with intensity ng (with respect to the Lebesgue measure) and it can be shown (see [12]) that

$$g(u) = 2u \int_u^1 f(x)dx =: (Gf)(u), \tag{1}$$

where $G : L^2([0; 1], dx) \to L^2([0; 1], du)$. From that observation we want to estimate f. As in [2], for mathematical tractability, we divide both sides of (1) by u^2, and obtain

$$h(u) := \frac{g(u)}{u^2} = \frac{2}{u} \int_u^1 f(x)dx =: (Kf)(u),$$

where $K : L^2([0; 1], dx) \to L^2([0; 1], d\mu)$, $d\mu = u^2 du$. As the operators K and G are compact Hilbert–Schmidt operators, their inverses are not bounded and the problem of unfolding f from linear sections, known in the literature as the Spektor–Lord–Willis (SLW) problem, is ill-posed in the Hadamard sense (cf. the related discussion in [12,2]). In the next sections, $\langle \cdot, \cdot \rangle$ and $[\cdot, \cdot]$ will denote the inner products in $L^2(dx)$ and $L^2(d\mu)$, respectively.

E-mail address: cmielbog@mat.agh.edu.pl.

0047-259X/$ – see front matter
doi:10.1016/j.jmva.2012.06.009

The SLW problem has applications in material sciences. For example in [11] there is a description of a method of measuring air-void systems in hardened concrete which leads to the SLW problem. The intensity of small air bubbles inside has an influence on strength of materials (for example on a frost-resistance). The SLW problem can also be used in metallurgy. In [1], a hot-deformation of some metals under different hot-working conditions was considered, and the grain diameter structure in several materials was analyzed. The distribution of that diameters has an influence on strength of materials and also on some other physical properties like thermal conductivity. We can also analyze that distribution to answer the question: "which processes of hot-working were used on this piece of metal"? If the intensity of the grain has several peaks it can be concluded that different characteristic processes were used. The SLW problem can also be applied for an analysis of the sintering process in spherical grain shape cases (see [7,9]). The grain size distribution is the most important microstructural parameter and has great influence on the material properties (for more information see [6] which is the sintering process review based on many publications). Notice that in practice we always examine finite objects so, in those cases, the assumption that the distribution of the random radius has a compact support is quite reasonable and one can assume the support, without loss of generality, to be the interval [0; 1].

The SLW problem was recently analyzed in [12] where B-spline sieved quasi-maximum likelihood estimators were used. Construction of a spectral estimator that is asymptotically rate minimax over a Sobolev-type class of functions can be found in [5,13]. In [2] a minimax estimator was constructed over Besov balls under restriction of the domain to radii larger than some positive minimal detection level ε. In this paper, we relax this assumption by replacing the domain restriction with some other assumptions on the local behavior of the estimated function in the vicinity of zero. We also propose a new method of selection of a smoothing parameter by empirical risk minimization and present the behavior of the new estimator with a data-driven choice of parameters on some examples in a numerical experiment.

2. WVD-based reproducing formula

In this section, we use the WVD (see, [3, Section 5.2]) of the operator K. The WVD construction details will be omitted, because they are the same as in [2]. Let ψ be a smooth mother wavelet that satisfies the conditions: supp $\psi = [0; N]$, $\int_{-\infty}^{\infty} \psi(x)dx = 0$, $\int_{-\infty}^{\infty} \psi^2(x)dx = 1$, $\psi \in C^2$ (it has two continuous derivatives) and $\|\psi'\|_{L^2(dx)} < \infty$. The functions $\psi_{jk}(x) = 2^{j/2}\psi(2^j x - k)$, with j and k integers form a complete, orthonormal system in $L^2(dx)$. Let ϕ be a father wavelet that satisfies the conditions: supp $\phi = [0; N]$, $\int_{-\infty}^{\infty} \phi(x)dx = 1$ and $\phi \in C^2$. If we assume, for a while, that $f \in L^2$, then it has the following inhomogeneous wavelet expansion (cf. [8, Chapter 3.2])

$$f = \sum_{k\in\mathbb{Z}} \langle f, \phi_{j_1 k}\rangle \phi_{j_1 k} + \sum_{j=j_1}^{\infty} \sum_{k\in\mathbb{Z}} \langle f, \psi_{jk}\rangle \psi_{jk},$$

where j_1 is a fixed integer and $\phi_{jk} = 2^{j/2}\phi(2^j x - k)$. Define

$$\gamma_{jk}(u) := \frac{\psi'_{jk}(u)}{2u} = 2^{\frac{3}{2}j} \frac{\psi'(2^j u - k)}{2u} \tag{2}$$

and

$$\tilde{\gamma}_{j_1 k}(u) = \frac{\phi'_{j_1 k}(u)}{2u} = 2^{\frac{3}{2}j_1} \frac{\phi'(2^{j_1} u - k)}{2u}. \tag{3}$$

Using (2) and (3), we have the reproducing formula (cf. [2])

$$f = \sum_{k\in\mathbb{Z}} [Kf, \tilde{\gamma}_{j_1 k}]\phi_{j_1 k} + \sum_{j=j_1}^{\infty} \sum_{k\in\mathbb{Z}} [Kf, \gamma_{jk}]\psi_{jk}.$$

Although $K\phi$ may not belong to $L^2(d\mu)$ the reproducing formula remains valid at least for functions f with only finite number of nonzero terms in the inhomogeneous wavelet expansion (cf. [3, p. 111]).

3. Minimax risk

Besov spaces can conveniently be defined in terms of wavelet coefficients. With $\sigma > 0$ and $p, q \geqslant 1$, a function $f = \sum_{k\in\mathbb{Z}} \alpha_{j_1 k}\phi_{j_1 k} + \sum_{j=j_1}^{\infty} \sum_{k\in\mathbb{Z}} \beta_{jk}\psi_{jk}$ belongs to the Besov ball $B_{\sigma pq}(M)$ if and only if

$$\|\alpha_{j_1 \cdot}\|_{l_p} + \left(\sum_{j=j_1}^{\infty} \left(2^{j(\sigma+1/2-1/p)} \|\beta_{j\cdot}\|_{l_p}\right)^q \right)^{1/q} \leqslant M.$$

Roughly speaking, the Besov space $B_{\sigma pq}$ consists of functions that "have $\lfloor \sigma \rfloor$ weak derivatives in L^p ", and the parameter q is of secondary importance, because, if $\sigma_1 > \sigma_2$, then $B_{\sigma_1 pq_1} \subset B_{\sigma_2 pq_2}$ for all $q_1, q_2 \geqslant 1$. In [2], we assumed that the spheres with radii $r < \varepsilon$ could not be observed, and all estimations were made on an interval separated from zero. In this paper

we show that, under some assumptions on the behavior of f in the vicinity of zero, a minimax estimator of the intensity function f on $[0;\,1]$ can also be constructed. From here we will write $\|\cdot\|_{L^2}$ rather than $\|\cdot\|_{L^2(dx)}$ (this is the norm defined by $\langle\cdot,\cdot\rangle$). Define

$$\mathcal{F}_{\sigma pq}(M,\varepsilon) = \Big\{f \in B_{\sigma pq}(M) : \operatorname{supp} f \subset [0;\,1],\ f \geqslant 0,\ f(x) = O\left(x^{(\frac{4}{3}\sigma-\frac{1}{2})_+}\right) \text{as } x \to 0^+,$$
$$\exists f_\varepsilon \in B_{\frac{4\sigma}{3}pq}(M)\ f_\varepsilon \cdot \mathbb{1}_{[0;\varepsilon]} = f \cdot \mathbb{1}_{[0;\varepsilon]}\Big\}.$$

This is the class of all nonnegative functions from $B_{\sigma pq}$ with the support in the interval $[0;\,1]$ with two additional assumptions near the origin: there exists $\varepsilon > 0$ such that on the interval $[0;\,\varepsilon]$ the "smoothness parameter" of the function f is equal to $4\sigma/3$, and the values of the function $f(x)$ approach zero at least as fast as $x^{(\frac{4}{3}\sigma-\frac{1}{2})_+}$ when x approaches zero. The reason for the assumptions near the origin will be discussed in the proof of Theorem 1.

Let us denote

$$\mathcal{K}_j^\phi(A) := \left\{k \in \mathbb{Z} : \operatorname{supp}\phi_{jk} \cap A \neq \emptyset\right\},$$
$$\mathcal{G}_j^\phi(A) := \left\{k \in \mathbb{Z} : \operatorname{supp}\phi_{jk} \subset A\right\},$$
$$\mathcal{H}_j^\phi(A) := \left\{k \in \mathbb{Z} : \operatorname{supp}\phi_{jk} \cap A \neq \emptyset,\ \operatorname{supp}\phi_{jk} \subset [0;\,\infty)\right\}.$$

$\mathcal{K}_j^\psi(A)$, $\mathcal{G}_j^\psi(A)$ and $\mathcal{H}_j^\psi(A)$ are defined analogously. If A is an interval e.g. $A = [a;\,b]$ we will write $\mathcal{K}_j^\phi[a;\,b]$ rather than $\mathcal{K}_j^\phi([a;\,b])$.

Let us take the observed Poisson process G^n with intensity function nh with respect to $d\mu$, and let ν_h^n denote the intensity measure of that process. We consider the following estimator of f on the interval $[0;\,1]$.

$$\hat{f}_n = \sum_{k\in\mathcal{K}^\phi_{j_1(n)}[0;1]} \hat{\alpha}_{j_1(n)k}\phi_{j_1(n)k} + \sum_{j=j_1(n)}^{j_2(n)}\sum_{k\in\mathcal{K}^\psi_{j_1(n)}(\varepsilon;1]} \delta_S(\hat{\beta}_{jk},\lambda_j)\psi_{jk} + \sum_{j=j_1(n)}^{j_2(n)}\sum_{k\in\mathcal{G}^\psi_{j_1(n)}[0;\varepsilon]} \delta_S(\hat{\beta}_{jk},\mu_j)\psi_{jk}, \tag{4}$$

where

$$2^{j_1(n)} \asymp n^{3/(8\sigma+12)}, \qquad \hat{\alpha}_{j_1(n)k} = \frac{1}{n}\int_0^1 \tilde{\gamma}_{j_1(n)k}\,dG^n, \qquad \hat{\beta}_{jk} = \frac{1}{n}\int_0^1 \gamma_{jk}\,dG^n, \tag{5}$$

and the nonnegative sequences (λ_j) and (μ_j) define soft-threshold rules

$$\delta_S(\hat{\beta}_{jk},\lambda_j) = \operatorname{sgn}(\hat{\beta}_{jk})(|\hat{\beta}_{jk}| - \lambda_j)_+, \qquad \delta_S(\hat{\beta}_{jk},\mu_j) = \operatorname{sgn}(\hat{\beta}_{jk})(|\hat{\beta}_{jk}| - \mu_j)_+$$

($a(n) \asymp b(n)$ means $a(n)/b(n)$ remains bounded and cut away from zero as $n \to \infty$). The sequence $(j_2(n))$ will be specified later. The choice of the threshold levels (λ_j) and (μ_j) will be commented (with a reference to [2]) in the proof of following theorem.

Theorem 1. *Let $p \geqslant 1$ and $\sigma > 3/(2p)$. Suppose that $\varepsilon \in \left(0;\,\frac{1}{2}\right)$. Then, there exist positive constants D_1 and D_2 such that*

$$\inf_{\tilde{f}_n}\sup_{f\in\mathcal{F}_{\sigma pq}(M,\varepsilon)} \mathbf{E}_f\|\tilde{f}_n - f\|_{L^2}^2 \geqslant D_1 n^{-\frac{2\sigma}{2\sigma+3}},$$

where $\tilde{f}_n$ denotes any estimator of the intensity function f, and

$$\sup_{f\in\mathcal{F}_{\sigma pq}(M,\varepsilon)} \mathbf{E}_f\|\hat{f}_n - f\|_{L^2}^2 \leqslant D_2 n^{-\frac{2\sigma}{2\sigma+3}},$$

where $\hat{f}_n$ is the estimator defined in (4).

The theorem remains valid, if $f \in g_0 + \mathcal{F}_{\sigma pq}(M,\varepsilon)$, where $g_0 \geqslant 0$ is a known and fixed function with support in $[0;\,1]$.

Proof of Theorem 1. For $f \in \mathcal{F}_{\sigma pq}(M,\varepsilon)$, there exist $(\alpha_{j_1,k})$ and $(\beta_{j,k})$ such that

$$f = \sum_{k\in\mathcal{K}^\phi_{j_1}[0;1]} \alpha_{j_1k}\phi_{j_1k} + \sum_{j=j_1}^{\infty}\sum_{k\in\mathcal{K}^\psi_j[0;1]} \beta_{jk}\psi_{jk}$$

on the interval $[0;\,1]$.

It is easy to see that the lower bound for the risk of any estimator $\tilde{f}_n$, obtained in [2, Section 3.1 Proposition 1] remains valid for $f \in \mathcal{F}_{\sigma pq}(M, \varepsilon)$. Indeed, let us define $\mathcal{G}_{\sigma pq}(j, \varepsilon) \subset \mathcal{F}_{\sigma pq}(M, \varepsilon)$ with

$$\mathcal{G}_{\sigma pq}(j, \varepsilon) := \left\{ f_\omega \geqslant 0 : f_\omega = f_0 + \delta_j \sum_{k \in \mathcal{G}_j^\varepsilon(\psi)} \omega_k \psi_{jk} \right\},$$

where $f_0 \in \mathcal{F}_{\sigma pq}(M/2, \varepsilon)$, $\int_{1/2}^1 f_0(x)dx > 0$, $\delta_j = \frac{M}{2} 2^{-j(\sigma+\frac{1}{2})}$, $\omega_k \in \{0, 1\}$, $\mathcal{G}_j^\varepsilon(\psi) := \left\{k \in \mathbb{Z} : \operatorname{supp} \psi_{jk} \subset \left(\varepsilon, \frac{1}{2}\right]\right\}$.

Notice that all functions from $\mathcal{G}_{\sigma pq}(j, \varepsilon)$ are equal to f_0 on the interval $[0; \varepsilon]$ and, hence, they satisfy near the origin all the assumptions the function f_0 does. Using the Assuad lemma, it can be shown that there exists an absolute constant $J \in \mathbb{N}$ such that for all $j \geqslant J$ we have

$$\sup_{f \in \mathcal{F}_{\sigma pq}(M,\varepsilon)} \mathbf{E} \|\tilde{f}_n - f\|_{L^2}^2 \geqslant \sup_{f_\omega \in \mathcal{G}_{\sigma pq}(j,\varepsilon)} \mathbf{E}_{f_\omega} \|\tilde{f}_n - f_\omega\|_{L^2}^2 \geqslant D_3 2^{-2j\sigma} \exp\left[-D_4 n 2^{-j(2\sigma+3)}\right].$$

If we take $2^j \asymp n^{1/(2\sigma+3)}$, then we obtain

$$\sup_{f \in \mathcal{F}_{\sigma pq}(M,\varepsilon)} \mathbf{E} \|\tilde{f}_n - f\|_{L^2}^2 \geqslant D_1 n^{-\frac{2\sigma}{2\sigma+3}}.$$

Now we will show that the estimator $\hat{f}_n$ achieves the optimal rate of convergence. Our estimator is a sum of three terms. The first two terms are almost the same as in [2], with one small difference: the first resolution level $j_1(n)$ depends on the size of the experiment. The last term consists of only those wavelets that have supports contained in the interval $[0; \varepsilon]$.

Notice that the estimator $\hat{f}_n$ does not use the wavelets with supports not entirely contained in $[0; \infty]$. The reason of this is that there is a problem with an upper bound for the error of the respective coefficients. However, we will show that with the assumption, $f(x) = O(x^{(4\sigma/3-1/2)_+})$ as $x \to 0^+$, the estimator $\hat{f}_n$ achieves the optimal rate of convergence in spite of lack of estimating those coefficients.

Let us evaluate the risk of the estimator (4) for $f \in \mathcal{F}_{\sigma pq}(M, \varepsilon)$.

$$\begin{aligned}
\mathbf{E}\|\hat{f}_n - f\|_{L^2}^2 \leqslant & \sum_{k \in \mathcal{H}_{j_1(n)}^\phi[0;1]} \mathbf{E}[\hat{\alpha}_{j_1(n)k} - \alpha_{j_1(n)k}]^2 + \sum_{j=j_1(n)}^{j_2(n)} \sum_{k \in \mathcal{K}_j^\psi(\varepsilon;1]} \mathbf{E}[\delta_S(\hat{\beta}_{jk}, \lambda_j) - \beta_{jk}]^2 \\
& + \sum_{j=j_1(n)}^{j_2(n)} \sum_{k \in \mathcal{G}_j^\psi[0;\varepsilon]} \mathbf{E}[\delta_S(\hat{\beta}_{jk}, \mu_j) - \beta_{jk}]^2 + \sum_{j=j_2(n)+1}^{\infty} \sum_{k \in \mathcal{H}_j^\psi[0;1]} \beta_{jk}^2 \\
& + \int_0^{N2^{-j_1(n)}} f^2(x)dx := \hat{L}_n(f) + \hat{S}_n(f, (\varepsilon; 1]) + \hat{S}_n(f, [0; \varepsilon]) + \hat{T}_n(f) + \hat{R}_n(f). \qquad (6)
\end{aligned}$$

The estimator $\hat{f}_n$ consists of three terms that estimate different components of the function f. Note that $\hat{R}_n(f)$ is an upper bound for the risk for a component, say f_1, that consists of only those wavelets with supports not contained in $[0; \infty]$. The support of f_1 is the interval $[0; (N-1)2^{-j_1(n)}]$ and the wavelets coefficients of that function are estimated in (4) by zeros, so the component f_1 is estimated by the zero function and the risk of that component is equal to $\|f_1\|_{L^2}$, which is smaller than $\|f \cdot \mathbb{1}_{[0;N2^{-j_1(n)}]}\|_{L^2}$.

It is known (see [10, Chapter 3.2]), that

$$\begin{aligned}
&\mathbf{E}\hat{\alpha}_{j_1(n)k} = \frac{1}{n} \int_0^1 \tilde{\gamma}_{j_1(n)k} d\nu_h^n = \langle f, \phi_{j_1(n)k} \rangle := \alpha_{j_1(n)k}, \\
&\mathbf{E}\hat{\beta}_{jk} = \frac{1}{n} \int_0^1 \gamma_{jk} d\nu_h^n = \langle f, \psi_{jk} \rangle := \beta_{jk}, \\
&\operatorname{Var} \hat{\alpha}_{j_1(n)k} = \frac{1}{n} \int_0^1 \tilde{\gamma}_{j_1(n)k}^2 d\nu_h. \\
&\operatorname{Var} \hat{\beta}_{jk} = \frac{1}{n} \int_0^1 \gamma_{jk}^2 d\nu_h.
\end{aligned}$$

Since $\phi \in C^2$ and $\operatorname{supp} \phi = [0; N]$, it is easy to see that

$$\lim_{x \to 0} \phi'(x)/x = 0,$$

so the function $\tilde{\gamma}$ is bounded on $[0;1]$. With $k \in \mathcal{H}_j^{\phi}[0;1]$ and using (3) we have

$$\int_0^1 \tilde{\gamma}_{j,k}^2 d\nu_h \leqslant 2^{3j} C_1. \tag{7}$$

Using (5) and (7) we obtain

$$\hat{L}_n(f) \leqslant C_2 \frac{1}{n} \sum_{k \in \mathcal{H}_{j_1(n)}^{\phi}[0;1]} 2^{3j_1(n)} \leqslant C_3 n^{-1} 2^{4j_1(n)} \leqslant C_4 n^{-\frac{2\sigma}{2\sigma+3}}. \tag{8}$$

For $\hat{T}_n(f)$, we have

$$\hat{T}_n(f) \leqslant \sup\{\hat{T}_n(f), f \in \mathcal{F}_{\sigma pq}(M,\varepsilon)\} = C_5 2^{-2j_2(n)\left(\sigma+\frac{1}{2}-\frac{1}{p}\right)}. \tag{9}$$

$\hat{S}_n(f,(\varepsilon;1])$ is obtained in [2] for fixed j_1. Here $j_1(n)$ increase with n to infinity, so $\hat{S}_n(f,(\varepsilon;1])$ is smaller for large n and evaluations from [2] remain valid, so that for an appropriate choice of (λ_j) we have

$$\hat{S}_n(f,(\varepsilon;1]) \leqslant C_6 n^{-2} \log^3 n\, 2^{4j_2(n)} + C_7 n^{-\frac{2\sigma}{2\sigma+3}}. \tag{10}$$

To evaluate $\hat{S}_n(f,[0;\varepsilon])$, we use a Gaussian approximation of $\hat{\beta}_{jk}$ in the same way as in [2]. The difference is in the evaluation of $\|\gamma_{jk}\|_{L^\infty}$. Here, for $k \in \mathcal{G}_j^{\psi}[0;1]$, we have

$$\|\gamma_{jk}\|_{L^\infty} \leqslant C_8 2^{\frac{5}{2}j}.$$

Consequently, it can be shown that for an appropriate choice of (μ_j) we have

$$\hat{S}_n(f,[0;\varepsilon]) \leqslant C_9 n^{-2} \log^3 n\, 2^{6j_2(n)} + C_{10} n^{-\frac{4\sigma/3}{4\sigma/3+2}}. \tag{11}$$

Here we can see why the assumption that the function f is smoother near the origin is essentially necessary. Since $\|\gamma_{jk}\|_{L^\infty}$ goes to infinity (when $j \to \infty$) faster when the support of γ_{jk} is close to the origin, we have to cut more wavelet coefficients there, but if we do that, the estimator is locally "oversmoothed" so we need this assumption to sustain the minimax optimality. From a technical point of view the factor $4/3$ seems to be the lowest (we can always take it higher, but if we take lower, then we lose optimality). Finally we evaluate $\hat{R}_n(f)$. Since the function f tends to zero at least as fast as $x^{4\sigma/3-1/2}$ when $x \to 0^+$ and $\sigma > 3/8$, or it is bounded, when $\sigma \leqslant 3/8$, we have

$$\hat{R}_n(f) \leqslant N 2^{-j_1(n)} \max_{x \in [0;N2^{-j_1(n)}]} f^2(x) \leqslant C_{11} 2^{-2j_1(n)4\sigma/3} \leqslant C_{12} n^{-\frac{2\sigma}{2\sigma+3}}. \tag{12}$$

It is easy to see that for $2^{j_1(n)} \asymp n^{3/(8\sigma+12)}$ and for $\sigma > 3/8$ the assumption $f(x) = O(1)$, as $x \to 0^+$ is not enough for $\hat{R}_n(f)$ to attain the rate $n^{-2\sigma/(2\sigma+3)}$. To achieve that rate, the function f needs to approach zero at least as fast as $x^{4\sigma/3-1/2}$. One can see that the rate $x^{4\sigma/3-1/2}$ cannot be lower, because the upper bound for $\hat{R}_n(f)$ depends on the rate of $j_1(n)$, which is set to keep the upper bound for $\hat{L}_n(f)$ smaller than $C_4 n^{-\frac{2\sigma}{2\sigma+3}}$. Now we can choose $j_2(n)$ such that

$$\frac{\sigma \log_2 n}{(2\sigma+3)(\sigma+1/2-1/p)} \ll j_2(n) \ll \frac{\sigma+3}{3(2\sigma+3)} \log_2 n - \frac{1}{2} \log_2 \log n, \tag{13}$$

where $a_n \ll b_n$ means that $\lim_{n\to\infty} b_n - a_n = \infty$. It is possible when

$$\frac{\sigma}{(2\sigma+3)(\sigma+1/2-1/p)} < \frac{\sigma+3}{3(2\sigma+3)}.$$

Since $p \geqslant 1$ it is easy to check, that this condition is true, if we assume that $\sigma > 3/(2p)$. With that choice of $j_2(n)$ an using (6) and (8)–(12) we have

$$\mathbf{E}\|\hat{f}_n - f\|_{L^2}^2 \leqslant D_2 n^{-\frac{2\sigma}{2\sigma+3}}$$

which completes the proof. □

4. Adaptive estimator

In the previous section, we presented an estimator that achieves the optimal rate of convergence. Unfortunately, there is a problem with practical applications of this estimator because of its dependence on Besov space parameters (see conditions (5) and (13)). In this section, we will present an estimator of the intensity function f that does not depend on the parameters

σ, p, q, which are unknown in practical estimation problems. We will show that the risk of the estimator achieves almost the optimal rate of convergence (to within logarithmic factors). Since this estimation procedure achieves optimal rates of convergence along the whole scale of Besov spaces, it is adaptive.

Denote

$$J = \left\{(\sigma, p, q) : \frac{3}{2p} < \sigma < r_0,\ 1 \leqslant p, q \leqslant \infty\right\},$$

where $r_0 > 3/2$ is fixed. Consider the following estimator

$$\tilde{f}_n^T = \sum_{k \in \mathcal{H}^{\phi}_{j_3(n)}[0;1]} \hat{\alpha}_{j_3(n)k}\phi_{j_3(n)k} + \sum_{j=j_3(n)}^{j_4(n)} \sum_{k \in \mathcal{K}_j^{\psi}(\varepsilon;1]} \delta_H(\hat{\beta}_{jk}, Tc_j)\psi_{jk}. + \sum_{j=j_3(n)}^{j_4(n)} \sum_{k \in \mathcal{G}_j^{\psi}[0;\varepsilon]} \delta_H(\hat{\beta}_{jk}, Td_j)\psi_{jk}, \tag{14}$$

where the coefficients $\hat{\alpha}_{j_3(n)k}$ and $\hat{\beta}_{jk}$ are defined in (5),

$$c_j = 2^j\sqrt{\frac{j}{n}}, \qquad d_j = 2^{3j/2}\sqrt{\frac{j}{n}}, \qquad 2^{j_3(n)} \asymp n^{3/(8r_0+12)}, \qquad 2^{j_4(n)} \asymp \frac{n^{1/2}}{\log_2 n}$$

and

$$\delta_H(\hat{\beta}_{jk}, T) = \begin{cases} \hat{\beta}_{jk}, & \text{if} |\hat{\beta}_{jk}| > T \\ 0, & \text{if } |\hat{\beta}_{jk}| \leqslant T \end{cases}$$

is a hard-threshold rule. We will prove the following theorem.

Theorem 2. *Let $(\sigma, p, q) \in J$. Then*

$$\forall\, \varepsilon \in \left(0; \frac{1}{2}\right) \quad \sup_{f \in \mathcal{F}_{\sigma pq}(M,\varepsilon)} \mathbf{E}_f \|\tilde{f}_n^T - f\|_{L^2}^2 \leqslant C\left(\frac{\log n}{n}\right)^{\frac{2\sigma}{2\sigma+3}}.$$

Proof of Theorem 2. Let us evaluate the risk of the estimator (14) for $f \in \mathcal{F}_{\sigma pq}(M, \varepsilon)$.

$$\begin{aligned}\mathbf{E}\|\tilde{f}_n^T - f\|_{L^2}^2 \leqslant & \sum_{k \in \mathcal{H}^{\phi}_{j_3(n)}[0;1]} \mathbf{E}[\hat{\alpha}_{j_3(n)k} - \alpha_{j_3(n)k}]^2 + \sum_{j=j_3(n)}^{j_4(n)} \sum_{k \in \mathcal{K}_j^{\psi}(\varepsilon;1]} \mathbf{E}[\delta_H(\hat{\beta}_{jk}, Tc_j) - \beta_{jk}]^2 \\ & + \sum_{j=j_3(n)}^{j_4(n)} \sum_{k \in \mathcal{G}_j^{\psi}[0;\varepsilon]} \mathbf{E}[\delta_H(\hat{\beta}_{jk}, Td_j) - \beta_{jk}]^2 + \sum_{j=j_4(n)+1}^{\infty} \sum_{k \in \mathcal{H}_j^{\phi}[0;1]} \beta_{jk}^2 \\ & + \int_0^{N2^{-j_3(n)}} f^2(x)dx := \tilde{L}_n(f) + \tilde{S}_n(f, (\varepsilon; 1]) + \tilde{S}_n(f, [0; \varepsilon]) + \tilde{T}_n(f) + \tilde{R}_n(f).\end{aligned} \tag{15}$$

Using (7) we obtain

$$\tilde{L}_n(f) \leqslant C_1\frac{1}{n} \sum_{k \in \mathcal{H}^{\phi}_{j_3(n)}[0;1]} 2^{3j_3(n)} \leqslant C_2 n^{-1}2^{4j_3(n)} \leqslant C_3 n^{-\frac{2r_0}{2r_0+3}} \leqslant C_3 n^{-\frac{2\sigma}{2\sigma+3}}. \tag{16}$$

For $\tilde{T}_n(f)$ we have

$$\tilde{T}_n(f) \leqslant C_4 2^{-2j_4(n)\left(\sigma+\frac{1}{2}-\frac{1}{p}\right)} \leqslant C_4 n^{-\frac{2\sigma}{2\sigma+3}}. \tag{17}$$

Let us evaluate $\tilde{S}_n(f, (\varepsilon; 1])$ and $\tilde{S}_n(f, [0; \varepsilon])$. We choose constant C_5 that for $k \in \mathcal{K}_j^{\psi}(\varepsilon; 1]$

$$\int_0^1 \gamma_{jk}^2 d\nu_h^n \leqslant C_5\varepsilon^{-2}2^{2j}n \quad \text{and} \quad \|\gamma_{jk}\|_{L^\infty} \leqslant C_5\varepsilon^{-1}2^{\frac{3}{2}j},$$

and for $k \in \mathcal{G}_j^{\psi}[0; \varepsilon]$

$$\int_0^1 \gamma_{jk}^2 d\nu_h^n \leqslant C_5\varepsilon^{-2}2^{3j}n \quad \text{and} \quad \|\gamma_{jk}\|_{L^\infty} \leqslant C_5\varepsilon^{-1}2^{\frac{5}{2}j}.$$

If we take $T = C_6\eta$ where $C_6^2 \geqslant 8C_5(1 + C_6/6)\log 2$ and $\eta = \max\{8\varepsilon^{-1}, (8r_0/3 + 2)\varepsilon^{-1}\}$ then using [2, Eq. (32)], [2, Lemma 3] and [4, Theorem 3] we have

$$\tilde{S}_n(f, (\varepsilon; 1]) \leqslant C_7 \left(\frac{\log n}{n}\right)^{\frac{2\sigma}{2\sigma+3}}, \tag{18}$$

and

$$\tilde{S}_n(f, [0; \varepsilon]) \leqslant C_8 \left(\frac{\log n}{n}\right)^{\frac{2\sigma}{2\sigma+3}}. \tag{19}$$

Finally we evaluate $\tilde{R}_n(f)$. Since the function f tends to zero at least as fast as $x^{4r_0/3-1/2}$, when $x \to 0^+$, we have

$$\tilde{R}_n(f) \leqslant N2^{-j_3(n)} \max_{x\in[0;N2^{-j_3(n)}]} f^2(x) \leqslant C_9 2^{-2j_3(n)4r_0/3} \leqslant C_{10} n^{-\frac{2\sigma}{2\sigma+3}}. \tag{20}$$

Using (15)–(20) we prove Theorem 2. □

5. Empirical risk minimization and numerical experiment

In this section, we will use an empirical risk minimization principle (ERM) for the choice of the parameter T in the estimator (14). We will also use that estimator in the numerical experiment. The best possible choice of the parameter T is that which minimizes the true L_2 error of the estimator $\tilde{f}_n^T$, which depends, however, on the unknown intensity f. The idea is now to approximate the unknown error, using the observations from the experiment, and to choose the parameter T that minimizes that approximated error. Let us denote

$$R = \|f - \tilde{f}_n^T\|_2^2 - \|f\|_2^2 = \|\tilde{f}_n^T\|_2^2 - 2\langle f, \tilde{f}_n^T\rangle.$$

Using (1) we have

$$R = \|\tilde{f}_n^T\|_2^2 - 2\langle G^{-1}g, \tilde{f}_n^T\rangle = \|\tilde{f}_n^T\|_2^2 + \int_0^1 \tilde{f}_n^T(x)d(g(x)/x).$$

To approximate the integral above we will use a regular partition of the interval $[0; 1]$. Let $0 = x_0 < x_1 < \cdots < x_M = 1$; then R may be approximated with

$$\hat{R} = \|\tilde{f}_n^T\|_2^2 + \sum_{i=1}^{M-1} \tilde{f}_n^T(x_i)\left(\frac{\hat{g}(x_{i+1})}{x_{i+1}} - \frac{\hat{g}(x_i)}{x_i}\right),$$

where $\hat{g}$ denotes some estimator of the intensity of the observed process. In the numerical experiment, an Epanechnikov kernel estimator of function g, with a window width $h = \frac{1}{2}n_1^{-1/5}$ was used, where n_1 denotes the number of the observed line segments. We use the kernel estimator instead of wavelet one for several reasons. First: kernel estimators are simple in implementation and they are numerically less complex than wavelet ones. Second: kernel estimators have good properties when the estimated functions are regular, and here the function g is the integral of function f.

Notice that the lower $j_4(n)$ level is, the higher chance for the ERM procedure to choose the best possible parameter T, because there are less β_{jk} parameters and consequently there are less options for cutting them off by the hard-threshold rule.

In the numerical example the Daubechies wavelets "DB8" (see [8, Chapter 7.1]), with the support length $N = 15$, were used. Four values of the experiment size n were used: 10^4, 10^5, 10^6 and 10^7. The minimal resolution levels were $j_3(10^4) = 3$, $j_3(10^5) = 3$, $j_3(10^6) = 4$, $j_3(10^7) = 4$ and the maximal resolution levels were $j_4(10^4) = 3$, $j_4(10^5) = 4$, $j_4(10^6) = 5$, $j_4(10^7) = 5$. The resolution level $j_3(n)$ should not be too low, because the estimator is made of only one "father wavelet" on the interval $[0; 2^{-j_3(n)}]$ which is rather inflexible. It cannot be too high, neither, because of the variance of the coefficients. The resolution level $j_4(n)$ should be as high as possible, but if it is too high, then the variances of the $\hat{\beta}_{j_4(n)k}$ parameters will be too high and the ERM procedure will choose such high parameter T that the hard-threshold rule will cut off all of the "mother wavelets" coefficients. The values used in the experiment have been selected subjectively to compromise those conflicting requirement. The minimization was performed through a grid search with $T \in [0; 10]$ and with step 0.01 and ε equal to 0.2. For wavelets derivatives calculation, a dyadic discretization of wavelets and a difference quotient were used. For obtaining wavelets derivative values in the points between the discretization points, a linear interpolation was used. The distance between discretization points of the functions ϕ_{jk}, ψ_{jk}, ϕ'_{jk}, ψ'_{jk} was $2^{-(j+10)}$.

For the simulation we set the expected number of sphere centers per unit volume to one, so the density of the random radius was equal to the intensity function f. The random radii ware generated from the following density functions:

- Beta(3, 2):

$$f(x) = 12x^2(1 - x).$$

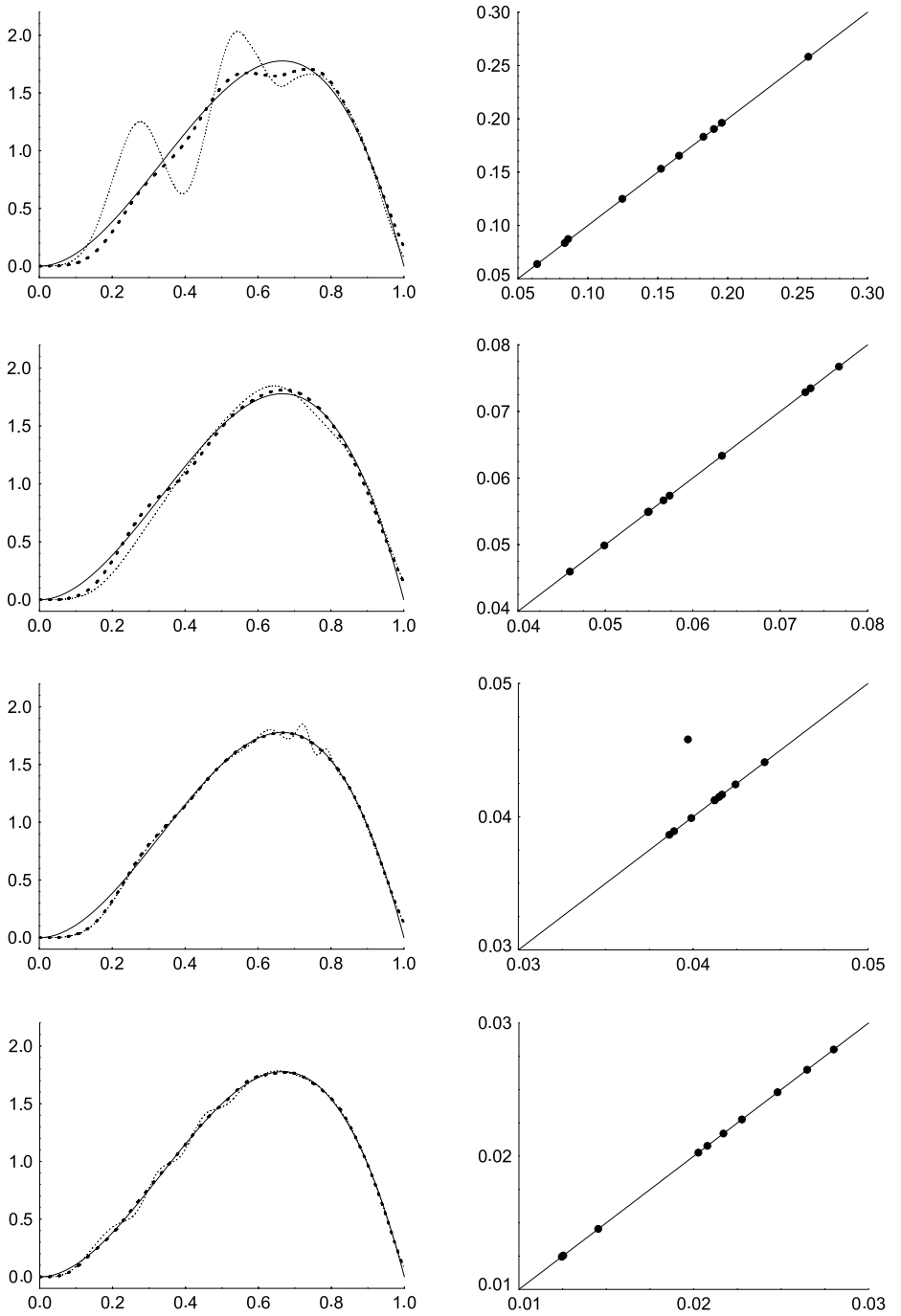

Fig. 1. Worst (dashed) and best (thick dashed) reconstruction (out of 10 data samples) of Beta(3, 2) intensity function (solid line) for the experiment size, from top: $n = 10^4$, $n = 10^5$, $n = 10^6$, $n = 10^7$. On the right side the scatterplots of L^2 error of ERM solutions versus those of the best possible solutions.

- Bimodal:

$$f(x) = \frac{28125}{512}x^2(0.8 - x)^2 \cdot \mathbb{1}_{[0;0.8]}(x) + \frac{9375}{8}(0.6 - x)^2(1 - x)^2 \cdot \mathbb{1}_{[0.6;1]}(x).$$

- Step function (cf. [5]):

$$f(x) = 0.6 \cdot \mathbb{1}_{[0;1/3]}(x) + 0.9 \cdot \mathbb{1}_{(1/3;3/4]}(x) + 1.7 \cdot \mathbb{1}_{(3/4;1]}(x).$$

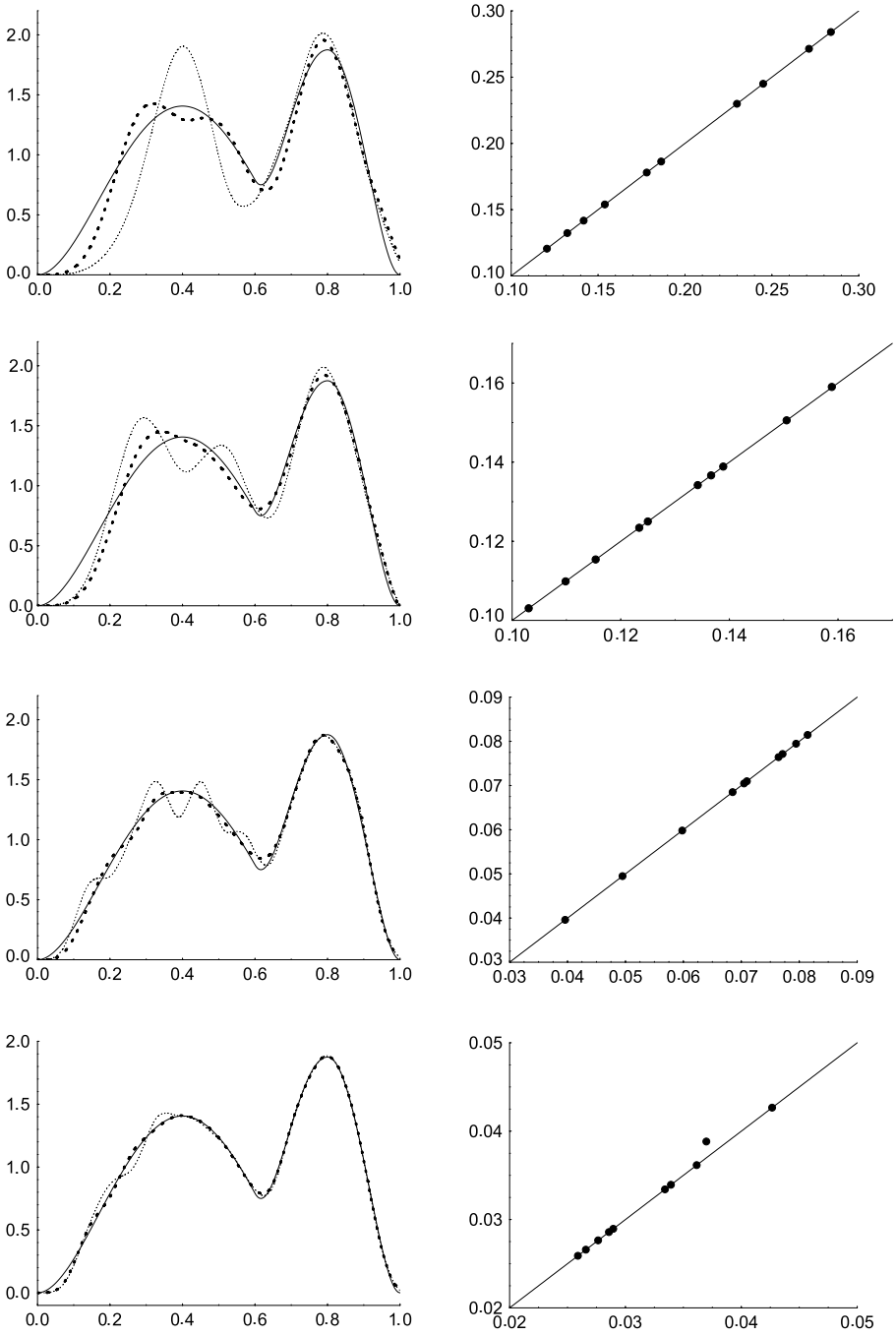

Fig. 2. Similar to Fig. 1 but for the Bimodal intensity function.

- Swapped Minerbo–Levy (cf. [5]):

$$f(x) = 4x^2 \cdot \mathbb{1}_{[0;0.5]}(x) + (2 - 4(1-x)^2) \cdot \mathbb{1}_{(0.5;1]}(x).$$

Let X_R be the radius of a random sphere, with the distance X_d of its center from the probing line. Then, for each radius, we generated X_d from the appropriate distribution (because the process of random sphere center is homogeneous, the density of X_d is $f_d(x) = 2x$ on the interval $[0, 1]$). Denote with X_r the radius of the intersection. If X_R is smaller than X_d, then we do not observe any line segment and the ball gets lost. In the other case $X_r = (X_R^2 - X_d^2)^{1/2}$. We estimate f using only observed line segment radii X_r, which means that the size of the observed sample is, in fact, smaller than n. In the estimation procedure n is treated as known and the normalization of f as unknown.

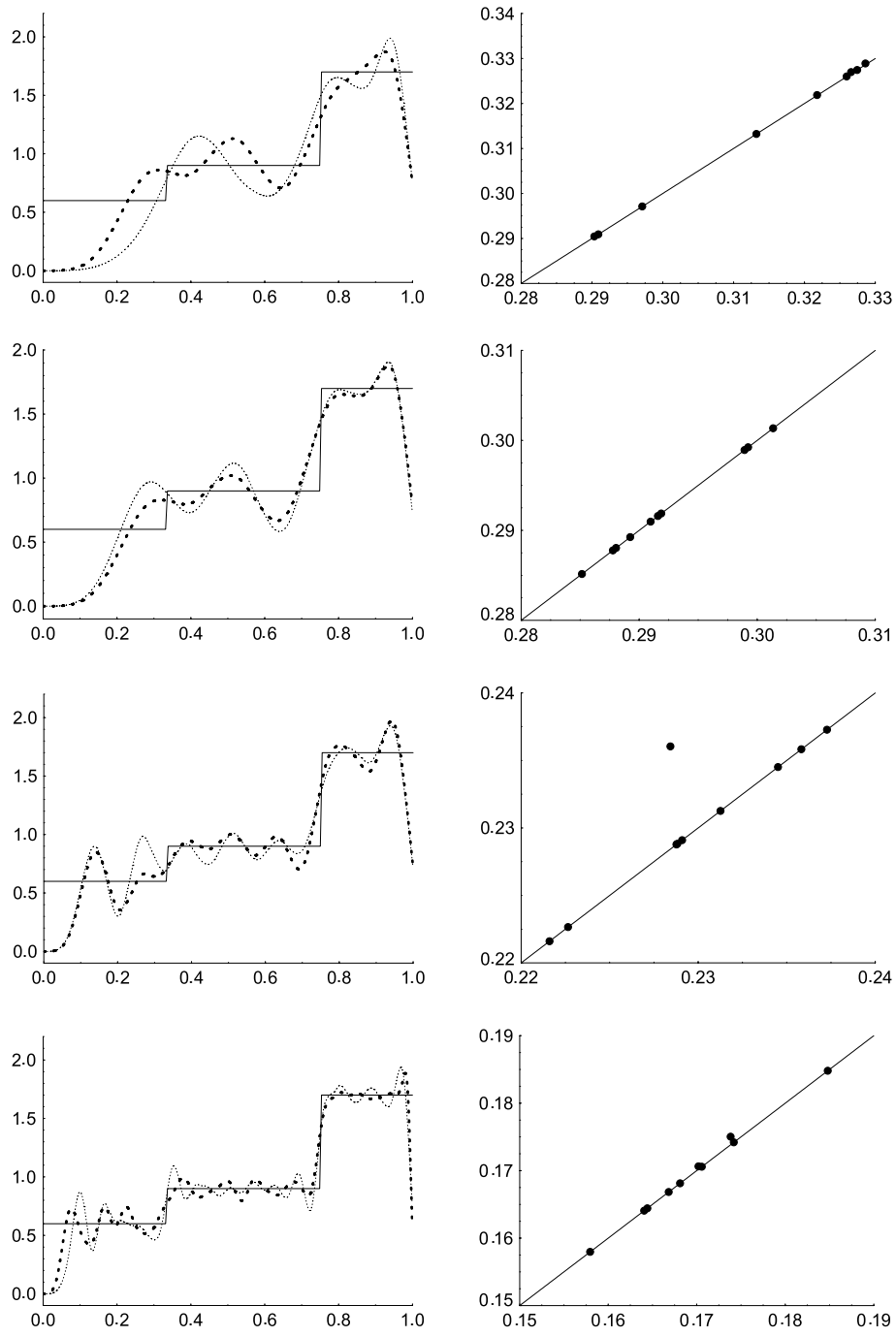

Fig. 3. Similar to Fig. 1 but for the Step intensity function.

The estimators $\tilde{f}_n^T$ were constructed for 10 artificial data samples generated for each function, with T selected to minimize $\hat{R}$. Also, the parameters T, that minimize the true L_2 error of the estimator were found on the same grid for each data sample. The best and worst of 10 data sample estimators are presented in Figs. 1–4.

Clear improvement of the estimator is seen, when n increases from 10^4 to 10^7. The ERM procedure seems to work very well—it chooses the best possible parameter T very often. It can also be observed, that for $n = 10^7$ the estimators of Beta(3, 2), Bimodal and Swapped Minerbo–Levy functions are very close to the true functions. The estimator of the step function seems to have higher risk but this particular function does not satisfy the conditions of Theorem 2 and was included in the experiment to check the behavior of the estimator in that case. The simulation results are quite similar to those in [5]

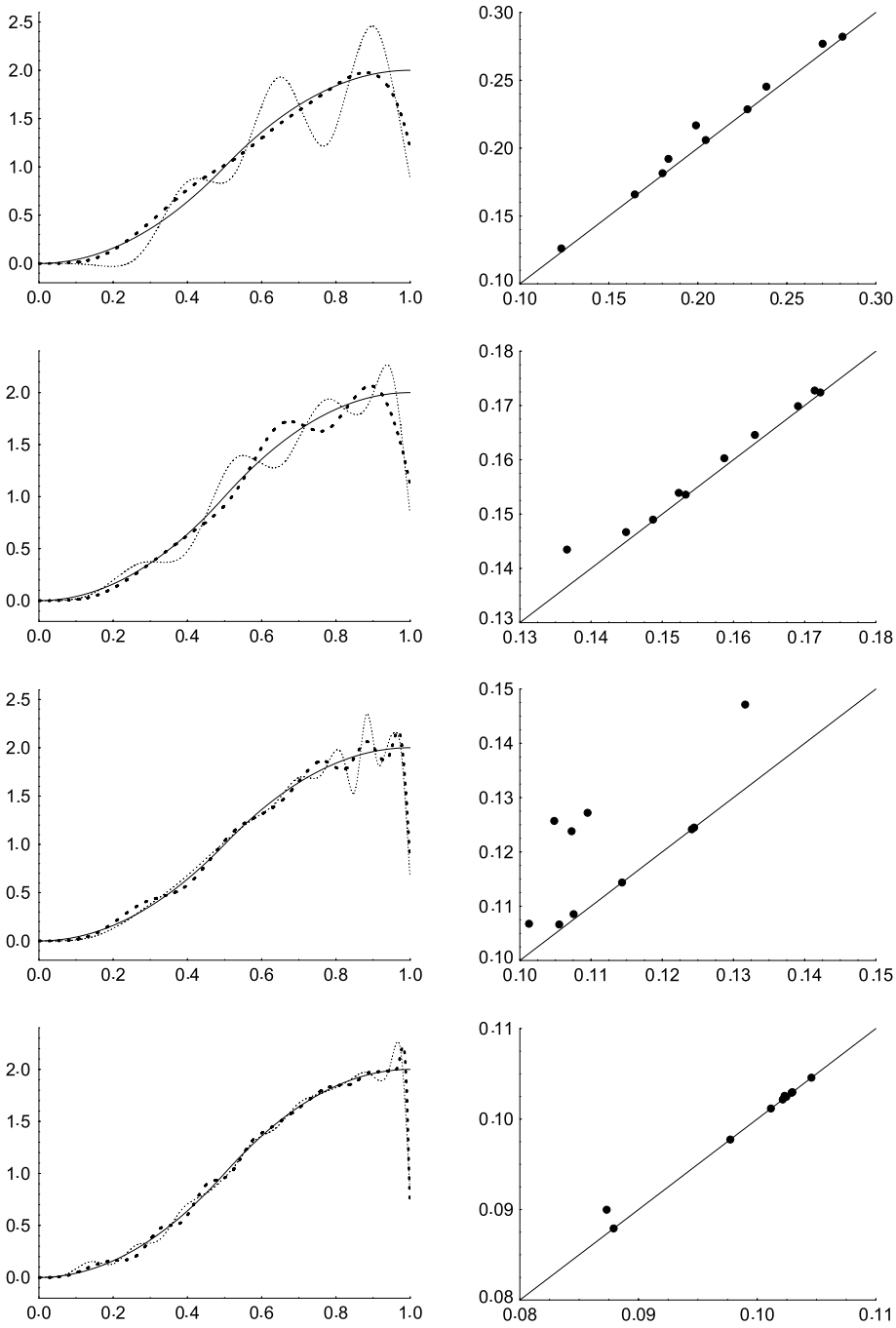

Fig. 4. Similar to Fig. 1 but for the Swapped Minerbo–Levy intensity function.

for $n = 10^4$. Unfortunately, there are no simulations for higher n in that paper. However, there are simulations for the experiment size $n = 2000$ which is too low for the wavelet estimator (for $n = 2000$ the lower resolution level $j_3(n)$ should not be higher than 2, because of the sample variance, and consequently the estimator would be made of only one "father wavelet" on the interval $[0; 1/4]$ and would be made of four "father wavelets" altogether). For $n = 2000$ the estimator $\tilde{f}_n^T$ works only for very regular functions, like Beta(3, 2). For the other functions the estimator $\tilde{f}_n^T$ had problems with their local behavior. Notice that our method is aimed at function classes different from those used in [5], the latter being expressed in terms of an expansion in series of singular functions of the specific folding operator. Here we have Besov classes which are independent of the operator. One can also give examples of functions that belong to our classes but not to the class defined in [5], for instance all discontinuous functions from some Besov space that satisfy the tail condition for model $\mathcal{F}_{\sigma pq}(M, \varepsilon)$.

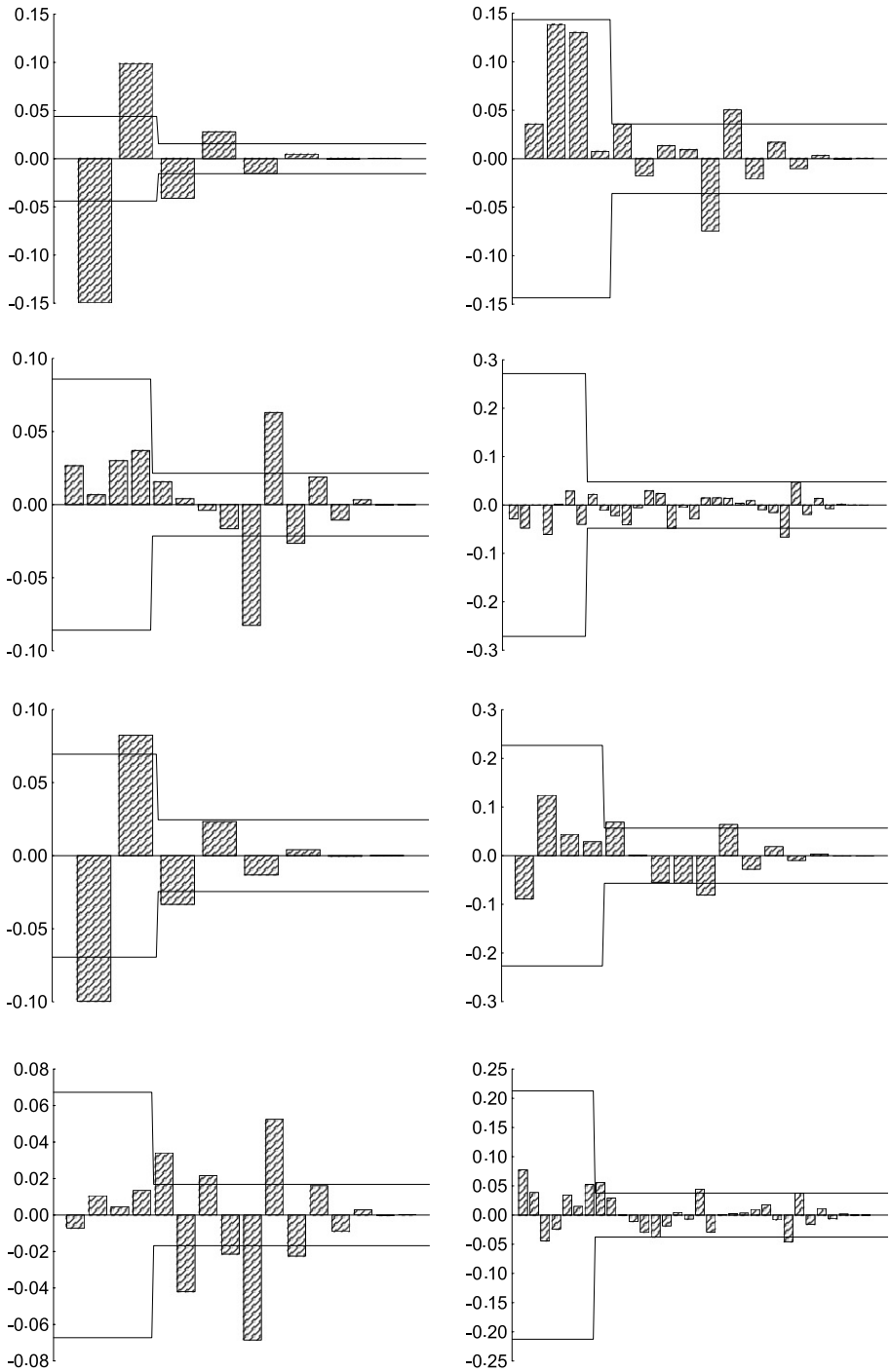

Fig. 5. An example of β_{jk} coefficients with hard-threshold levels. From top: Swapped Minerbo–Levy intensity function for $n = 10^5$ on the left side $j = 3$, on the right side $j = 4$; Swapped Minerbo–Levy intensity function for $n = 10^7$ on the left side $j = 4$, on the right side $j = 5$; step intensity function for $n = 10^5$ on the left side $j = 3$, on the right side $j = 4$; step intensity function for $n = 10^7$ on the left side $j = 4$, on the right side $j = 5$.

Examples of β_{jk} coefficients with hard-threshold levels are presented in Fig. 5 (there are 2^j coefficients on the resolution level j). It can be seen that the threshold levels on the interval $[0; \varepsilon]$ are higher than those on the interval $(\varepsilon; 1]$ (see the threshold levels in (14)). It also can be observed that the more regular function f is, the lower values of β_{jk} coefficients are and there is a greater chance to cut them off (for the Beta(3, 2) function the values of β_{jk} coefficients are very low and the ERM procedure chooses such parameter T that the hard-threshold rule cuts all of them).

Acknowledgments

The author would like to thank the anonymous reviewers for their valuable comments and suggestions to improve the paper.

The author wishes to express his gratitude to Profesor Zbigniew Szkutnik for many helpful comments and his active interest in the publication of this paper.

This research is supported by MNiSW Grant N-N201-609440.

The author would also like to convey thanks to the Polish Ministry of Science and Higher Education for providing the financial support.

References

[1] M. Barthel, P. Klimanek, D. Stoyan, Stereological substructure analysis in hot-deformed metals from TEM-images, Czechoslovak Journal of Physics Section B 35 (1985) 265–268.
[2] B. Ćmiel, Adaptive wavelet shrinkage approach to the Spector–Lord–Willis problem, Journal of Multivariate Analysis 101 (2010) 1458–1470.
[3] D.L. Donoho, Nonlinear solution of linear inverse problems by wavelet–vaguelette decomposition, Applied and Computational Harmonic Analysis 2 (1995) 101–126.
[4] D.L. Donoho, I.M. Johnstone, G. Kerkyacharian, D. Picard, Universal near minimaxity of wavelet shrinkage, in: Festschrift for Lucien Le Cam, Springer-Verlag, New York, 1996.
[5] A. Dudek, Z. Szkutnik, Minimax unfolding spheres size distribution from linear sections, Statistica Sinica 18 (2008) 1063–1080.
[6] R.M. German, P. Suri, S.J. Park, Review: liquid phase sintering, Journal of Materials Science 44 (2009) 1–39.
[7] J.-H. Han, D.-Y. Kim, Determination of three-dimensional grain size distribution by linear intercept measurement, Acta Materialia 46 (1998) 2021–2028.
[8] W. Härdle, G. Kerkyacharian, D. Picard, A.B. Tsybakov, Wavelets, Approximation and Statistical Applications, Springer-Verlag, New York, 1998.
[9] J.E. Hilliard, L.R. Lawson, Stereology and Stochastic Geometry, Kluwer, Dordrecht, 2003.
[10] J.F.C. Kingman, Poisson Processes, University Press, Oxford, 1993.
[11] J.F. Lamond, Significance of tests and properties of concrete and concrete-making materials, ASTM International (2006).
[12] Z. Szkutnik, Unfolding spheres size distribution from linear sections with *B*-splines and EMDS algorithm, Opuscula Mathematica 27 (2007) 151–165.
[13] Z. Szkutnik, A note on minimax rates of convergence in the Spektor–Lord–Willis problem, Opuscula Mathematica 30 (2010) 203–207.