Opuscula Mathematica • Vol. 27 • No. 1 • 2007

Zbigniew Szkutnik

UNFOLDING SPHERES SIZE DISTRIBUTION FROM LINEAR SECTIONS WITH B-SPLINES AND EMDS ALGORITHM

Abstract. The stereological problem of unfolding spheres size distribution from linear sections is formulated as a problem of inverse estimation of a Poisson process intensity function. A singular value expansion of the corresponding integral operator is given. The theory of recently proposed B-spline sieved quasi-maximum likelihood estimators is modified to make it applicable to the current problem. Strong L^2-consistency is proved and convergence rates are given. The estimators are implemented with the recently proposed EMDS algorithm. Promising performance of this new methodology in finite samples is illustrated with a numerical example. Data grouping effects are also discussed.

Keywords: inverse problem, singular value expansion, stereology, discretization, quasi-maximum likelihood estimator.

Mathematics Subject Classification: 62G05, 45Q05.

1. THE UNFOLDING PROBLEM

A population of spheres embedded in a medium is modeled with a Poisson process Ψ_1 of points (x, y, z, R) in $\mathbb{R}^3 \times (0, \infty)$. The centers (x, y, z) of the spheres form a homogeneous Poisson process in $\mathbb{R}^3$ with the expected number of c points per unit volume. The random spheres radii R have a distribution Q, independent of the center. The mean measure of Ψ_1 is thus $\nu_1 = c \cdot \lambda_3 \otimes Q$. (Here and in what follows λ_k stands for the Lebesgue measure in $\mathbb{R}^k$.)

The spheres cannot be observed directly. Instead, a random linear section through the medium is observed, i.e., for a randomly selected straight line, one observes the line segments that are intersections of the line with the spheres. Our derivation of the folding operator is similar to that given in [5], pp. 47–48, for a related Wicksell's problem. Without loss of generality, assume that the straight line is the z-axis. For $D = \{(x, y, z, R) : x^2 + y^2 \leq R^2\}$, denote by $\Psi_2(\cdot) := \Psi_1(\cdot \cap D)$ the truncation of

Ψ_1 to those spheres that are intersected by the z-axis. Ψ_2 is again a Poisson process with the mean measure $\nu_2(\cdot) = \nu_1(\cdot \cap D)$; see, e.g., [5], p. 8.

Let Φ be the point process of the observed linear sections, i.e., the point process in $\mathbb{R}^2$ with points (z, r) that represent the centers z and radii r of the observed line segments (one-dimensional balls). The points of Φ are thus obtained from the points of Ψ_2 through the transformation $h(x, y, z, R) = (z, \sqrt{R^2 - x^2 - y^2})$. Therefore, Φ is a Poisson process with the mean measure $\nu_\Phi(\cdot) = \nu_2[h^{-1}(\cdot)]$; see, e.g., [5], p. 13. For any Borel set $B \subset \mathbb{R}$ and $t > 0$, one obtains

$$\begin{aligned} \nu_\Phi(B \times [0,t]) &= \nu_2\left(\left\{(x,y,z,R) : z \in B,\ \sqrt{R^2 - x^2 - y^2} \le t\right\}\right) = \\ &= \nu_1\left(\left\{(x,y,z,R) : z \in B,\ \sqrt{R^2 - x^2 - y^2} \le t,\ x^2 + y^2 \le R^2\right\}\right) = \\ &= c \cdot \lambda_1(B) \cdot (\lambda_2 \otimes Q)\left(\left\{(x,y,z,R) : R^2 - t^2 \le x^2 + y^2 \le R^2\right\}\right) = \\ &= c \cdot \lambda_1(B) \cdot \pi \int_0^\infty \left[R^2 - \max\{0, R^2 - t^2\}\right] \mathrm{d}Q(R)\,. \end{aligned}$$

Noting that

$$R^2 - \max\{0, R^2 - t^2\} = \int_0^t \mathbf{1}_{[0,R]}(r) \cdot 2r\mathrm{d}r\,,$$

one gets, changing the order of integration,

$$\begin{aligned} \nu_\Phi(B \times [0,t]) &= \pi c \lambda_1(B) \int_0^\infty \int_0^t \mathbf{1}_{[0,R]}(r) \cdot 2r\mathrm{d}r\mathrm{d}Q(R) = \\ &= \pi c \lambda_1(B) \int_0^t \left[2r \int_r^\infty \mathrm{d}Q(R)\right] \mathrm{d}r. \end{aligned}$$

This means that, if B is the observed portion of the linear section through the medium, then the intensity function of the Poisson process on $[0, \infty)$ of the radii of observed sections has an intensity function of the form $2\pi c \lambda_1(B) r \int_r^\infty \mathrm{d}Q(R)$ with respect to λ_1. Assume that there is an upper bound, say 1, for R and that $Q << \lambda_1$ with $\mathrm{d}Q/\mathrm{d}\lambda_1 = q$. Denote cq with f and the 'size of the experiment' $\pi\lambda_1(B)$ with t. One then observes a Poisson process of radii of sections with an intensity function $t \cdot g(r)$, where

$$g(r) = 2r \int_r^1 f(R)\mathrm{d}R \tag{1}$$

and the final goal is to unfold f. Notice that the definition of the 'size of the experiment' is quite natural: t equals the volume of the cylinder to which the centers of the intersected balls must belong. Also notice that the function f to be unfolded does not have to be a probability density. This means that both the shape of the distribution and the intensity c have to be estimated.

Equations equivalent to (1) were first derived by Spektor ([7]) and Lord and Willis ([4]) as models of some measurements in material sciences. For an application in metallurgy, see, e.g., [1]. The problem, called in the sequel the SLW problem, was also discussed in [8], p. 296–299, along with traditionally used algorithms based on various discretizations of equation (1), and the (rather discouraging) performance of the algorithms was illustrated with a numerical example. Since then, to the best of our knowledge, there have been no further significant contributions to the problem.

The SLW problem is known to be a rather hard ill-posed inverse problem, essentially harder than the related and better-known Wicksell's stereological problem of unfolding spheres size distribution from planar sections. The solution of (1) takes the form:

$$f(R) = \frac{1}{2}\left[\frac{g(R)}{R^2} - \frac{g'(R)}{R}\right],$$

which explains the statistical difficulty of the problem – inverse estimation of f in $L^2(\mathrm{d}R)$ roughly corresponds to the direct estimation of the intensity g in $L^2(R^{-4}\mathrm{d}R)$ and of its derivative g' in $L^2(R^{-2}\mathrm{d}R)$.

The aim of this paper is to study the potential of a more formal, alternative approach to the SLW problem – the construction of nonparametric, sieved quasi-maximum likelihood estimators. In Section 2, the difficulty of the SLW problem is quantified with the decay rate of the singular values of the integral operator defined in (1)–the result needed for the analysis of the asymptotics of the estimators. In Section 3, the construction of sieved quasi-maximum likelihood estimators is discussed and general theorems on L^2-consistency and convergence rates are given and then applied to the SLW problem. A numerical example is given in Section 4. Proofs and some auxiliary results are deferred to the Appendix.

2. SINGULAR VALUES AND SINGULAR FUNCTIONS OF THE FOLDING OPERATOR

The kernel $k(y,x) = 2y\mathbf{1}_{\{y<x\}}$ of the operator $(\mathcal{K}f)(y) = \int_0^1 k(y,x)f(x)\mathrm{d}x$ defined by equation (1) is square-integrable in $[0,1]^2$, which implies that $\mathcal{K}$, considered as an operator in $L^2([0,1],\lambda_1)$, is a Hilbert-Schmidt operator. Consequently, as an inverse of a compact operator, $\mathcal{K}^{-1}$ is not bounded and the unfolding problem is ill-posed in the Hadamard sense. The degree of ill-posedness can be measured with the decay rate of the singular values σ_i of $\mathcal{K}$, written in the nonincreasing order. It will be shown below that they decay as i^{-1}. This shows that the SLW problem is indeed essentially harder than the Wicksell's problem, for which the singular values of the corresponding Abel-type operator are known to decay as $i^{-1/2}$, with suitably chosen dominating measures.

The singular values and the right singular functions of $\mathcal{K}$ can be found, respectively, as square roots of the eigenvalues and as the eigenfunctions of the self-adjoint operator $\mathcal{K}^*\mathcal{K}$, which is an integral operator of the form

$$(\mathcal{K}^*\mathcal{K}f)(x) = \frac{4}{3}\int_0^1 \min^3(x,y)f(y)\mathrm{d}y = \frac{4}{3}\int_0^x y^3 f(y)\mathrm{d}y + \frac{4}{3}\int_x^1 x^3 f(y)\mathrm{d}y\,.$$

Differentiation of the eigenequation $(\mathcal{K}^*\mathcal{K}f)(x) = \eta f(x)$ with respect to x gives

$$4x^2\int_x^1 f(y)\mathrm{d}y = \eta f'(x)\,. \tag{2}$$

Setting $x = 0$ in the eigenequation gives $f(0) = 0$ and setting $x = 1$ in equation (2) gives $f'(1) = 0$. Division of (2) by x^2 and another differentiation with respect to x leads to a differential eigenvalue problem

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

with $\mu = 4/\eta$.

The solution of this differential equation takes the form (cf. [3], Part 3, Ch. II, Eq. 2.162(1a)):

$$f(x) = [C_1 J_{3/4}(\sqrt{\mu}x^2/2) + C_2 J_{-3/4}(\sqrt{\mu}x^2/2)] \cdot x^{3/2},$$

where $J_\nu(\cdot)$ denotes rank ν Bessel function of the first kind, i.e.

$$\begin{aligned} J_\nu(z) &= \frac{z^\nu}{2^\nu\,\Gamma(\nu+1)}\left(1 - \frac{z^2}{2(2\nu+2)} + \frac{z^4}{2\cdot 4(2\nu+2)(2\nu+4)} - \ldots\right) = \\ &= \sum_{k=0}^{\infty} \frac{(-1)^k (z/2)^{\nu+2k}}{k!\,\Gamma(\nu+k+1)}\,. \end{aligned} \tag{3}$$

Since $J_\nu(z) \asymp z^\nu$, as $z \to 0$, one obtains $x^{3/2}J_{-3/4}(\sqrt{\mu}x^2/2) \asymp 1$ and $x^{3/2}J_{3/4}(\sqrt{\mu}x^2/2) \to 0$, as $x \to 0$, and the boundary condition $f(0) = 0$ implies that $C_2 = 0$. It is well known (see, e.g., [13], Ch. 17.21) that $[z^\nu J_\nu(z)]' = z^\nu J_{\nu-1}(z)$. Hence, with $F(y) := y^{3/4}J_{3/4}(y)$, we obtain

$$f'(x) = C_1 \left(\frac{2}{\sqrt{\mu}}\right)^{3/4} \frac{d}{dx} F(\sqrt{\mu}x^2) = C_1 \sqrt{\mu}\, x^{5/2}\, J_{-3/4}(\sqrt{\mu}x^2/2)\,,$$

which implies that $f'(1) = 0$ if and only if $J_{-1/4}(\sqrt{\mu}/2) = 0$.

For $|z| \to \infty$, one has $J_\nu(z) = \sqrt{2/(\pi z)}[\cos(z - \nu\pi/2 - \pi/4) + O(1/z)]$ (see, e.g., [13], Ch. 17.5), so that, for $\mu_i \to \infty$,

$$J_{-1/4}(\sqrt{\mu_i}/2) = \frac{2}{\pi^{1/2}\mu_i^{1/4}}\left[-\sin\left(\frac{\sqrt{\mu_i}}{2} - \frac{5\pi}{8}\right) + H(\mu_i)\right]$$

with a function $H(\cdot)$ such that

$$H(\mu_i) = O(1/\sqrt{\mu_i})\,. \tag{4}$$

Hence, $J_{-1/4}(\sqrt{\mu_i}/2) = 0$ if

$$\frac{\sqrt{\mu_i}}{2} - \frac{5\pi}{8} = i\pi + \Delta_i \tag{5}$$

with $\Delta_i \to 0$ such that

$$\sin(i\pi + \Delta_i) = (-1)^i \sin\Delta_i = H(\mu_i)\,. \tag{6}$$

Then, because of (4), (5) and (6),

$$\frac{\sqrt{\mu_i}}{2} - \frac{5\pi}{8} = i\pi + \frac{\Delta_i}{\sin \Delta_i} \sin \Delta_i = i\pi + O(1/\sqrt{\mu_i})\,,$$

which implies that $\sqrt{\mu_i} \asymp i$ and

$$\mu_i = (5\pi/4 + 2i\pi)^2 + O(1/\mu_i) + O(i/\sqrt{\mu_i}) = (5\pi/4 + 2i\pi)^2 + O(1)\,.$$

Consequently,

$$\eta_i = \frac{4}{\mu_i} = \frac{4}{(5\pi/4 + 2i\pi)^2 + O(1)} \asymp i^{-2}\,,$$

i.e., the singular values σ_i of the SLW operator $\mathcal{K}$ are exactly of the order of i^{-1}.

With z_i, $i = 1, 2, \ldots$ denoting the positive zeroes of $J_{-1/4}(z)$, the right singular functions are $\phi_i(x) = A_i x^{3/2} J_{3/4}(z_i x^2)$ and the normalizing constants $A_i = 2/|J_{3/4}(z_i)| = 2/|J'_{-1/4}(z_i)|$ can easily be computed using the integral formulas given, e.g., in [13], Ch. 17, Ex. 18. Those formulas can also be used to prove directly that ϕ_i, $i = 1, 2, \ldots$ indeed form an orthonormal system.

The left singular functions $\psi_i(y) = A_i y^{3/2} J_{-1/4}(z_i y^2)$ can now be obtained from the equation $\mathcal{K}\phi_i = \sigma_i \psi_i$, using representation (3). Again, integral formulas from [13], Ch.17, Ex.19 can be used to prove directly that ψ_i, $i = 1, 2, \ldots$ form an orthonormal system.

The calculations are summarized as

Proposition 1. *Let z_i, $i = 1, 2, \ldots$ be the positive zeroes of $J_{-1/4}(z)$ and let $A_i = 2/|J_{3/4}(z_i)| = 2/|J'_{-1/4}(z_i)|$. The singular values of the SLW operator, considered as an operator in $L^2([0,1], \lambda_1)$, are equal to $\sigma_i = z_i^{-1} \asymp i^{-1}$ with the corresponding right singular functions $\phi_i(x) = A_i x^{3/2} J_{3/4}(z_i x^2)$ and left singular functions $\psi_i(y) = A_i y^{3/2} J_{-1/4}(z_i y^2)$.*

3. SIEVED QUASI-MAXIMUM LIKELIHOOD ESTIMATORS

As an alternative to the traditional algorithms, described in [8], the SLW problem may be solved with a sieved quasi-maximum likelihood approach. For a general inverse problem, with B-spline sieves in the solution space and with discrete, binned data, this approach was studied in detail in [12]. Following that paper, let $[0,1] = B_1 \cup \cdots \cup B_m$ be a partition of the data space into disjoint bins. The observed data $\mathbf{n} = [n_1, \ldots, n_m]$ consist of the counts n_i of the line segments radii observed in the bins B_i, respectively.

The order p, B-spline sieve in the solution space is defined as follows. First, a set of equidistant knots is defined by $x_k = kh$, $k = -p+1, -p+2, \ldots, n$ with $h = 1/(n-p+1)$. Notice that $x_0 = 0$ and $x_{n-p+1} = 1$, so that, in total, $2p-2$ knots are outside the interval $[0,1]$. Then, the order p, B-spline sieve is defined as $U_n = \mathrm{Span}\{u_j, j = 1, \ldots, n\}$, with $u_j(x) = Q_p((x - x_{j-p})/h)\mathbf{1}_{[0,1]}(x)$, where

$$Q_p(x) = \frac{1}{(p-1)!} \sum_{i=0}^{p} (-1)^i \binom{p}{i} (x-i)_+^{p-1}\,.$$

$\{u_j\}$ is a basis of the linear space of order p (degree $p-1$) splines on $[0,1]$ with $n-p$ internal, equidistant knots of multiplicity one (cf. [6], Theorem 4.9).

The data binning can also be expressed in terms of sieves. Let $v_i(y) = \mathbf{1}_{B_i}(y),\ i = 1,\dots,m$ be indicator functions of the bins B_i. In the observation space one then has a histogram sieve $V_m = \mathrm{Span}\{v_i, i = 1,\dots,m\}$. Denote with $\mathcal{P}_m^V$ and $\mathcal{P}_n^U$ the $L^2([0,1],\lambda_1)$ projections onto V_m and U_n. Discretization replaces the operator $\mathcal{K}$ with a finite-dimensional operator $\mathcal{K}_{mn} = \mathcal{P}_m^V \mathcal{K} \mathcal{P}_n^U$.

Define a $m \times n$ matrix $\mathbf{C} = [c_{ij}]$ with

$$c_{ij} = \int_{B_i} \int_0^1 k(x,y) u_j(x) dx dy = \langle \mathcal{K} u_j, v_i \rangle \,,$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product in $L^2([0,1],\lambda_1)$. With a parametric set $\Theta_n \subset \mathbb{R}^n$, one then has a Poisson regression model for $\mathbf{n}$

$$P^t_{\mathbf{g}}(\mathbf{n}) = \prod_{i=1}^m (t g_i)^{n_i} (n_i!)^{-1} e^{-t g_i}$$

with $\mathbf{g} = [g_1,\dots,g_m]^T = \mathbf{C}\boldsymbol{\theta},\ \boldsymbol{\theta} \in \Theta_n$. The vector $\mathbf{g}$ represents the expected counts in the data space bins, and $\boldsymbol{\theta} = [\theta_1,\dots,\theta_n]^T$ represents the projection $\mathcal{P}_n^U f = \sum_{j=1}^n \theta_j u_j$. The vector $\boldsymbol{\theta}$ that corresponds to the true f will be denoted with $\boldsymbol{\theta}^0$, and the true vector of intensities with $\mathbf{g}^0 = [g_1^0, ..., g_m^0]^T$.

With $\gamma(t) \in (0,1]$ and with $\hat{\boldsymbol{\theta}} = [\hat{\theta}_1, ..., \hat{\theta}_n]^T$, we call

$$\hat{f}_t(x) = \sum_{j=1}^n \hat{\theta}_j u_j(x)$$

a quasi-maximum likelihood (QML) B-spline sieve estimator of f if

$$P^t_{\mathbf{C}\hat{\boldsymbol{\theta}}}(\mathbf{n}) \geq \gamma(t) \sup_{\boldsymbol{\theta} \in \Theta_n} P^t_{\mathbf{C}\boldsymbol{\theta}}(\mathbf{n}) \,.$$

As t increases, the discretization indices n and m are increased as well. For simplicity, the dependence of m and n on t is not marked explicitly in the notation. The same holds true for the matrix $\mathbf{C}$ and several other quantities.

It turns out that, due to discretization effects, it is necessary to modify the matrix $\mathbf{C}$ in order to obtain strongly L^2-consistent estimators. As in [12], let $\mathbf{G}$ be the Gram matrix of the functions $\{u_j\}$ and let $\mathbf{T} := \mathrm{diag}(\lambda_1(B_i))$. Write the singular value decomposition $\mathbf{T}^{-1/2}\mathbf{C}\mathbf{G}^{-1/2} = \mathbf{V}\mathrm{diag}(s_i)\mathbf{W}^T$, where $\mathbf{V}$ and $\mathbf{W} = [\mathbf{w}_1 \vdots \dots \vdots \mathbf{w}_n]$ are matrices with orthonormal columns and $\mathbf{w}_i$ denotes the ith column of $\mathbf{W}$. The numbers $s_1 \geq s_2 \geq \dots \geq s_n$ are then the singular values of $\mathcal{K}_{mn}$, and they approximate the singular values of $\mathcal{K}$ from below (see [12]). A modified or regularized matrix $\mathbf{C}_r$ that replaces $\mathbf{C}$ in the definition of the QML estimators is defined as

$$\mathbf{C}_r = \mathbf{T}^{1/2}\mathbf{V}\mathrm{diag}(r_i)\mathbf{W}^T\mathbf{G}^{1/2} \,,$$

where

$$r_i = \max\left\{s_i,\ C_0 n^{-(p-\alpha)/2}\right\}$$

and $\alpha < p$ and C_0 are some positive parameters. Under suitable assumptions, the QML B-spline sieve estimators with the matrix $\mathbf{C}_r$ in place of $\mathbf{C}$ may be proved to be strongly L^2-consistent and the convergence rates can be obtained (Theorems 3 and 4 in [12]). Those results are, however, not directly applicable to the SLW problem, because of a restrictive assumption of all data bins being of the same size, i.e., $\lambda_1(B_i) = \lambda_1(B_k)$, $i,k = 1,\ldots,m$, which is hard to satisfy for the SLW problem together with assumption C2 in Theorem 3 in [12]. Therefore, in this paper, we first generalize Theorems 1, 3 and 4 from [12] to cover also the case of non-uniform data binnings, and only then apply them to the SLW problem.

In the sequel, for a vector $\mathbf{x} = [x_1,\ldots,x_n] \in \mathbb{R}^n$, $\|\mathbf{x}\|$ stands for its Euclidean norm, $\|\mathbf{x}\|_1 = \sum_i |x_i|$ denotes its ℓ^1-norm and C is used as a generic constant.

With some arbitrary $m \times n$ matrix $\mathbf{A}$, consider a QML estimator $\hat{f}_t$, constructed with $\mathbf{A}$ in place of $\mathbf{C}$. Let $\lambda_{min}(\mathbf{A}^T\mathbf{A})$ be the minimal eigenvalue of $\mathbf{A}^T\mathbf{A}$ and $\lambda_{max}(\mathbf{G})$ the maximal eigenvalue of $\mathbf{G}$.

Theorem 1. *Assume that:*

A1. $m \geq n$ *and* $\log\gamma(t)^{-1} = O(m\log mt)$.
A2. $g_i^0 \asymp m^{-1}$ *and* $g_i \asymp m^{-1}$, $i = 1,...,m$, *for* $\mathbf{g} = \mathbf{A}\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \Theta_n$.
A3. $m = o(t)$ *and* $\lambda_{\max}(\mathbf{G})/\lambda_{\min}(\mathbf{A}^T\mathbf{A}) = O(t^\beta)$ *for some* $0 < \beta < 1$.
A4. $\|\mathbf{A}\boldsymbol{\theta}^0 - \mathbf{g}^0\|_1 = o(m\lambda_{\min}(\mathbf{A}^T\mathbf{A})/\lambda_{\max}(\mathbf{G}))$.

Then, with probability one, $\|\hat{f}_t - f\|_{L^2} \to 0$ *as* $t \to \infty$*, for all* f *such that* $\boldsymbol{\theta}^0 \in \Theta_n$ *for sufficiently large* n.

Notice that A4 is slightly weaker than the corresponding assumption $\|\mathbf{A}\boldsymbol{\theta}^0 - \mathbf{g}^0\| = o(m^{1/2}\lambda_{\min}(\mathbf{A}^T\mathbf{A})/\lambda_{\max}(\mathbf{G}))$ in [12], because $\|\mathbf{A}\boldsymbol{\theta}^0 - \mathbf{g}^0\|_1 \leq m^{1/2}\|\mathbf{A}\boldsymbol{\theta}^0 - \mathbf{g}^0\|$. In addition to other advantages discussed in the sequel, this small change allows for a more explicit interpretation of A4, with the minimal bin size involved only (cf. formula (7) in [12], in which the maximal bin size is used as well). To this end, set $\mathbf{A} = \mathbf{C}$, assume that $\min_i \lambda_1(B_i) \asymp m^{-1}$ and recall that $\lambda_{min}(\mathbf{G}) \asymp n^{-1}$ and $\lambda_{max}(\mathbf{G}) \asymp n^{-1}$ ([12], Lemma 2). The first part of Lemma 1 in [12] then gives $mn\lambda_{min}(\mathbf{C}^T\mathbf{C}) \geq C\lambda_{min}(\mathcal{K}_{mn}^*\mathcal{K}_{mn})$. Further,

$$\begin{aligned}\|\mathbf{C}\boldsymbol{\theta}^0 - \mathbf{g}^0\|_1 &= \|\mathcal{P}_m^V\mathcal{K}\mathcal{P}_n^U f - \mathcal{P}_m^V\mathcal{K}f\|_{L^1} \leq \|\mathcal{P}_m^V\mathcal{K}\mathcal{P}_n^U f - \mathcal{P}_m^V\mathcal{K}f\|_{L^2} = \\ &= O\left(\|\mathcal{P}_n^U f - f\|_{L^2}\right)\end{aligned} \tag{7}$$

(cf. [9], p.8, and use the Hölder inequality and the boundedness of $\mathcal{P}_m^V$ and $\mathcal{K}$). Consequently, with $\mathbf{A} = \mathbf{C}$ and $\min_i \lambda_1(B_i) \asymp m^{-1}$, it is sufficient for A4 that $\|\mathcal{P}_n^U f - f\|_{L^2} = o(\lambda_{min}(\mathcal{K}_{mn}^*\mathcal{K}_{mn}))$, which shows that A4 is indeed a crucial feasibility condition, as discussed in detail in [12].

Assume that

$$\Theta_n \subset \left\{\boldsymbol{\theta} \in \mathbb{R}^n : \sum_{i=1}^n i^{2a}(\mathbf{w}_i^T\mathbf{G}^{1/2}\boldsymbol{\theta})^2 < M\right\} \tag{8}$$

with some positive constants M and a. Condition (8) may be interpreted as a discrete version of the requirement that the Fourier coefficients of f with respect to right singular functions of $\mathcal{K}$ decay at a certain rate; cf. a related discussion in [12]. The following theorem is a generalized version of Theorem 3 in that paper. The assumption of all data bins being of the same size is replaced with a condition on the smallest bin size only. The largest bin size is allowed to decrease at an arbitrary rate. Moreover, the generalized theorem covers a broader range of the operator regularization parameter α.

Denote by W_2^p the Sobolev space of functions on $[0,1]$ with square integrable p-th derivative and let $\|\mathcal{K}\|_{HS}$ be the Hilbert-Schmidt norm.

Theorem 2. *Let $\hat{f}_t$ be a QML order p, B-spline sieve estimator of f constructed with the matrix $\mathbf{C}_r$ in place of $\mathbf{C}$, with parametric sets satisfying (8) and with data binning such that $C_1 \le m\lambda_1(B_i) \le C_2 m^{\Delta}$, $i = 1, \dots, m$, with some $C_1, C_2 > 0$ and $\Delta \in (0,1)$. Assume that the singular values σ_i of $\mathcal{K}$ decay as i^{-b} and that:*

B1. *$m \ge n$ and $\log \gamma(t)^{-1} = O(m \log mt)$.*
B2. *$g_i^0 \asymp m^{-1}$ and $g_i \asymp m^{-1}$, $i = 1, ..., m$, for $\mathbf{g} = \mathbf{C}\boldsymbol{\theta}$, $\boldsymbol{\theta} \in \Theta_n$.*
B3. *$\|\mathcal{K} - \mathcal{K}_{mn}\|_{HS} = O(n^{-r})$ with some $r > 0$.*

If either ("weak regularization regime")

B4. *$0 < \alpha < p - 2r$, $m^{\Delta} = o(n^{2ar/b-(p-\alpha)})$, $m^{\Delta+1} = o(n^{2ar/b+p-\alpha})$ and $mn^{p-\alpha} = O(t^{\beta})$ for some $\beta \in (0,1)$,*

or ("strong regularization regime")

B4'. *$p - 2r \le \alpha < p$, $m^{\Delta} = o(n^{(p-\alpha)(a-b)/b})$, $m^{\Delta+1} = o(n^{(p-\alpha)(a+b)/b})$ and $mn^{p-\alpha} = O(t^{\beta})$ for some $\beta \in (0,1)$,*

then, with probability one, $\|\hat{f}_t - f\|_{L^2} \to 0$ as $t \to \infty$, for all $f \in S_2^p$ such that $\boldsymbol{\theta}^0 \in \Theta_n$ for sufficiently large n.

Because $m \ge n$, the weak regularization regime is possible only if

$$p - \frac{2ar}{b} - \Delta < \alpha < p - \max\left\{2r, \Delta + 1 - \frac{2ar}{b}\right\} \tag{9}$$

and with

$$a > \frac{b}{2r}\left(\Delta + \max\{2r, 1/2\}\right), \tag{10}$$

which ensures that (9) gives a non-empty interval for α.

Similarly, the strong regularization regime is possible only if

$$p - 2r \le \alpha < p - \frac{b}{a+b}\max\left\{\Delta + 1, \Delta\frac{a+b}{a-b}\right\} \tag{11}$$

and with

$$a \ge \frac{b}{2r}\max\{\Delta + 2r, \Delta - 2r + 1\}. \tag{12}$$

With $p \le 2r$, only the strong regime is possible and $\alpha > 0$ provides a lower bound for α. In this case, one has a non-empty interval for α only if $a > b(\Delta + p)/p$.

With $\Delta = 0$ in the strong regularization regime, one obtains Theorem 3 from [12] as a special case.

For a fixed value of a, which implicitly defines the size of the function class to which f may belong, the parameters α and β and the discretization rates may be optimized to produce the fastest convergence rates. The following theorem describes the dependence of the convergence rate on the parameter α and allows, in any particular application, to choose α in the optimal way. For simplicity, only the case $m \asymp n$ is covered. It can be shown, however, that $n = o(m)$ does not lead to any improvements. Note that, with $m \asymp n$, the last part of B4 and B4' becomes $m \asymp n \asymp t^{\beta/(p-\alpha+1)}$.

Define the mean integrated square error of $\hat{f}_t$ as $\mathrm{MISE}(\hat{f}_t) = \mathrm{E}\|\hat{f}_t - f\|^2_{L^2}$.

Theorem 3. *Under the assumptions of Theorem 2, with* $m \asymp n \asymp t^{\beta/(p-\alpha+1)}$ *and with any positive* D, *MISE*$(\hat{f}_t) = O(t^{-s}\log t)$ *as* $t \to \infty$, *uniformly for* $f \in W_2^p$ *such that* $\|D^p f\|_{L^2} \le D$ *and* $\boldsymbol{\theta}^0 \in \Theta_n$ *for sufficiently large* n.

In the weak regularization regime, $s = 1 - \beta = \alpha/(p+1)$, *if* $\alpha \le 2ra/b - p - \Delta$ *and* $s = 1 - \beta = [2ra - b\Delta - b(p-\alpha)]/[2ra - b\Delta + b(p-\alpha) + 2b]$ *for larger* α. *In both cases* s *increases with* α.

In the strong regularization regime, $s = 1-\beta = \alpha/(p+1)$ *and* s *increases with* α, *if* $\alpha \le [p(a-b)-b\Delta]/(a+b)$, *and* $s = 1-\beta = [(p-\alpha)(a-b)-b\Delta]/[(p-\alpha)(a+b)+b(2-\Delta)]$ *and* s *decreases with* α, *for larger values of* α.

Setting $\Delta = 0$ in the strong regularization regime, one obtains Theorem 4 in [12] as a special case.

The first part of assumption B2 essentially means that all data bins should be approximately equally populated, which usually leads to a non-uniform binning in the data space. In the sequel, a special binning will be constructed for the SLW problem, suitable for functions f that are bounded and cut away from zero. For such functions, if $B_1 = [0, y_1]$ and $B_i = (y_{i-1}, y_i]$, $i = 2, \ldots, m$ with $y_m = 1$, one gets

$$g_i^0 = \int_{B_i} \int_0^1 2y\mathbf{1}_{\{y<x\}} f(x)dydx \asymp H(y_i) - H(y_{i-1})$$

with $H(y) = y^2(3-2y)$. Hence, if b_i are selected to satisfy $H(b_i) = i/m$, then $g_i^0 \asymp m^{-1}$ for $i = 1, \ldots, m$. Notice that $H'(y)$ takes its maximal value $3/2$ at $y = 1/2$ and $H'(0) = H'(1) = 0$. This means that the central bins are the smallest ones and $\min_i \lambda_1(B_i) \asymp m^{-1}$, as postulated in Theorem 2. The size of the largest bins tends, however, to zero at a slower rate ($\lambda_1(B_1) \asymp m^{-1/2}$), which means that $\Delta = 1/2$ should be set in Theorems 2 and 3 and shows that the work invested in generalizing the theorems was indeed necesary, in order to make them applicable to the SLW problem with functions f bounded and cut away from zero.

It then follows from Lemma 1 (see the Appendix) that, with the special binning defined by $H(\cdot)$, $\|\mathcal{K} - \mathcal{K}_{mn}\|_{HS} = O(n^{-1/4})$. In this setup, the properties of $\hat{f}_t$ in the SLW problem can be summarized as

Corollary 1. *Let a QML order* p, *B-spline sieve estimator* $\hat{f}_t$ *for* f *in the SLW problem be constructed with the matrix* $\mathbf{C}_r$ *in place of* $\mathbf{C}$, *with data binning defined by the function* $H(\cdot)$ *and with parametric sets satisfying* (8) *and such that* $0 < c \le$

$\sum_{j=1}^{n} \theta_j u_j(x) \le d$ *for some constants* c *and* d *and for* $x \in [0,1]$*. Assume that B1 holds true and that* $f \in S_2^p$ *is bounded and cut away from zero and such that* $\boldsymbol{\theta}^0 \in \Theta_n$ *for sufficiently large* n*. Then the best rates are obtained in the strong regularization regime:*

1. *If* $2 < a \le 4p$*, then* $\mathrm{MISE}(\hat{f}_t) = O(t^{-(a-2)/(a+4)} \log t)$*, with* $m \asymp n \asymp t^{4/(a+4)}$ *and* $\alpha = p - 1/2$.
2. *If* $a > 4p$*, then* $\mathrm{MISE}(\hat{f}_t) = O(t^{-[p(a-1)-1/2]/[(p+1)(a+1)]} \log t)$*, with* $m \asymp n \asymp t^{1/(p+1)}$ *and* $\alpha = [p(a-1) - 1/2]/(a+1)$.

In both cases $\hat{f}_t$ *is strongly* L^2*-consistent.*

Whether the rates given in Corollary are minimax is an open question, because no lower bounds for the minimax risk are known for the non-standard class of functions to which f is assumed to belong.

If f might be arbitrarily close to zero or unbounded, the special binning defined through the function $H(\cdot)$ need not, of course, lead to all g_i^0 of the same order. "Approximately equally populated data bins" remains, however, a paradigm in applications to real data sets.

It should be noticed that with uniform data binning one obtains $\|\mathcal{K} - \mathcal{K}_{mn}\|_{HS} = O(n^{-1/2})$, which leads to faster convergence rates. With $r = 1/2$ and $\Delta = 0$, the weak regime is possible with $a > 1$ and $p - a < \alpha < p - 1$, (cf. (9) and (10)), and the strong regime is possible with $a > 1$ and $p - 1 \le \alpha < p - 1/(a+1)$, (cf. (11) and (12)). Then, $s = (a-1)/(a+3)$, if $a < 2p + 1$, and $s = p(a-1)/[(p+1)(a+1)]$, if $a \ge 2p + 1$, and the rates are again obtained in the strong regime. It is, however, not quite clear how to express any natural conditions on f that may ensure B2 with the uniform data binning.

Also notice that, for "small" a (or "large" p), the convergence rates depend neither on the order of the splines, nor on the smoothness of f, both expressed in terms of p. This may be attributed to discretization effects (cf. a related discussion in [12]) and considered a drawback of the maximum likelihood approach to the analysis of binned data.

4. NUMERICAL EXAMPLE

The QML B-spline sieve estimators may be computed by means of the EMDS algorithm, described in detail in [11, 12]. In order to illustrate this approach and to compare its performance with more traditional methods, the SLW problem with data taken from Table 11.3 in [8], p. 298, was solved. The data formed an artificial sample of 1,000 points, grouped in 13 intervals of equal lengths, and were generated from a Rayleigh density. For the present example the range was rescaled to the $(0,1)$ interval. Additionally, to make our results comparable with those in Table 11.3, the unfolded function was normalized to be a probability density function.

In the implementation of the EMDS algorithm, a discrete approximation of the folding operator was needed. Let $B_i = (b_{i-1}, b_i]$, $i = 1, \ldots, m$, $b_0 = 0$, $b_m = 1$, be the data bins. For the EMDS implementation, the domain of the solution was also partitioned into a (large) number of subintervals $(a_{j-1}, a_j]$, $j = 1, \ldots, s$, $a_0 = 0$, $a_s = 1$. The discrete approximation of the operator was then represented by a matrix $[\bar{c}_{ji}]$, with $\bar{c}_{ji} = 2\int_{b_{j-1}}^{b_j}\int_{a_{i-1}}^{a_j} y\mathbf{1}_{\{y<x\}}(y)\mathrm{d}x\mathrm{d}y$, and elementary calculation gave $\bar{c}_{ij}$ in the form:

$$
\begin{array}{ll}
0 & \text{if } a_i \le b_{j-1} \\
\frac{1}{3}a_i^3 + \frac{2}{3}b_{j-1}^3 - b_{j-1}^2 a_i & \text{if } b_{j-1} < a_i \le b_j,\ a_{i-1} \le b_{j-1} \\
\frac{1}{3}(a_i^3 - a_{i-1}^3) - b_{j-1}^2(a_i - a_{i-1}) & \text{if } b_{j-1} < a_i \le b_j,\ a_{i-1} > b_{j-1} \\
a_i(b_j^2 - b_{j-1}^2) - \frac{2}{3}(b_j^3 - b_{j-1}^3) & \text{if } a_i > b_j,\ a_{i-1} \le b_{j-1} \\
\frac{1}{3}(b_j^3 - a_{i-1}^3) - b_{j-1}^2(b_j - a_{i-1}) & \\
\qquad +(a_i - b_j)(b_j^2 - b_{j-1}^2) & \text{if } a_i > b_j,\ b_{j-1} < a_{i-1} \le bj \\
(a_i - a_{i-1})(b_j^2 - b_{j-1}^2) & \text{if } a_{i-1} > b_j.
\end{array}
$$

Figure 1 shows the true function (smooth, solid line), the solution obtained with the EMDS algoritm with a sieve spanned by 13 cubic B-splines (solid, step-like line) and the solution obtained with a two-step algorithm proposed in [1] (dotted line). The latter is based on the last column in Table 11.3 in [8], and was also rescaled to the $(0, 1)$ interval.

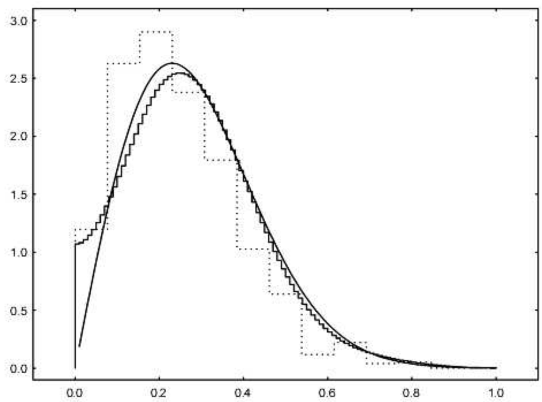

Fig. 1. True Rayleigh density (solid), the QML estimator (solid, step-like) and the Barthel-Klimanek-Stoyan estimator (dotted). The step-like representation of the QML estimator is due to its implementation via the EMDS algorithm

The parameters used in the EMDS algorithm (cf. [12]) were: $s = 100$, $J = 19$, $a = 2$ and $edf = 13$. $\mathbf{C}_r = \mathbf{C}$ was set and the edf parameter was selected to minimize a GCV-like criterion, as described in [11, 12]. It should be noticed that $edf = 13$ means that no so-called projection smoothing was applied.

Although the QML solution is clearly much more accurate than that obtained in [8] with the method of Barthel ([1]), more extensive simulation studies are needed to further investigate the potential of the QML approach to the SLW problem.

5. APPENDIX

Proof of Theorem 1. It may be proved (see [9], Corollary to Proposition 1) that, under A1 and A2, for $\epsilon > 0$ and $t > 6m$

$$\mathrm{P}\left(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\| > \epsilon\right) \le F \exp\left[-\left(4C\epsilon^2 m\lambda_{\min}(\mathbf{A}^T\mathbf{A}) - O(\|\mathbf{A}\boldsymbol{\theta}^0 - \mathbf{g}^0\|_1)\right) t\right],$$

where $F = F(m,t)$ and $\log F = O(m \log mt)$. Using that, a minor modification of the proof to Theorem 1 from [12] gives the thesis. □

Proof of Theorem 2. It will be proved that the assumptions of Theorem 1 are satisfied with $\mathbf{A} = \mathbf{C}_r$. Using Lemma 1 in [10] and then the Ostrowski theorem, as in [12], notice first that

$$\lambda_{min}(\mathbf{C}_r^T\mathbf{C}_r) = s^2_{min}(\mathbf{C}_r) \ge C \min_i \lambda_1(B_i)\, s^2_{min}\left(\mathbf{V}\mathrm{diag}(r_i)\mathbf{W}^T\mathbf{G}^{1/2}\right) =$$
$$= Cm^{-1}\lambda_{min}\left(\mathbf{G}^{1/2}\mathbf{W}\mathrm{diag}(r_i^2)\mathbf{W}^T\mathbf{G}^{1/2}\right) \ge C(mn)^{-1}n^{-(p-\alpha)},$$

where $s_{min}(\cdot)$ stands for the minimal singular value of a matrix. This gives

$$mn\lambda_{min}\left(\mathbf{C}_r^T\mathbf{C}_r\right) \ge Cn^{-(p-\alpha)}. \tag{13}$$

Assumption A3 takes the form

$$m = o(t) \text{ and } n^{-1} = O\left(t^\beta \lambda_{min}\left(\mathbf{C}_r^T\mathbf{C}_r\right)\right),$$

which is satisfied, because of (13) and the last part of B4 or B4'.

For A4, using (7) and the approximation rate n^{-p} of functions from W_2^p with order p, B-splines (Theorems 6.27 and 2.59 in [6]), write

$$\|\mathbf{C}_r\boldsymbol{\theta}^0 - \mathbf{g}^0\|_1 \le \|\mathbf{C}\boldsymbol{\theta}^0 - \mathbf{g}^0\|_1 + \|(\mathbf{C}_r - \mathbf{C})\boldsymbol{\theta}^0\|_1 \le$$
$$\le O\left(\|\mathcal{P}_n^U f - f\|_{L^2}\right) + m^{1/2}\|(\mathbf{C}_r - \mathbf{C})\boldsymbol{\theta}^0\| =$$
$$= O(n^{-p}) + m^{1/2}\|(\mathbf{C}_r - \mathbf{C})\boldsymbol{\theta}^0\|.$$

In view of (13), it is then sufficient for A4 that $m^{1/2}\|(\mathbf{C}_r - \mathbf{C})\boldsymbol{\theta}^0\| = o(n^{-(p-\alpha)})$. Denote $\delta_i = r_i - s_i$. Then, using the assumption on the data bins size and (8),

$$m^{1/2}\|(\mathbf{C}_r - \mathbf{C})\boldsymbol{\theta}^0\| \le C_2^{1/2} m^{\Delta/2}\|\mathrm{diag}(\delta_i)\mathbf{W}^T\mathbf{G}^{1/2}\boldsymbol{\theta}^0\| \le Cm^{\Delta/2}\left[\max_{1\le i\le n}\frac{\delta_i^2}{i^{2a}}\right]^{1/2}$$

and, reasoning as in the proof of Theorem 3 in [12], one obtains that it is sufficient for A4 that $m^\Delta n^{p-\alpha-2a\gamma/b} = o(1)$ with $\gamma = \min\{(p-\alpha)/2, r\}$, which is clearly satisfied in both weak and strong regularization regime.

In order to show that the second part of A2 holds true with $\mathbf{A} = \mathbf{C}_r$ (as needed for an application of Theorem 1) if it is true with $\mathbf{A} = \mathbf{C}$ (as assumed in the second part of B2) notice that

$$m\|(\mathbf{C}_r - \mathbf{C})\boldsymbol{\theta}\| \leq Cm^{(\Delta+1)/2}n^{-(p-\alpha)/2-a\gamma/b}$$

and (cf. [12]) it is sufficient to show that $m^{\Delta+1} = o(n^{p-\alpha+2a\gamma/b})$, which is obviously true in both regularization regimes. This completes the proof. □

Proof of Theorem 3. Write

$$\mathrm{MISE}(\hat{f}_t) = \|f - \mathcal{P}_n^U f\|_{L^2}^2 + \mathrm{E}\|\hat{f}_t - \mathcal{P}_n^U f\|_{L^2}^2 = O\left(n^{-2p}\right) + \int_0^\infty \mathrm{P}\left(\|\hat{f}_t - \mathcal{P}_n^U f\|_{L^2}^2 > x\right)\mathrm{d}x$$

and, because $\|\hat{f}_t - \mathcal{P}_n^U f\|_{L^2}^2 \leq \lambda_{max}(\mathbf{G})\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|^2 \leq Cn^{-1}\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\|^2$ (cf. [12], p. 214 and Lemma 2), one obtains

$$\begin{aligned}\mathrm{P}\left(\|\hat{f}_t - \mathcal{P}_n^U f\|_{L^2}^2 > x\right) &\leq \mathrm{P}\left(\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0\| > C(nx)^{1/2}\right) \leq \\ &\leq O\left(m \log mt\right) \exp\left[-\left(4C_1 x n^{-(p-\alpha)} - O\left(m^{1/2}\|(\mathbf{C}_r - \mathbf{C})\boldsymbol{\theta}^0\| + n^{-p}\right)\right) t\right],\end{aligned}$$

as in the proofs of Theorems 1 and 2. Further (cf. the proof of Theorem 2 above and of Theorem 3 in [12]),

$$m^{1/2}\|(\mathbf{C}_r - \mathbf{C})\boldsymbol{\theta}^0\| \leq Cm^{\Delta/2}n^{-[(p-\alpha)/2+\gamma a/b]} = Cn^{-[(p-\alpha)/2+\gamma a/b-\Delta/2]},$$

with $\gamma = \min\{(p-\alpha)/2, r\}$. Hence,

$$\mathrm{P}\left(\|\hat{f}_t - \mathcal{P}_n^U f\|_{L^2}^2 > x\right) \leq \exp\left[-\left(4C_1 x n^{-(p-\alpha)} - C_2 m t^{-1}\log mt - C_3 n^{-\delta}\right) t\right] \quad (14)$$

and $\delta = \min\{p, (p-\alpha)/2 + ra/b - \Delta/2\}$ in the weak regularization regime, and $\delta = \min\{p, (p-\alpha)(a+b)/(2b) - \Delta/2\}$ in the strong regularization regime.

Consider the strong regime first. If $\alpha \leq [p(a-b) - b\Delta]/(a+b)$, then $\delta = p$ and, reasoning as in the proof of Theorem 4 in [12], one obtains $s = \min\{\alpha\beta/(p-\alpha+1), 1-\beta\}$, which is maximal if $s = 1-\beta = \alpha/(p+1)$. If $\alpha > [p(a-b) - b\Delta]/(a+b)$, then $\delta = (p-\alpha)(a+b)/(2b) - \Delta/2$ and, reasoning as before, one obtains $s = \min\{1-\beta, \beta[(p-\alpha)(a-b)/(2b) - \Delta/2]/(p-\alpha+1)\}$. Balancing the two terms, one obtains the optimal s in the form given in the theorem and it is elementary to check that this optimal s decreases with increasing α.

In the weak regularization regime, if $\alpha \leq 2ra/b - p - \Delta$, then $\delta = p$ and one obtains $s = \alpha/(p+1)$, as in the strong regime. If $\alpha > 2ra/b - p - \Delta$, then $\delta = (p-\alpha)/2 + ra/b - \Delta/2$ and the last term in the exponent in (14) becomes negligible, if

$$x > n^{(p-\alpha)/2 - ra/b + \Delta/2}\log t = t^{-\beta[ra/b - \Delta/2 - (p-\alpha)/2]/(p-\alpha+1)}\log t.$$

As in [12], this leads to $s = \min\{1-\beta, \beta[ra/b - \Delta/2 - (p-\alpha)/2]/(p-\alpha+1)\}$ and, after balancing the two terms, to the optimal s in the form given in the theorem. Clearly, the optimal s increases with increasing α. This completes the proof. □

Proof of Corollary 1. The first part of B2 is, of course, fulfilled with the binning defined through the function $H(\cdot)$. For its second part, write

$$g_i = \sum_{j=1}^{n} c_{ij}\theta_j = \int_{B_i}\int_0^1 2y\mathbf{1}_{\{y<x\}}\sum_{j=1}^{n}\theta_j u_j(x)\mathrm{d}x\mathrm{d}y$$

and notice that this is again of the same order as $H(b_i)-H(b_{i-1}) \asymp m^{-1}$. With $a > 2$, the weak regularization regime is possible with $\max\{0, p-a/2+1/2\} < \alpha < p-1/2$, (cf. (9) and (10)), and the strong regime is possible with $p-1/2 \le \alpha < p-3/[2(a+1)]$, (cf. (11) and (12)). The conclusion then follows from considering two cases, in which $[p(a-1)-1/2]/(a+1)$ does, or does not belong to that interval, respectively. □

Lemma 1. *Let Δ_x be the mesh size of the set of x-knots and $\Delta_y = \max_j(y_j - y_{j-1})$ be the size of the largest data bin. Then, $\|\mathcal{K}-\mathcal{K}_{mn}\|_{HS}^2 = O(\Delta_x+\Delta_y)$ as $m, n \longrightarrow \infty$.*

Proof. The degenerated kernel k_{mn} of the finite-dimensional operator $\mathcal{K}_{mn}$ is the orthogonal projection in $L^2([0,1]^2, \lambda_2)$ of $k(y,x) = 2y\mathbf{1}_{\{y<x\}}$ onto the space spanned by tensor-product splines $u_j(x)\mathbf{1}_{B_i}(y)$, where $j = 1, \dots, n$ and $i = 1, \dots, m$. With $B_i = (y_{i-1}, y_i]$, one obtains

$$\|\mathcal{K}-\mathcal{K}_{mn}\|_{HS}^2 = \sum_{i=1}^{m}\int_0^1\int_{y_{i-1}}^{y_i}(k-k_{mn})^2\mathrm{d}y\mathrm{d}x.$$

Define $r(i) := \max\{k : x_k \le y_{i-1}\}$ and $s(i) := \min\{k : x_k \ge y_i\}$. The best L^2-approximation is not worse than

$$\tilde{k}(y,x) = \sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}u_j(x)\mathbf{1}_{B_i}(y),$$

with $a_{ij} = 0$, if $j < r(i)+p$ and $a_{ij} = y_{i-1}$, if $j \ge r(i)+p$. Notice that $u_j(x)$ is zero outside the interval $[x_{j-p}, x_j]$ and recall that B-splines u_j form a partition of unity; that is $\sum_j u_j = 1$. Define $S_i^{(1)} := B_i \times [0, x_{r(i)}]$, $S_i^{(2)} := B_i \times [x_{r(i)}, x_{s(i)+p-1}]$ and $S_i^{(3)} = B_i \times [x_{s(i)+p-1}, 1]$. In $S_i^{(1)}$, both k and $\tilde{k}$ are zero. In $S_i^{(2)}$, both k and $\tilde{k}$ are between 0 and y_i. In $S_i^{(3)}$, $\tilde{k}(y,x) = y_{i-1}$ and $y_{i-1} \le k(y,x) \le y_i$. Consequently,

$$\begin{aligned}\|\mathcal{K}-\mathcal{K}_{mn}\|_{HS}^2 &\le \sum_{i=1}^{m}\left[\int_{y_{i-1}}^{y_i}\int_{x_{r(i)}}^{x_{s(i)+p-1}} y_i^2\mathrm{d}x\mathrm{d}y + \int_{y_{i-1}}^{y_i}\int_{x_{s(i)+p-1}}^{1}(y_i-y_{i-1})^2\mathrm{d}x\mathrm{d}y\right] \le \\ &\le \sum_{i=1}^{m}(y_i-y_{i-1})\left[(x_{s(i)+p-1}-x_{r(i)}) + (y_i-y_{i-1})^2\right] \le \\ &\le \sum_{i=1}^{m}(y_i-y_{i-1})\left[\Delta_y + (p+1)\Delta_x + \Delta_y^2\right] = O(\Delta_x+\Delta_y),\end{aligned}$$

which completes the proof. □

Acknowledgements
Work supported by AGH local Grant 10.420.03.

REFERENCES

[1] M. Barthel, P. Klimanek, D. Stoyan, *Stereological substructure analysis in hot-deformed metals from TEM-images*, Czech. J. Phys. B **35** (1985), 265–268.

[2] I.M. Johnstone, B.W. Silverman, *Discretization effects in statistical inverse problems*, J. Complexity **7** (1991), 1–34.

[3] E. Kamke, *Differentialgleichungen. Lösungsmethoden und Lösungen* (6th ed., Leipzig 1959), Russian edition: Nauka, Moscow, 1971.

[4] G.W. Lord, T.F. Willis, *Calculation of air bubble distribution from results of a Rosiwal traverse of aerated concrete*, A.S.T.M. Bull. **56** (1951), 177–187.

[5] R.D. Reiss, *A course on point processes*, Springer, New York, 1993.

[6] L.L. Schumaker, *Spline functions: basic theory*, Krieger Publishing Company, Malabar, Florida, 1993.

[7] A.G. Spektor *Analysis of distribution of spherical particles in non-transparent structures*, Zavodsk. Lab. **16** (1950), 173–177.

[8] D. Stoyan, W.S. Kendall, L. Mecke, *Stochastic geometry and its applications*, Akademie-Verlag, Berlin, 1987.

[9] Z. Szkutnik, *Unfolding intensity function of a Poisson process in models with approximately specified folding operator*, Metrika **52** (2000), 1–26.

[10] Z. Szkutnik, *A note on quasi-maximum likelihood solutions to an inverse problem for Poisson processes*, Statist. Probab. Lett. **60** (2002), 253–263.

[11] Z. Szkutnik, *Doubly smoothed EM algorithm for statistical inverse problems*, J. Amer. Statist. Assoc. **98** (2003), 178–190.

[12] Z. Szkutnik, *B-splines and discretization in an inverse problem for Poisson processes*, J. Multiv. Anal. **93** (2005), 198–221.

[13] E.T. Whittaker, G.N. Watson, *A course of modern analysis. Part 2*, University Press, Cambridge, 1963.

Zbigniew Szkutnik
szkutnik@agh.edu.pl

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

Received: September 7, 2006.