

MINIMAX UNFOLDING OF THE SPHERES SIZE DISTRIBUTION FROM LINEAR SECTIONS

Anna Dudek and Zbigniew Szkutnik

AGH University of Science and Technology

Abstract: The stereological problem of unfolding the spheres size distribution from linear sections is analysed as a statistical inverse problem of estimation of a Poisson process intensity function from indirectly observed and binned data. Using suitably constructed singular value decomposition of the folding operator, a spectral estimator is constructed that is, up to a logarithmic factor, asymptotically rate minimax over a Sobolev-type class of functions. Finite sample behaviour of the estimator is demonstrated in a small numerical experiment.

Key words and phrases: Discretization, empirical risk minimization, ill-posed inverse problem, rate minimaxity, singular value decomposition, stereology.

1. Introduction

Consider a population of spheres embedded in an opaque medium. Assume that the centers of the spheres form a homogeneous Poisson process on \mathbb{R}^3 with the expected number of c points per unit volume. The radii x of the spheres are random with a distribution Q on $[0, 1]$, independent of the centers and absolutely continuous (with respect to dx) with a probability density function $q(x)$. An experimenter is interested in both c and q , but the spheres are not directly observable. Instead, a linear section through the medium is taken and the experimenter observes the line segments that are intersections of the line with the spheres. It can be shown (see, e.g., Szkutnik (2007)) that the observed radii y of the line segments form a Poisson point process on $[0, 1]$ with intensity function (with respect to dy) of the form $nh(y)$, where n is the 'size of the experiment', related to the total length of the observed portion of the linear section through the medium, and

$$h(y) = 2y \int_y^1 u(x) dx, \quad (1.1)$$

with $u(x) = cq(x)$. Notice that $nu(x)$ is the intensity function (with respect to dx) of the unobservable Poisson point process of spheres radii. Given observed sections radii, the goal is to unfold u . Asymptotics will be studied with n increasing to infinity.

Equations similar to (1.1) were first derived by Spektor (1950) and Lord and Willis (1951) as models of some measurements in material sciences. For an application in metallurgy see, e.g., Barthel, Klimanek and Stoyan (1985). In the sequel, the problem of unfolding u from linear sections data will be called the Spektor-Lord-Willis (SLW) problem. Equation (1.1) may also serve in isotropic cases as a model of linear intercept measurements on polished metallographic sections (cf., methodological remarks in Hilliard and Lawson (2003, p.117)), and a discussion of practical importance of such measurements in modelling, e.g., sintering processes in Han and Kim (1998)). The SLW problem was also studied in some detail by Stoyan, Kendall and Mecke (1987, pp.296-299), who reviewed some heuristic algorithms traditionally used for unfolding u , and discussed relations to the better known Wicksell problem of unfolding u from planar sections, and by Szkutnik (2007), who proposed a strongly consistent, sieved maximum likelihood estimator based on B-splines and the EMDS algorithm, and studied its convergence rates in $L^2([0, 1], dx)$.

Clearly, the operator defined by (1.1) is a compact Hilbert-Schmidt operator in $L^2([0, 1], dx)$. Consequently, its inverse is not bounded and the unfolding problem is ill-posed in the Hadamard sense. As noted by Szkutnik (2007), inverse estimation of u in $L^2([0, 1], dx)$ roughly corresponds to direct estimation of the intensity h in $L^2([0, 1], y^{-4} dy)$ and of its derivative in $L^2([0, 1], y^{-2} dy)$, which illustrates the statistical difficulty of the problem.

In the sequel, it will be assumed that the observed data are given in a discrete, binned form. Let $[0, 1] = B_1 \cup \dots \cup B_N$ be a partition of the data space into N disjoint bins. The observed data $[n_1, \dots, n_N]$ consist of the counts n_j of the line segments radii observed in the bins B_j . Discretization effects in linear inverse problems with compact operators were studied by Johnstone and Silverman (1991), henceforth JS91, who proposed a spectral type estimator that is asymptotically rate minimax, at least among linear estimators, over classes of functions defined in terms of singular functions of the folding operator. Their results do not directly apply, however, to the SLW problem with u considered as an element of $L^2([0, 1], dx)$. Therefore, in Section 2, the SLW problem is reconsidered with suitably chosen dominating measures, and a singular value decomposition (SVD) of the folding operator is derived. This is used in Section 3 for the derivation of lower bounds for the convergence rates of any estimator, and in Section 4 for the construction of the estimator and for proving its minimaxity with Johnstone and Silverman techniques. Finite sample behaviour of the estimator with a data-driven choice of parameters is studied in a simulation experiment described in Section 5.

2. SVD of the Operator, Discretization and Classes of Functions

Ill-posedness of the SLW problem implies that some sort of regularization is necessary. Our approach is to diagonalize the folding operator by finding its

SVD and to suitably dump the estimated Fourier coefficients of the unfolded function with respect to the basis of singular functions—the idea studied in detail by Johnstone and Silverman (1990, 1991). Let us recall that the SVD of a compact (and, for simplicity, invertible and such that $\overline{\text{Im } \mathcal{K}} = H_2$) operator \mathcal{K} , acting between two Hilbert spaces H_1 and H_2 , is a triple consisting of a sequence $\{b_\nu\}$ of positive numbers and of two orthonormal Schauder bases: $\{\phi_\nu\}$ in H_1 and $\{\psi_\nu\}$ in H_2 , such that $\mathcal{K}\phi_\nu = b_\nu\psi_\nu$. The b_ν 's are called the singular values of \mathcal{K} . If H_1 and H_2 are function spaces, then $\{\phi_\nu\}$ and $\{\psi_\nu\}$ are called, respectively, right and left singular functions.

The SVD of the operator (1.1), acting in $L^2([0, 1], dx)$, was constructed by Szkutnik (2007). With z_ν being the positive zeroes of the Bessel function $J_{-1/4}(z)$, it was found that the b_ν decay asymptotically as ν^{-1} , and that $\phi_\nu(x)$ and $\psi_\nu(y)$ are proportional, respectively, to $x^{3/2}J_{3/4}(z_\nu x^2)$ and $y^{3/2}J_{-1/4}(z_\nu y^2)$.

It is essential for the applicability of the Johnstone and Silverman technique that the discretization operator that projects L^2 -functions onto the subspace of step functions, constant in the bins B_j , satisfies the “matching SVD assumption”: given any ν_1 and ν_2 , the projections of ψ_{ν_1} and ψ_{ν_2} are either parallel or orthogonal. The SVD obtained by Szkutnik (2007) does not seem tractable in this respect. Therefore, in order to obtain a more tractable SVD, we change the dominating measures, both in the data space and in the solution space, and consider the folding operator \mathcal{K} as an operator from $L^2([0, 1], d\mu(x))$ to $L^2([0, 1], d\lambda(y))$. Another crucial postulate is that the Poisson process intensity function in the data space (with respect to λ) be bounded, which allows for setting an upper bound for the ratio of the so-called surrogate risk and the true risk (cf., JS91, p.9). For that, it is necessary that the left singular functions be bounded. In order to satisfy both requirements, we take $d\mu(x) = xdx$ and $d\lambda(y) = ydy$. The functions u and h are then replaced with $f(x) = u(x)/x$ and $g(y) = h(y)/y$, and the operator given in (1.1) becomes

$$g(y) = (\mathcal{K}f)(y) = 2 \int_y^1 f(x)d\mu(x). \quad (2.1)$$

The SVD of this operator can be found by solving the eigenproblem $\mathcal{K}^*\mathcal{K}f = \gamma f$, which is equivalent to the differential eigenvalue problem

$$\begin{cases} x^2 f'' - x f' + 4\gamma^{-1} x^4 f = 0 \\ f(0) = f'(1) = 0 \end{cases}.$$

This leads to Bessel functions of order $1/2$ which are expressible in terms of elementary functions. Standard calculations, similar to those presented in Szkutnik (2007), lead to the following.

Proposition 1. *The singular values of the operator (2.1) acting from $L^2([0, 1], \mu)$ to $L^2([0, 1], \lambda)$ are $b_\nu = 2/[\pi(2\nu + 1)]$, $\nu = 0, 1, \dots$, with the right singular functions $\phi_\nu(x) = 2 \sin[(2\nu + 1)\pi x^2/2]$ and the left singular functions $\psi_\nu(y) = 2 \cos[(2\nu + 1)\pi y^2/2]$.*

Let $B_j = [(j-1)/N)^{1/2}, (j/N)^{1/2}]$, $j = 1, \dots, N$, be the binning in the data space. Then $\lambda(B_j) = 1/(2N)$ for all j . The discretization operator, say \mathcal{P}_N , maps g to a step function $(\mathcal{P}_N g)(y) = \sum_j c_j \mathbf{1}_{B_j}(y)$, with $c_j = \int_{B_j} g d\lambda / \lambda(B_j)$. For two step functions, we have

$$\left\langle \sum_j c_j \mathbf{1}_{B_j}, \sum_j d_j \mathbf{1}_{B_j} \right\rangle_{L^2} = \frac{1}{2N} \sum_j c_j d_j. \tag{2.2}$$

In the sequel, to simplify notation, we identify step functions $\sum_j c_j \mathbf{1}_{B_j}$ with N -dimensional vectors $[c_1, \dots, c_N]$, considered as elements of \mathbb{R}^N with the rescaled inner product given in (2.2). It is easy to verify that, with

$$(\chi_\ell)_j = 2 \cos \frac{(2\ell + 1)(2j - 1)\pi}{4N} \quad \ell = 0, \dots, N - 1, \quad j = 1, \dots, N, \tag{2.3}$$

the vectors χ_ℓ form an orthonormal basis in that space. Moreover,

$$(\mathcal{P}_N \psi_\nu)_j = \operatorname{sinc} \frac{(2\nu + 1)\pi}{4N} \cdot 2 \cos \frac{(2\nu + 1)(2j - 1)\pi}{4N}. \tag{2.4}$$

This implies that if, for some integer k , either

$$\nu_2 - \nu_1 = 4Nk \quad \text{or} \quad \nu_2 + \nu_1 = 4Nk - 1, \tag{2.5}$$

or

$$\nu_2 - \nu_1 = 2N(2k + 1) \quad \text{or} \quad \nu_2 + \nu_1 = 2N(2k + 1) - 1, \tag{2.6}$$

then $\mathcal{P}_N \psi_{\nu_2}$ and $\mathcal{P}_N \psi_{\nu_1}$ are multiples of the same basis vector. (In case of (2.5), the cosines in (2.4) are the same for ν_1 and ν_2 and for all j ; in case of (2.6), they only differ in sign.) Define $[\nu] = \min\{\nu', 2N - 1 - \nu'\}$, where ν' is the residue of ν modulo $2N$. Then $[\nu] \in \{0, \dots, N - 1\}$ and $\mathcal{P}_N \psi_\nu = \gamma_\nu \chi_{[\nu]}$, with

$$\gamma_\nu = (-1)^{\mathbf{1}_{[N, 3N)}(\nu'')} \operatorname{sinc} \frac{(2\nu + 1)\pi}{4N},$$

where ν'' is the residue of ν modulo $4N$. This means that the matching SVD assumption is fulfilled and $\mathcal{P}_N \mathcal{K} \phi_\nu = b_\nu \gamma_\nu \chi_{[\nu]}$, which illustrates the smoothing action of $\mathcal{P}_N \mathcal{K}$ along the directions spanned by ϕ_ν 's. Define, for $\ell = 0, \dots, N - 1$, $\Gamma_\ell = \{\nu : \mathcal{P}_N \psi_\nu \text{ and } \mathcal{P}_N \psi_\ell \text{ are parallel}\}$ and consider

$$|b_\nu \gamma_\nu| = \frac{8N}{\pi^2(2\nu + 1)^2} \left| \operatorname{sinc} \frac{(2\nu + 1)\pi}{4N} \right|.$$

It follows from (2.5) and (2.6) that

$$\Gamma_\ell = \{2Ni + \ell : i = 0, 1, \dots\} \cup \{2Ni - 1 - \ell : i = 1, 2, \dots\}, \tag{2.7}$$

and it is easily seen that the moduli of the sine terms are constant within Γ_ℓ . Hence $\max_{\nu \in \Gamma_\ell} |b_\nu \gamma_\nu| = |b_\ell \gamma_\ell|$, which means that, among all components of f mapped by the operator $\mathcal{P}_N \mathcal{K}$ to the same one-dimensional subspace spanned by χ_ℓ , the least smoothed component is ϕ_ℓ . This will be crucial for the construction of the estimator in Section 4.

The function f is assumed to belong to the class

$$\mathcal{F}_{a,C} = \left\{ \sum_{\nu=0}^{\infty} f_\nu \phi_\nu : f_0 = 1, \sum_{\nu=1}^{\infty} (2\nu + 1)^{2a} f_\nu^2 \leq C^2 \right\}, \tag{2.8}$$

with some $a > 1/2$ and some C . Regularity of the functions belonging to $\mathcal{F}_{a,C}$ is described by the following proposition, proved in the Appendix.

Proposition 2. *Let k be a natural number.*

- a. *If $f \in \mathcal{F}_{a,C}$ with $a > k + 1/2$, then f is k times continuously differentiable in $[0, 1]$.*
- b. *If $f \in \mathcal{F}_{k,C}$, then f has k weak derivatives that are square integrable in $[0, 1]$ with respect to $dm(x) := x^{1/2} dx$.*

Since $a > 1/2$, it follows that $f \in \mathcal{F}_{a,C}$ is necessarily continuous on $[0, 1]$. Moreover, it is clear that $f(0) = 0$ and, using inequality (A.1) in the Appendix, it can be shown that $f(x) = O(x^2)$ as $x \rightarrow 0$, if $a > 3/2$. We also have the following.

Lemma 1. *For every $g = \mathcal{K}f$, with $f \in \mathcal{F}_{a,C}$,*

$$\left| \frac{g(y)}{b_0 \psi_0} - 1 \right| \leq \frac{C}{b_0} \sqrt{\sum_{k=1}^{\infty} \frac{1}{(2k + 1)^{2a}}}. \tag{2.9}$$

Proof. Obvious modification of the proof of Proposition 1 in JS91, combined with the bound $|\psi_\nu(x)/\psi_0(x)| \leq 2\nu + 1$, gives the result.

Lemma 1 implies, in particular, that if the constant C is small enough for the right-hand side of (2.9) to be smaller than one, then g is nonnegative, and there exist constants c_1 and c_2 such that, for any $g_1, g_2 \in \mathcal{K}\mathcal{F}_{a,C}$, one has $0 < c_1 \leq g_1/g_2 \leq c_2$.

3. Lower Bounds for the Convergence Rates

Define the risk of an estimator \tilde{f}_n as the mean integrated square error

$$M(\tilde{f}_n, f) = E_f \| \tilde{f}_n - f \|^2, \tag{3.1}$$

where $\|\cdot\|$ denotes the $L^2([0,1],\mu)$ norm. With $f \in \mathcal{F}_{a,C}$ one would expect the minimax convergence rates $n^{-2a/(2a+3)}$ (cf., e.g., Johnstone and Silverman (1990), or van Rooij and Ruymgaart (1996)). This is indeed the case, up to a log-factor, but some technical difficulties have to be resolved. The methodology based on the modulus of continuity and Fano's Lemma, as developed by Johnstone and Silverman (1990), does not work in our case because $g(1) = 0$ for all $g \in \mathcal{K}\mathcal{F}_{a,C}$, which causes problems with upper bounding the Kullback-Leibler divergence between densities by the corresponding L^2 -norm. With the approach based on van Trees inequality and developed by van Rooij and Ruymgaart (1996), we were able to obtain $n^{-2a/(2a+3)}(\log n)^{-1}$. This may be sharpened to $(n \log n)^{-2a/(2a+3)}$, using the Assouad cube technique. For any probability measures P, Q , denote by $\rho(P, Q)$ the Hellinger affinity between them; for two finite, binary sequences ω, ω' of the same length, denote by $\Delta(\omega, \omega')$ their Hamming distance. We use the following version of the Assouad Lemma (cf., Assouad (1983), or Birgé (2006)).

Lemma 2. *Let $\{P_\omega, \omega \in \mathcal{D}\}$ be a family of distributions indexed by $\mathcal{D} = \{0, 1\}^m$, and X_1, \dots, X_n an i.i.d. sample from a distribution in the family. Assume that $\rho(P_\omega, P_{\omega'}) \geq \bar{\rho}$ for each pair $(\omega, \omega') \in \mathcal{D}^2$ such that $\Delta(\omega, \omega') = 1$. Then, for any estimator $\hat{\omega}(X_1, \dots, X_n)$ with values in \mathcal{D} , $\sup_{\omega \in \mathcal{D}} \mathbb{E}_\omega [\Delta(\hat{\omega}, \omega)] \geq m\bar{\rho}^{2n}/4$, where \mathbb{E}_ω denotes the expectation when the X_i have the distribution P_ω .*

Proposition 3. *For the class of estimators*

$$\mathcal{T} = \{\tilde{f}_n : \mathbb{E}_f \|\tilde{f}_n\|^2 < \infty, f \in \mathcal{F}_{a,C}\}, \quad (3.2)$$

there exists a constant c such that

$$\inf_{\tilde{f}_n \in \mathcal{T}} \sup_{f \in \mathcal{F}_{a,C}} M(\tilde{f}_n, f) \geq c(n \log n)^{-\frac{2a}{2a+3}}.$$

Proof. In order to obtain a good lower bound, one should construct a possibly large number of well separated f_i 's in $\mathcal{F}_{a,C}$ for which the corresponding data distributions are close to each other. It is convenient to define f_j 's in terms of the singular functions, in order to describe the action of \mathcal{K} on f_j 's in a tractable way. Let b_k, ϕ_k and ψ_k be as in Proposition 1. For an integer $m = m(n)$, let $\omega = (\omega_1, \dots, \omega_m)$ with $\omega_i \in \{0, 1\}$, and set

$$f_\omega = \phi_0 + \delta_m \sum_{i=m}^{2m-1} \omega_{i-m+1} \phi_i$$

for some positive δ_m . Notice that $f_\omega \in \mathcal{F}_{a,C}$ for all ω , if $\delta_m^2 \sum_{i=m}^{2m-1} (2i+1)^{2a} \leq C^2$. It is thus sufficient that $(4m-1)^{2a+1} \leq C^2 \delta_m^{-2}$, and we take

$$\delta_m^2 \asymp m^{-(2a+1)}. \quad (3.3)$$

With $g_\omega = \mathcal{K}f_\omega$, one observes a Poisson process \mathcal{N}_{ng_ω} with intensity function ng_ω or, equivalently, n i.i.d. copies of a Poisson process \mathcal{N}_{g_ω} . Take $f_0 = \phi_0$, $g_0 = \mathcal{K}f_0$, and denote by $\mathcal{L}(\mathcal{N}_g)$ the distribution of \mathcal{N}_g . Then the Hellinger affinity between the distributions takes the form

$$\rho(\mathcal{L}(\mathcal{N}_{g_\omega}), \mathcal{L}(\mathcal{N}_{g_{\omega'}})) = \int \sqrt{\frac{d\mathcal{L}(\mathcal{N}_{g_\omega})}{d\mathcal{L}(\mathcal{N}_{g_0})} \frac{d\mathcal{L}(\mathcal{N}_{g_{\omega'}})}{d\mathcal{L}(\mathcal{N}_{g_0})}} d\mathcal{L}(\mathcal{N}_{g_0}) = \exp[-H^2(g_\omega, g_{\omega'})],$$

where $H^2(g_\omega, g_{\omega'}) = \int_0^1 (\sqrt{g_\omega} - \sqrt{g_{\omega'}})^2 d\lambda/2$ (cf., e.g., Reiss (1993, Chap.3.2)). With $\Delta(\omega, \omega') = 1$, one has $g_{\omega'} = g_\omega + \delta_m b_k \psi_k$, for some k between m and $2m - 1$, and with $g_\omega = b_0 \psi_0 + \delta_m \sum_{i=m}^{2m-1} \omega_{i-m+1} b_i \psi_i$. Standard calculation gives

$$H^2(g_\omega, g_{\omega'}) = \frac{\delta_m^2 b_k^2}{2b_0} \int_0^1 \frac{\psi_k^2}{\psi_0} \left(\sqrt{\frac{g_{\omega'}}{b_0 \psi_0}} + \sqrt{\frac{g_\omega}{b_0 \psi_0}} \right)^{-2} d\lambda.$$

The second factor under the integral is bounded and cut away from zero (see the remark after Lemma 1; C may be assumed small without loss of generality). The integral $\int \psi_k^2/\psi_0 d\lambda$ evaluates to $[\gamma + \psi(3/2 + 2k) + \log 4]/2$, where γ is Euler's constant and $\psi(\cdot)$ is the digamma function. Since (cf., Gradshteyn and Ryzhik (1980)) $\psi(3/2 + 2k) = -0.58 \dots + 2[\sum_{i=1}^{2k+1} 1/(2i - 1) - \log 2]$, one easily obtains $\int \psi_k^2/\psi_0 d\lambda \asymp \log(2k + 1)$ and, using (3.3),

$$H^2(g_\omega, g_{\omega'}) \asymp \delta_m^2 b_k^2 \log(2k + 1) \asymp \delta_m^2 m^{-2} \log m \asymp m^{-(2a+3)} \log m. \tag{3.4}$$

Now, for any estimator \tilde{f}_n of f , take $\tilde{\omega} \in \mathcal{D} = \{0, 1\}^m$ such that $\|f_{\tilde{\omega}} - \tilde{f}_n\| = \min_{\omega \in \mathcal{D}} \|f_\omega - \tilde{f}_n\|$. Then $\|f_{\tilde{\omega}} - f_\omega\| \leq \|f_{\tilde{\omega}} - \tilde{f}_n\| + \|f_\omega - \tilde{f}_n\|$ and

$$\begin{aligned} \sup_{f \in \mathcal{F}_{a,C}} \mathbf{E}_f \|\tilde{f}_n - f\|^2 &\geq \max_{\omega \in \mathcal{D}} \mathbf{E}_{f_\omega} \|\tilde{f}_n - f_\omega\|^2 \geq \frac{1}{4} \max_{\omega \in \mathcal{D}} \mathbf{E}_{f_\omega} \|f_{\tilde{\omega}} - f_\omega\|^2 \\ &= \frac{\delta_m^2}{4} \max_{\omega \in \mathcal{D}} \mathbf{E}_{f_\omega} [\Delta(\tilde{\omega}, \omega)] \geq \frac{\delta_m^2 m \bar{\rho}^{2n}}{16} \asymp m^{-2a} \bar{\rho}^{2n}, \end{aligned} \tag{3.5}$$

because of the Assouad Lemma and (3.3). Take $m \asymp (n \log n)^{1/(2a+3)}$. Then $\log m \asymp \log n$, $nm^{-(2a+3)} \log m \asymp 1$, and it follows from (3.4) that $\bar{\rho}^{2n} \asymp 1$, which gives $\sup_{f \in \mathcal{F}_{a,C}} \mathbf{E}_f \|\tilde{f}_n - f\|^2 \geq c(n \log n)^{-2a/(2a+3)}$ and completes the proof.

It does not seem possible to get rid of the log-factor in the present derivation. The logarithm enters because $H^2(g_\omega, g_{\omega'})$ is of the order of $m^{-(2a+3)} \log m$, since $\int \psi_k^2/\psi_0 d\lambda$ is of the order of $\log(2k + 1)$. A similar statement is true for the method of van Rooij and Ruymgaart. It will be seen in the next Section that $n^{-2a/(2a+3)}$ is the minimax rate for linear estimators. We tend to believe that the log-factor may not be superfluous, and may be related to the restrictions $f(0) = 0$ and $g(1) = 0$, which make the estimation problem slightly easier,

statistically speaking. If it is so, then a suitably constructed nonlinear estimator might give a logarithmic speed-up of the convergence.

4. The Estimator and its Rates of Convergence

Let $Z_j = n_j/(n\lambda(B_j)) = 2Nn_j/n$, $j = 1, \dots, N$, be normalized bin counts and (cf., (2.2)) $\tilde{Z}_\ell = \sum_{j=1}^N Z_j(\chi_\ell)_j/(2N) = n^{-1} \sum_{j=1}^N n_j(\chi_\ell)_j$, $\ell = 0, \dots, N-1$, be the coordinates of the vector $[Z_1, \dots, Z_N]$ with respect to the basis $\{\chi_\ell\}$. Following JS91, we define an estimator \hat{f}_n of f as

$$\hat{f}_n(x) = \phi_0(x) + \sum_{\nu=1}^{N-1} \tilde{Z}_\nu \frac{(1 - \alpha^{\frac{1}{2}}(2\nu+1)^a)_+}{b_\nu \gamma_\nu} \phi_\nu(x), \quad (4.1)$$

with α chosen to ensure that

$$n^{-1} \sum_{\nu=1}^{N-1} b_\nu^{-2} \gamma_\nu^{-2} (2\nu+1)^{2a} (\alpha^{-\frac{1}{2}}(2\nu+1)^{-a} - 1)_+ = C^2. \quad (4.2)$$

The estimator \hat{f}_n is a linear combination of the first N right singular functions of the operator \mathcal{K} ; it is a Fourier-type estimator, in which higher order components are neglected. \hat{f}_n is also a member of the class \mathcal{T}_N of linear estimators that contains estimators of the form $\tilde{f}_n(x) = \phi_0(x) + \sum_{\nu=1}^{N-1} c_\nu \phi_\nu(x)$, with c_ν depending linearly on the data $[n_1, \dots, n_N]$.

Because the risk (3.1) is difficult to handle, Johnstone and Silverman proposed, for $\tilde{f}_n \in \mathcal{T}_N$, to work with a surrogate risk $M^*(\tilde{f}_n, f)$ obtained from (3.1) through simplifying its variance term (for details, see JS91, p.8). To have $M(\tilde{f}_n, f)$ and $M^*(\tilde{f}_n, f)$ asymptotically equivalent, we need to show that the ratio $M(\tilde{f}_n, f)/M^*(\tilde{f}_n, f)$ is bounded above, and below away from zero. This can be done similarly to JS91, p.21, using our Lemma 1. Since g is nonnegative, provided C is small enough, an upper bound for $M(\tilde{f}_n, f)/M^*(\tilde{f}_n, f)$ follows immediately, as in JS91. A positive lower bound for that ratio can again be obtained as in JS91, using the inequality

$$\int_0^1 \psi_\nu^2 \psi_0 d\lambda = \frac{2\nu^2 + \nu + \frac{5}{16}}{\pi\nu^2 + \nu + \frac{3}{16}} \geq \frac{2}{\pi}.$$

The following theorem is the main result of this article and shows that, if the discretization rate is not too slow, then \hat{f}_n is, essentially, asymptotically rate minimax.

Theorem 1. *Let $n = O(N^{2a+3})$. Then for any C such that the right-hand side of (2.9) is smaller than one, \hat{f}_n is, up to a logarithmic factor, asymptotically rate minimax over $\mathcal{F}_{a,C}$ in the class of essentially all estimators, as defined in*

(3.2), with $M(\hat{f}_n, f) \asymp n^{-2a/(2a+3)}$, uniformly for $f \in \mathcal{F}_{a,C}$. In the class of linear estimators, \hat{f}_n is exactly asymptotically rate minimax.

Proof. The lower bound $(n \log n)^{-2a/(2a+3)}$ is given in Proposition 3, and Lemma 1 in JS91 asserts that $n^{-2a/(2a+3)}$ is the minimax rate in the class of linear estimators. It is thus sufficient to show that \hat{f}_n achieves the rate $n^{-2a/(2a+3)}$ and, moreover, in view of the results discussed above, it is sufficient to work with the surrogate risk $M^*(\hat{f}_n, f)$. The reasoning will closely follow that on pages 13-17 of JS91. We sketch the main idea, omit most of the details, and report only some intermediate results that are different from those in JS91.

Denote the minimax risk over a class of estimators \mathcal{T} and a function class \mathcal{F} by

$$\mathcal{M}(\mathcal{T}, \mathcal{F}) = \inf_{\hat{f}_n \in \mathcal{T}} \sup_{f \in \mathcal{F}} M^*(\hat{f}_n, f),$$

and define by $\mathcal{F}_{a,C}^L = \{f \in \mathcal{F}_{a,C} : f_N = f_{N+1} = \dots = 0\}$ the class of “low-frequency” members of $\mathcal{F}_{a,C}$. With Γ_ν given by (2.7) write

$$S_N(\nu) = \sum_{\rho \in \Gamma_\nu \setminus \{\nu\}} \frac{1}{(2\rho + 1)^{2a}},$$

$$\varepsilon(N) = C^2 \left[\max_{\nu \in \{1, 2, \dots, N-1\}} S_N(\nu) + \frac{1}{(2N + 1)^{2a}} \right].$$

It then follows from Theorem 1 in JS91 that if

$$\varepsilon(N) = O(\mathcal{M}(\mathcal{T}_N, \mathcal{F}_{a,C}^L)), \tag{4.3}$$

then $\mathcal{M}(\mathcal{T}_N, \mathcal{F}_{a,C})$ and $\mathcal{M}(\mathcal{T}_N, \mathcal{F}_{a,C}^L)$ are asymptotically equivalent, so that $\varepsilon(N)$ may be called a “high-frequency effect”. Moreover, Lemma 1 in JS91 gives an explicit form of $\mathcal{M}(\mathcal{T}_N, \mathcal{F}_{a,C}^L)$ and asserts, under (4.3), the rate minimaxity of \hat{f}_n in the class of linear estimators. It is thus sufficient to show that (4.3) is fulfilled with the assumed discretization rate, and that $\mathcal{M}(\mathcal{T}_N, \mathcal{F}_{a,C}^L)$ approaches zero at the rate $n^{-2a/(2a+3)}$.

It can easily be shown that

$$S_N(\nu) = \sum_{j=1}^{\infty} [(2\nu + 1 + 4Nj)^{-2a} + (4Nj - 2\nu - 1)^{-2a}] \leq N^{-2a} \sum_{k=1}^{\infty} k^{-2a}.$$

Therefore, $\varepsilon(N) = O(N^{-2a})$, because $a > 1/2$.

Define $\nu_\alpha = \lfloor (\alpha^{-1/(2a)} - 1)/2 \rfloor$. The choice of α satisfying (4.2) is asymptotically equivalent to the choice of ν_α that satisfy

$$\frac{1}{n} \sum_{\nu=1}^{\nu_\alpha-1} \frac{(2\nu + 1)^{2a+2}}{\text{sinc}^2[\frac{\pi(2\nu+1)}{4N}]} \left[\left(\frac{2\nu_\alpha}{2\nu + 1} \right)^a - 1 \right] = \frac{4}{\pi^2} C^2 \tag{4.4}$$

if $\nu_\alpha \leq N - 1$, or to the choice of α that satisfy

$$\frac{1}{n} \sum_{\nu=1}^{N-1} \frac{(2\nu+1)^{2a+2}}{\operatorname{sinc}^2[\frac{\pi(2\nu+1)}{4N}]} \left[\frac{\alpha^{-\frac{1}{2}}}{(2\nu+1)^a} - 1 \right] = \frac{4}{\pi^2} C^2 \quad (4.5)$$

if $\nu_\alpha > N - 1$. To make the choice of ν_α or α , two cases are considered.

1. The discretization index N tends to infinity faster than $n^{1/(2a+3)}$, i.e., $n = o(N^{2a+3})$. Because

$$\frac{1}{n} \sum_{\nu=1}^{M-1} \frac{(2\nu+1)^{2a+2}}{\operatorname{sinc}^2[\frac{\pi(2\nu+1)}{4N}]} \left[\left(\frac{2M}{2\nu+1} \right)^a - 1 \right] \sim \frac{4}{\pi^2} C^2,$$

with $(2M)^{2a+3} = 8n(a+3)(2a+3)C^2/(a\pi^2)$, $\nu_\alpha \sim M$ can be taken in (4.4). ($a_n \sim b_n$ means that $a_n/b_n \rightarrow 1$, as $n \rightarrow \infty$.) The low-frequency minimax error

$$\mathcal{M}(\mathcal{T}_N, \mathcal{F}_{a,C}^L) = \frac{\pi^2}{4n} \sum_{\nu=1}^{M-1} \frac{(2\nu+1)^2}{\operatorname{sinc}^2[\frac{\pi(2\nu+1)}{4N}]} \left[1 - \left(\frac{2\nu+1}{2M} \right)^a \right]$$

is then proportional to $n^{-2a/(2a+3)}$, which implies that $\epsilon(N) = o(\mathcal{M}(\mathcal{T}_N, \mathcal{F}_{a,C}^L))$.

2. The discretization index tends to infinity at the rate $n^{1/(2a+3)}$. Let $n = c(4N)^{2a+3}$, for a positive constant c . Two sub-cases are considered.

(a) $c < J_{a,a+2}(1/2)\pi^2/(8C^2)$, where $J_{p,q}(x) = \int_0^x (x^p - u^p)u^q/\operatorname{sinc}^2(\pi u)du$ is an increasing function of $x \in [0, 1/2)$, with $p, q > 0$. One then takes $\nu_\alpha = \lfloor 2NU \rfloor$ in (4.4), where U is the solution of the equation $J_{a,a+2}(U) = 8cC^2/\pi^2$.

(b) $c > J_{a,a+2}(1/2)\pi^2/(8C^2)$. In this case, $\alpha = (2r)^{-2a}I_{a+2}^2(8cC^2/\pi^2 + I_{2a+2})^{-2}$ is suitable for (4.5), where $I_p = \int_0^{1/2} u^p/\operatorname{sinc}^2(\pi u)du$.

In both sub-cases one gets $\epsilon(N)$ and $\mathcal{M}(\mathcal{T}_N, \mathcal{F}_{a,C}^L)$ proportional to $n^{-2a/(2a+3)}$.

5. Empirical Risk Minimization and Numerical Experiment

In practical applications, it is more natural to estimate the intensity function with respect to the Lebesgue measure (denoted by $u(x)$ in Section 1) rather than with respect to μ . Therefore, in the numerical experiment, a modified estimator was used. Since $u(x) = xf(x)$, the estimator given by (4.1) was multiplied by x to get an estimator of u . Also, the restriction $f_0 = 1$ is rather artificial and it may be replaced for greater flexibility with, say, $f_0 = D$. The constant D should

be $\int f\phi_0 d\mu$ and may be estimated from the data, if it is not known *a priori*. It follows from (2.1) that $f(x) = (\mathcal{K}^{-1}g)(x) = -(2x)^{-1}g'(x)$. Substituting that into the integral defining D , using the explicit form of ϕ_0 and integrating by parts, one easily obtains $D = \pi \int_0^1 h(y) \cos(\pi y^2/2) dy$. A natural estimator for D can thus be constructed as

$$\hat{D} = \pi \sum_{j=1}^N \frac{n_j}{n} \cos \frac{\pi y_j^2}{2},$$

where y_j is the midpoint of the j th data bin.

The estimator used in simulations and recommended for practical usage had the form

$$\hat{u}_n(x) = x\hat{D}\phi_0(x) + \sum_{\nu=1}^{N-1} \tilde{Z}_\nu \frac{(1 - \alpha^{\frac{1}{2}}(2\nu + 1)^a)_+}{b_\nu \gamma_\nu} x\phi_\nu(x). \tag{5.1}$$

A nontrivial problem is how to adaptively select the parameters. N is usually fixed by an experimental setup. Given N , the choice of a and α may be based on minimization of an empirical analogue of the risk function - a common idea behind, e.g., cross-validation and Mallows C_p (cf., e.g., Efromovich (1999, Chap. 7.4)). The surrogate risk $M^*(\hat{f}_n, f)$, asymptotically equivalent to $M(\hat{f}_n, f)$, may be written in the form (cf., JS91,p.8)

$$M^*(\hat{f}_n, f) = \sum_{\nu=1}^{N-1} \left(T_\nu G_\nu - \frac{g_\nu}{b_\nu} \right)^2 + \frac{1}{n} \sum_{\nu=1}^{N-1} T_\nu^2 + c,$$

where $T_\nu = (1 - \alpha^{1/2}(2\nu + 1)^a)_+ / (\gamma_\nu b_\nu)$, $g_\nu = b_\nu f_\nu$, $G_\nu = \sum_{k \in \Gamma_\nu} \gamma_k g_k$ with Γ_ν defined in (2.7), and c does not depend on a or on α . If the g_k 's decrease quickly, then $G_\nu \approx \gamma_\nu g_\nu$ and

$$M^*(\hat{f}_n, f) \approx \sum_{\nu=1}^{N-1} \left(T_\nu - \frac{1}{b_\nu \gamma_\nu} \right)^2 G_\nu^2 + \frac{1}{n} \sum_{\nu=1}^{N-1} T_\nu^2 + c.$$

It is known (JS91,p.7) that $E_f(\tilde{Z}_\nu) = G_\nu$ and $\text{Var}_f(\tilde{Z}_\nu) = n^{-2} \sum_{j=1}^N E_f(n_j)(\chi_\nu)_j^2$, so that $\hat{G}_\nu^2 = \tilde{Z}_\nu^2 - n^{-2} \sum_{j=1}^N n_j(\chi_\nu)_j^2$ is an unbiased estimator of G_ν^2 . Consequently, neglecting constant terms, a and α are chosen to minimize

$$\begin{aligned} & \sum_{\nu=1}^{N-1} \left(T_\nu^2 - \frac{2T_\nu}{\gamma_\nu b_\nu} \right) \hat{G}_\nu^2 + \frac{1}{n} \sum_{\nu=1}^{N-1} T_\nu^2 \\ &= - \sum_{\nu=1}^{N-1} \frac{(1 - \alpha(2\nu + 1)^{2a})_+}{b_\nu^2 \gamma_\nu^2} \hat{G}_\nu^2 + \frac{1}{n} \sum_{\nu=1}^{N-1} \frac{(1 - \alpha^{\frac{1}{2}}(2\nu + 1)^a)_+^2}{b_\nu^2 \gamma_\nu^2}, \end{aligned} \tag{5.2}$$

and the \hat{u}_n that corresponds to the minimum is taken as the final solution and referred to as the ERM estimator.

The first term in (5.2) is, up to an additive constant, the “squared bias term” and the second one is the “variance term”. If, for instance, a is kept fixed and α decreases, then the bias term decreases and the variance term tends to increase. Minimization of (5.2) provides a bias-variance tradeoff.

In the simulations, two values of the experiment size n were used: 2,000 and 10,000. The value of N was always 40. The minimization was performed through a grid search with $(a, \alpha) \in [0.5, 1.5] \times [0.01, 0.1]$, and with steps 0.1 and 0.01, respectively. The points that led to more than 10 components in (5.1), as well as those leading to negative solutions, were excluded from the search, as a safety measure against occasional “pathological” data that may lead to heavily oscillating solutions.

Data samples were generated from the following intensity functions:

- Swapped Minerbo-Levy A (SML-A):

$$u(x) = 2[1 - 2(1 - x)^2]\mathbf{1}_{(0,0.5]} + 4[1 - (1 - x)^2]\mathbf{1}_{(0.5,1)};$$
- Swapped Minerbo-Levy B (SML-B):

$$u(x) = 1.241(2x - x^2)^{-3/2} \exp[1.21(1 - (2x - x^2)^{-1})];$$
- Normal mixture (NM): $0.7 \cdot N(0.7, 0.08) + 0.3 \cdot N(0.35, 0.08);$
- Beta(4, 2): $u(x) = 20x^3(1 - x);$
- Step function (SF): $u(x) = 0.6 \mathbf{1}_{[0,1/3]} + 0.9 \mathbf{1}_{(1/3,0.75]} + 1.7 \mathbf{1}_{(0.75,1]}.$

The first two functions are taken from Minerbo and Levy (1969) and swapped to satisfy $u(0) = 0$. The step function neither satisfies that condition, nor is it continuous. It was included in the simulation experiment to check the behaviour of the estimator when the assumptions of the theory are violated.

For each function and for each experiment size, 10 artificial data samples were generated and the estimators \hat{u}_n were constructed with a and α selected to minimize the empirical risk. Additionally, for each of the data samples, the best possible values of a and α were found, i.e., those that minimize the (numerically computed) L^2 distance between \hat{u}_n and the true u . The best and worst cases (out of 10 data samples) are presented in Figures 5.1–5.5.

As expected, clear improvement is seen when n increases from 2,000 to 10,000. The promising behaviour of the proposed ERM procedure calls for a more thorough, theoretical study of its adaptivity properties. This is, however, outside the scope of the present paper.

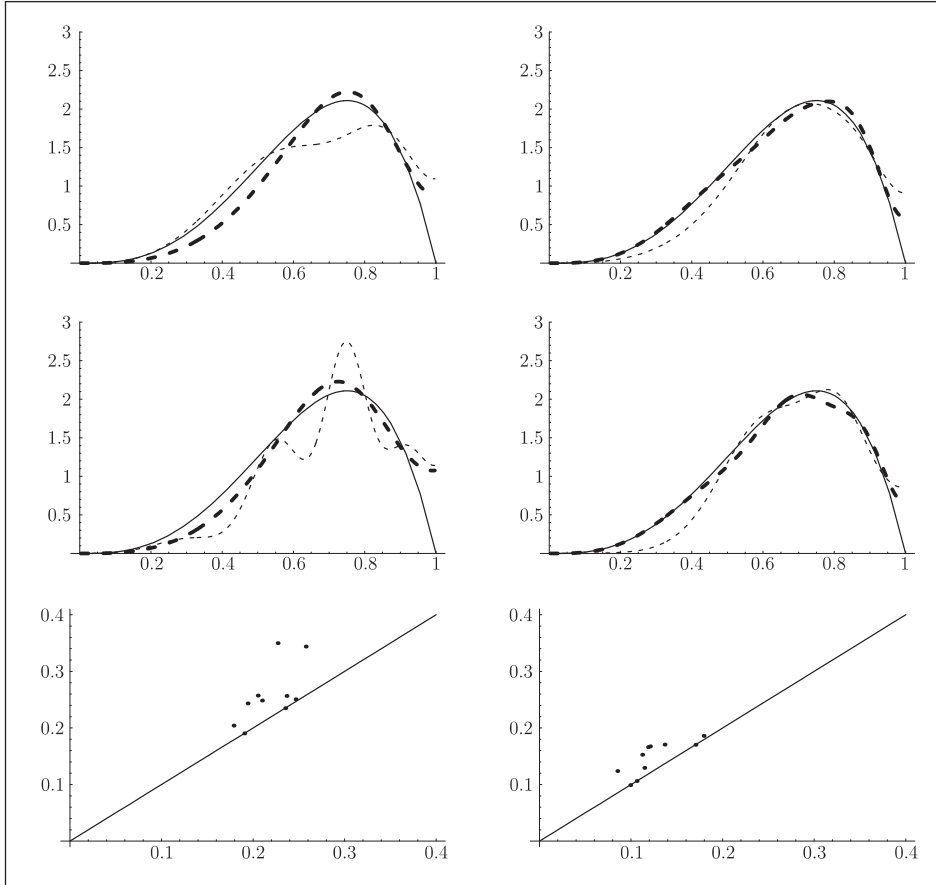


Figure 5.1. Worst (dashed) and best (thick dashed) reconstruction (out of 10 data samples) of a Beta(4, 2) intensity function (solid line) for the experiment size $n = 2,000$ (left) and $n = 10,000$ (right). From top: the best possible solutions, the solutions obtained through minimization of the empirical risk (ERM) and scatterplots of L^2 errors of ERM solutions versus those of the best possible solutions. (The closer the points to the bisecting line, the more efficient the ERM procedure for selection of a and α .)

Acknowledgement

Work partially supported by AGH local grant 10.420.03. Constructive criticism of anonymous reviewers is kindly acknowledged.

Appendix

Proof of Proposition 2. Let F be defined by $f(x) = F(x^2)$. Clearly, $f \rightarrow F$ is an isometric bijection between $L^2(xdx)$ and $L^2(dz/2)$. The right singular functions $\{\phi_\nu\}$ form an orthonormal basis in $L^2(xdx)$, as a set of eigenfunctions

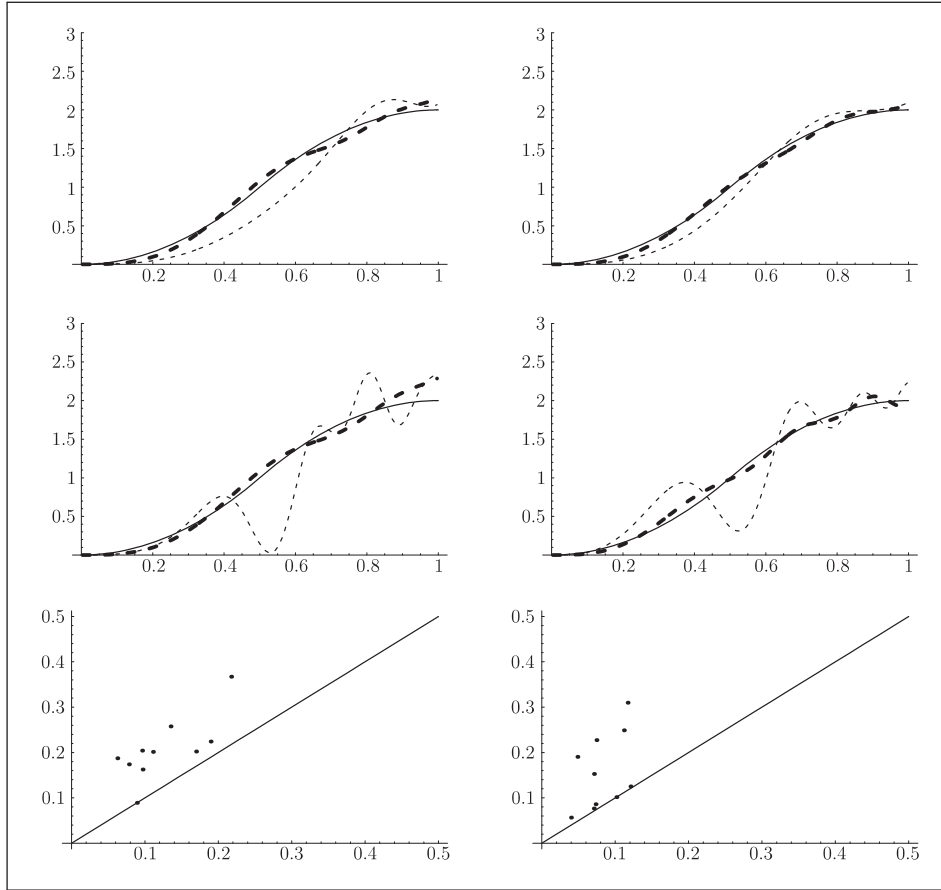


Figure 5.2. Similar to Fig. 5.1, but for the SML-A intensity function.

of $\mathcal{K}^*\mathcal{K}$, which is defined on $L^2(xdx)$. Consequently, $u_\nu(z) := 2 \sin[(2\nu + 1)\pi z/2]$, $\nu = 0, 1, \dots$, form an orthonormal basis in $L^2(dz/2)$. Similarly, the left singular functions $\{\psi_\nu\}$ form an orthonormal basis in $L^2(xdx)$, as a set of eigenfunctions of $\mathcal{K}\mathcal{K}^*$, which is defined on $L^2(xdx)$ and, consequently, $v_\nu(z) := 2 \cos[(2\nu + 1)\pi z/2]$, $\nu = 0, 1, \dots$, form another orthonormal basis in $L^2(dz/2)$. Obviously, $f = \sum_\nu f_\nu \phi_\nu$, if and only if $F = \sum_\nu f_\nu u_\nu$.

For the proof of a , notice that, by the Schwarz inequality,

$$\sum_{\nu=0}^{\infty} (2\nu + 1)^k |f_\nu| \leq \left[\sum_{\nu=0}^{\infty} (2\nu + 1)^{2a} |f_\nu|^2 \right]^{\frac{1}{2}} \left[\sum_{\nu=0}^{\infty} \frac{1}{(2\nu + 1)^{2(a-k)}} \right]^{\frac{1}{2}}. \quad (\text{A.1})$$

If $f \in \mathcal{F}_{a,C}$ with $a > k+1/2$, then $\sum_\nu (2\nu+1)^k |f_\nu| < \infty$, which allows for termwise differentiation, so that $F^{(k)} = \sum_\nu f_\nu u_\nu^{(k)}$, and $F^{(k)}$ is continuous because the series

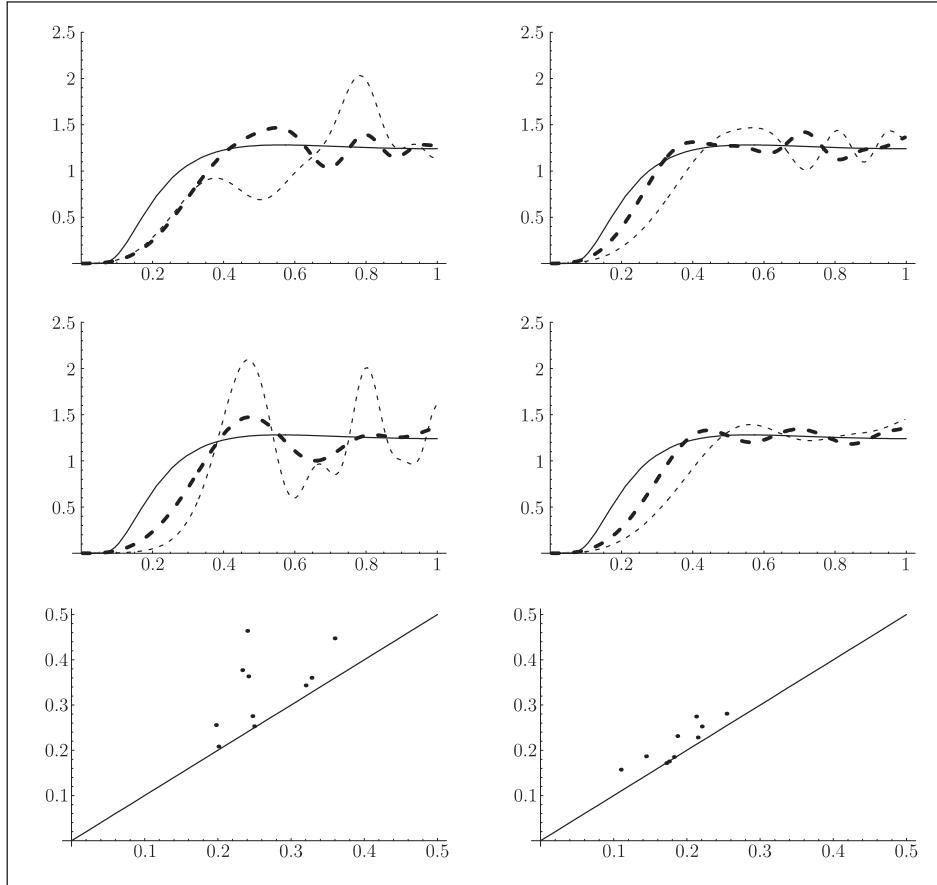


Figure 5.3. Similar to Fig. 5.1, but for the SML-B intensity function.

converges uniformly in $[0, 1]$. Consequently, f itself is then k times continuously differentiable in $[0, 1]$.

For the proof of b , we first show that F has k weak derivatives that are square integrable in $[0, 1]$ with respect to dx . Write $F = \sum_{\nu} f_{\nu} u_{\nu}$ and take $D^k F := \sum_{\nu} f_{\nu} u_{\nu}^{(k)}$. Then, for odd k , $D^k F = \epsilon_k \sum_{\nu} f_{\nu} [(2\nu+1)\pi/2]^k v_{\nu}$ with $\epsilon_k = (-1)^{\lfloor k/2 \rfloor}$, and $D^k F \in L^2(dz/2)$ if $f \in \mathcal{F}_{k,C}$. For even k , $D^k F = \epsilon_k \sum_{\nu} f_{\nu} [(2\nu+1)\pi/2]^k u_{\nu}$ and, again, $D^k F \in L^2(dz/2)$ if $f \in \mathcal{F}_{k,C}$. Since $F \in L^2(dx)$, it suffices to show that, for any $s \in C_0^{\infty}(0, 1)$, the space of infinitely differentiable functions with compact support contained in $(0, 1)$,

$$\int_0^1 F(x) s^{(k)}(x) dx = (-1)^k \int_0^1 (D^k F)(x) s(x) dx. \tag{A.2}$$

For $m \in \mathbb{N}$, write $F = \sum_{\nu=0}^m f_{\nu} u_{\nu} + R_m$ with $\|R_m\|_{L^2(dx)} \rightarrow 0$ when $m \rightarrow \infty$, because of the Parseval equality. Then, integrating by parts k times and using

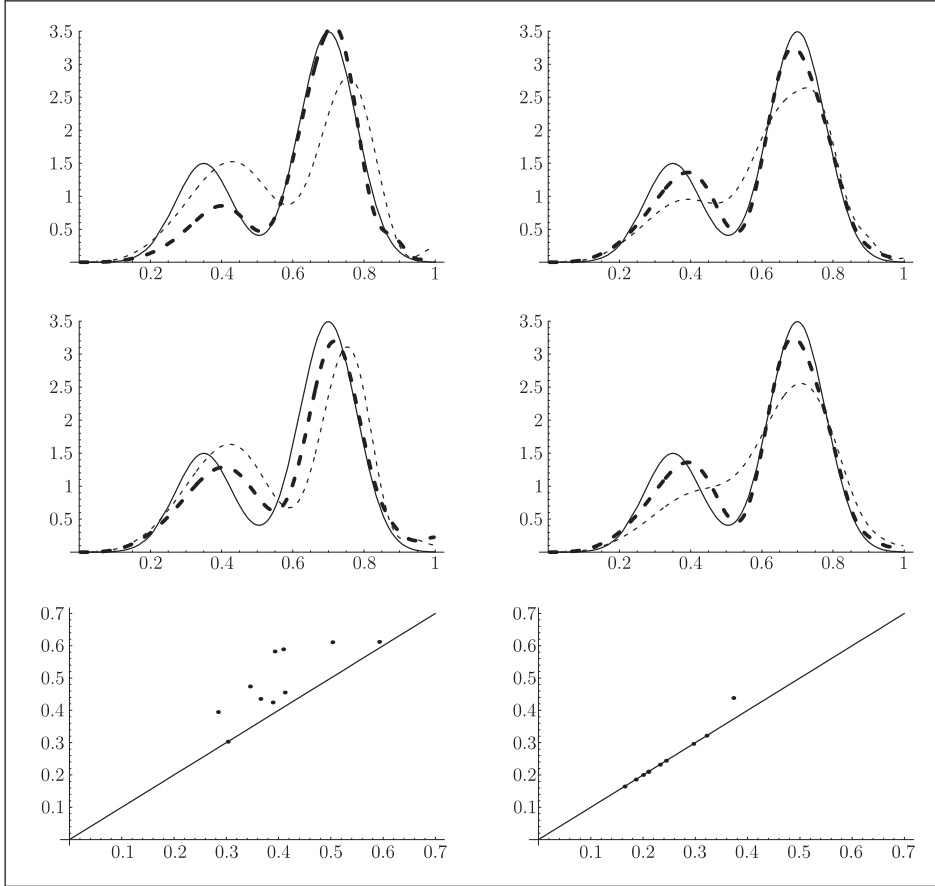


Figure 5.4. Similar to Fig. 5.1, but for the NM intensity function.

the Schwarz inequality,

$$\int_0^1 F s^{(k)} dx = (-1)^k \int_0^1 \sum_{\nu=0}^m f_\nu u_\nu^{(k)} s dx + O(\|R_m\|_{L^2(dx)}).$$

For odd k , write $D^k F = \epsilon_k \sum_{\nu=0}^m f_\nu [(2\nu + 1)\pi/2]^k v_\nu + \tilde{R}_m$ with $\|\tilde{R}_m\|_{L^2(dx)} \rightarrow 0$ when $m \rightarrow \infty$, because of the Parseval equality, and further, again by the Schwarz inequality,

$$\begin{aligned} \int_0^1 F s^{(k)} dx &= (-1)^k \int_0^1 (D^k F - \tilde{R}_m) s dx + O(\|R_m\|_{L^2(dx)}) \\ &= (-1)^k \int_0^1 (D^k F) s dx + O(\|\tilde{R}_m\|_{L^2(dx)}) + O(\|R_m\|_{L^2(dx)}). \end{aligned}$$

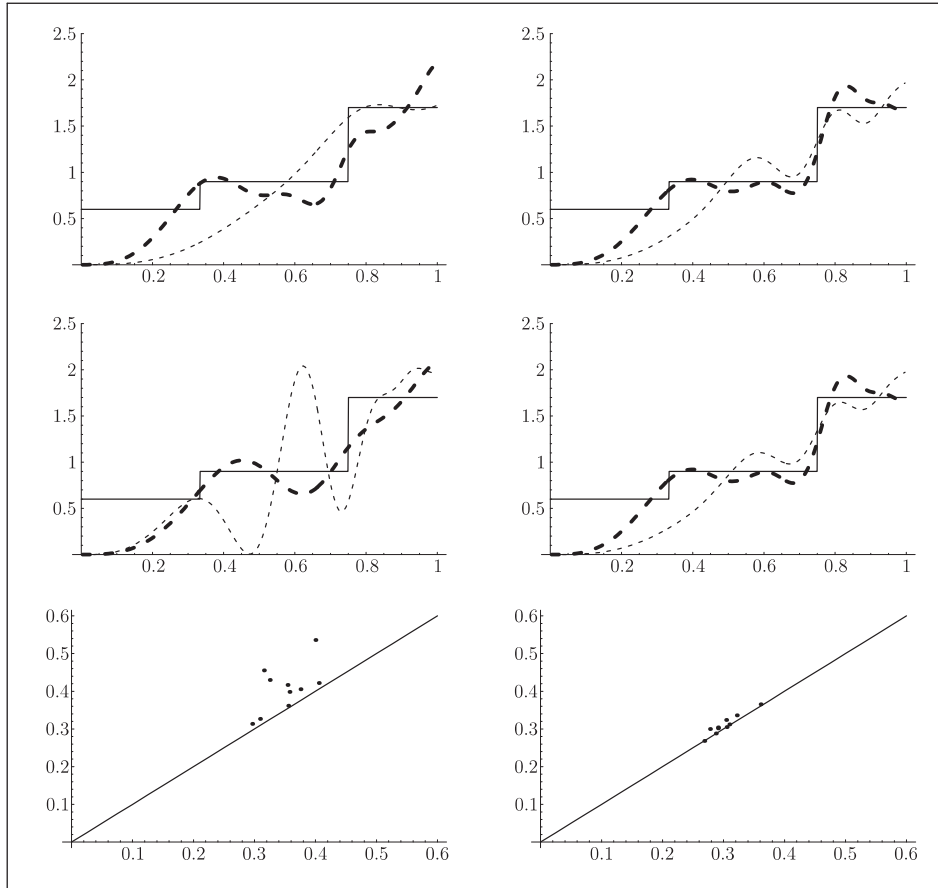


Figure 5.5. Similar to Fig. 5.1, but for the SF intensity function.

Equality (A.2) is then obtained when $m \rightarrow \infty$. For even k the reasoning is almost identical, with v_ν replaced by u_ν in the representation of $D^k F$.

The conclusion about f now follows in a fairly standard way, similar to the proof of Theorem 3.41 in Adams and Fournier (2003, p.78). One first proves the chain rule for the weak differentiation of $F(x^2)$ by approximating the $W^{k,1}(0,1)$ function F with $C^\infty(0,1)$ functions (in the Sobolev norm). An upper bound for $\int (D^k f)^2 dm$ then follows easily, and the integration with respect to $dm(x)$ rather than dx helps to handle the singularity of $(\sqrt{x})'$ at zero.

References

Adams, R. A. and Fournier, J. J. F. (2003). *Sobolev Spaces*. Academic Press: Elsevier, Amsterdam.

- Assouad, P. (1983). Deux remarques sur l'estimation. *C. R. Acad. Sc. Paris Sér. I Math.* **296**, 1021-1024.
- Barthel, M., Klimanek, P. and Stoyan, D. (1985). Stereological substructure analysis in hot-deformed metals from TEM-images. *Czech. J. Phys. B* **35**, 265-268.
- Birgé, L. (2006). The Brouwer Lecture 2005. Statistical estimation with model selection. (available at arXiv:math.ST/0605187 v2)
- Efromovich, S. (1999). *Nonparametric Curve Estimation*. Springer, New York.
- Gradshteyn, I. S. and Ryzhik, G. E. (1980). *Tables of Integrals, Series and Products*. Academic Press, New York.
- Han, J.-H. and Kim, D.-Y. (1998). Determination of three-dimensional grain size distribution by linear intercept measurement. *Acta Mater.* **46**, 2021-2028.
- Hilliard, J. E. and Lawson L. R. (2003). *Stereology and Stochastic Geometry*. Kluwer, Dordrecht.
- Johnstone, I. M. and Silverman, B. W. (1990). Speed of estimation in positron emission tomography and related inverse problems. *Ann. Statist.* **18**, 251-280.
- Johnstone, I. M. and Silverman, B. W. (1991). Discretization effects in statistical inverse problems. *J. Complexity* **7**, 1-34.
- Lord, G. W. and Willis, T. F. (1951). Calculation of air bubble distribution from results of a Rosiwal traverse of aerated concrete. *A.S.T.M. Bull.* **56**, 177-187.
- Minerbo, G. N. and Levy, M. E. (1969). Inversion of Abel's integral equation by means of orthogonal polynomials. *SIAM J. Numer. Anal.* **6**, 598-616.
- Reiss, R. D. (1993). *A Course on Point Processes*. Springer, New York.
- van Rooij, A. C. M. and Ruymgaart, F. H. (1996). Asymptotic minimax rates for abstract linear estimators. *J. Statist. Plann. Inference* **53**, 389-402.
- Spektor, A. G. (1950). Analysis of distribution of spherical particles in non-transparent structures. *Zavodskaja Laboratorija* **16**, 173-177.
- Stoyan, D., Kendall, W. S. and Mecke, J. (1987). *Stochastic Geometry and its Applications*. Akademie-Verlag, Berlin.
- Szkutnik, Z. (2007). Unfolding spheres size distribution from linear sections with B-splines and EMDS algorithm. *Opuscula Mathematica* **27**, 151-165.

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

E-mail: aedudek@agh.edu.pl

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

E-mail: szkutnik@agh.edu.pl

(Received July 2006; accepted April 2007)