

MAXIMUM SMOOTHED LIKELIHOOD DENSITY ESTIMATION FOR INVERSE PROBLEMS

BY P. P. B. EGGERMONT AND V. N. LARICCIA

University of Delaware

We consider the problem of estimating a pdf f from samples X_1, X_2, \dots, X_n of a random variable with pdf $\mathcal{K}f$, where \mathcal{K} is a compact integral operator. We employ a maximum smoothed likelihood formalism inspired by a nonlinearly smoothed version of the EMS algorithm of Silverman, Jones, Wilson and Nychka. We show that this nonlinearly smoothed algorithm is itself an EM algorithm, which helps explain the strong convergence properties of the algorithm. For the case of (standard) density estimation, that is, the case where \mathcal{K} is the identity, the method yields the standard kernel density estimators. The maximum smoothed likelihood density estimation technique is a regularization technique. We prove an inequality which implies the stability and convergence of the regularization method for the large sample asymptotic problem. Under minimal assumptions it also implies the a.s. convergence of the finite sample density estimate via a uniform version of the strong law of large numbers. Under extra regularity conditions we get a.s. convergence rates via a uniform version of the law of the iterated logarithm (under stronger conditions than usual).

1. Introduction. We consider the problem of estimating the probability density function f^* of the random variable Y with values in $\Omega \subset \mathbb{R}^d$, from the samples X_i , $i = 1, 2, \dots, n$, of the random variable X with values in $\Sigma \subset \mathbb{R}^d$, with probability density function g , which is related to f^* via a (linear) integral equation $\mathcal{K}f^*(x) = g(x)$, $x \in \Sigma$, or

$$(1.1) \quad \int_{\Omega} k(x, y) f^*(y) dy = g(x), \quad x \in \Sigma.$$

It is assumed that Ω and Σ are bounded sets. For inverse problems related to Fredholm integral equations of the first kind this is usually the case. We further assume that k is continuous and positive on $\Sigma \times \Omega$ and satisfies the normalization condition

$$(1.2) \quad \int_{\Sigma} k(x, y) dx = 1 \quad \text{for all } y \in \Omega.$$

Examples of problems of this type include the deconvolution problem [see Mendelsohn and Rice (1982) and Carroll and Hall (1988)] and stereology [see Silverman, Jones, Wilson and Nychka (1990) and Wilson (1989)]. See also

Received December 1992; revised May 1994.

AMS 1991 *subject classifications*. 62G07, 62G20, 65R30.

Key words and phrases. Nonparametric density estimation, maximum smoothed likelihood, integral equation of the first kind, regularization, smoothing, almost sure convergence.

Vardi and Lee (1993). There is a close similarity with regression-type problems in positron emission tomography [see Vardi, Shepp and Kaufman (1985)].

The problem of maximum likelihood estimation may be formulated as

$$(1.3) \quad \text{minimize } l_n(f) =_{\text{def}} -\frac{1}{n} \sum_{i=1}^n \log([\mathcal{K}f](X_i)) + \int_{\Omega} f(y) dy,$$

subject to $f \geq 0$. The added term $\int_{\Omega} f(y) dy$ guarantees that the optimal f is indeed a probability density function [cf. Silverman (1982)]. We consider the following algorithm for the solution of this minimization problem, with $f_1(y) = 1/|\Omega|$, for $y \in \Omega$,

$$(1.4) \quad f_{p+1}(y) = f_p(y) \cdot \frac{1}{n} \sum_{i=1}^n \frac{k(X_i, y)}{[\mathcal{K}f_p](X_i)}, \quad p \geq 2, y \in \Omega.$$

This is actually an EM algorithm, as shown by Vardi, Shepp and Kaufman (1985). Consequently, the algorithm (1.4) has very strong monotonicity properties, to wit,

$$(1.5) \quad l_n(f_p) - l_n(f_{p+1}) \geq D(f_{p+1}|f_p) \geq 0, \quad p \geq 1,$$

$$(1.6) \quad D(f_0|f_p) - D(f_0|f_{p+1}) \geq l_n(f_p) - l_n(f_0) \geq 0, \quad p \geq 1,$$

provided the problem (1.3) has a solution $f_0 \in L^1(\Omega)$. Here the Kullback-Leibler informational divergence between the nonnegative functions $f, \varphi \in L^1(\Omega)$ is defined as $D(f|\varphi) = \int_{\Omega} f(y) \log(f(y)/\varphi(y)) + \varphi(y) - f(y) dy$ [see Kemperman (1967)]. The trouble with all this is that (1.3) need not have an absolutely continuous solution. In fact, (1.3) is an ill-posed problem. Standard methods of dealing with this are maximum penalized likelihood methods [see Silverman (1982), Green (1990) and Cox and O'Sullivan (1990)]. We consider a variation of the smoothed EM approach suggested by Silverman, Jones, Wilson and Nychka (1990). It is interesting to note that this has a natural connection with the roughness penalty approach to the usual case of density estimation advocated by Good and Gaskins (1971).

The smoothed EM approach of Silverman, Jones, Wilson and Nychka (1990) deals with the ill-posedness of (1.3) by adding a smoothing step to the algorithm (1.4). Let $s_h(x, y)$ be a smoothing kernel on $\Omega \times \Omega$, with $s_h(z, y) = s_h(y, z)$, and define the smoothing operator S by

$$(1.7) \quad Sf(y) = \int_{\Omega} s_h(y, z) f(z) dz, \quad y \in \Omega,$$

for all $f \in L^1(\Omega)$. Typically, we will have

$$(1.8) \quad s_h(x, y) = h^{-d} A(h^{-1}(x - y)),$$

for (x, y) away from the boundary of $\Omega \times \Omega$, and corrections near the boundary. Here A is any pdf on \mathbb{R}^d with compact support and with $A(x) = A(-x)$ for all $x \in \mathbb{R}^d$. Note that in the limit as $h \rightarrow 0$ no smoothing is applied

at all. Now Silverman, Jones, Wilson and Nychka (1990) replace the algorithm (1.4) by the EMS algorithm

$$(1.9) \quad f_{p+1}(y) = S \left\{ f_p(\cdot) \frac{1}{n} \sum_{i=1}^n \frac{k(X_i, \cdot)}{[\mathcal{H}f_p](X_i)} \right\}(y), \quad p \geq 1, y \in \Omega.$$

Qualitatively, this algorithm works quite well (after discretization) but is hard to understand from a quantitative point of view. Although some results are known [see Kay (1992) and Latham and Anderssen (1993)], we prefer to apply the nonlinearly smoothed version of Eggermont (1992), that is, define the nonlinear smoothing operator \mathcal{N} by

$$(1.10) \quad [\mathcal{N}(f)](y) = \exp([S\{\log f\}](y)), \quad y \in \Omega,$$

and modify the algorithm (1.9) by adding the nonlinear smoothing step:

$$(1.11) \quad f_{p+1}(y) = S \left\{ [\mathcal{N}f_p](\cdot) \frac{1}{n} \sum_{i=1}^n \frac{k(X_i, \cdot)}{[\mathcal{N}f_p](X_i)} \right\}(y), \quad p \geq 1, y \in \Omega.$$

We refer to this algorithm as the NEMS algorithm. This algorithm has the remarkable property that it minimizes the modified negative log-likelihood problem

$$(1.12) \quad \text{minimize} \quad L_n(f) =_{\text{def}} -\frac{1}{n} \sum_{i=1}^n \log([\mathcal{N}(f)](X_i)) + \int_{\Omega} f(y) dy,$$

subject to $f \geq 0$, which actually has a continuous solution f^n . Moreover, the NEMS algorithm has properties similar to the original EM algorithm (1.4), namely, for $p \geq 1$,

$$(1.13) \quad L_n(f_p) - L_n(f_{p+1}) \geq D(f_{p+1}|f_p) \geq 0,$$

$$(1.14) \quad D(f^n|f_p) - D(S^{-1}f^n|S^{-1}f_{p+1}) \geq L_n(f_p) - L_n(f^n) \geq 0.$$

These inequalities guarantee that the f_p generated by algorithm (1.11) converge to the solution f^n in the $C(\Omega)$ -norm. The proofs of these two inequalities, as well as the proof of the existence of a unique, continuous solution f^n of (1.12), are identical to those given for the large sample asymptotic problem in Eggermont (1992) and are therefore omitted. Inequalities (1.13) and (1.14) suggest that (1.11) is itself an EM algorithm. This is indeed the case (see Section 3).

Besides the convergence of the NEMS algorithm (1.11) to the solution f^n of the minimization problem (1.12), there is also the convergence of f^n to f^* as $n \rightarrow \infty$. The main thrust of this paper is in showing this convergence under minimal regularity conditions, as well as in providing rates of convergence under strong(er) regularity conditions. It seems obvious that convergence can only occur if $h \rightarrow 0$, that is, if the smoothing operator S (and \mathcal{N}) tends to the identity operator as $n \rightarrow \infty$. The easiest way to quantify all this is first to

consider the convergence of f^n to f_S , the solution of the limiting form of (1.12) with infinite sample size for a fixed smoothing operator S ,

$$(1.15) \quad \text{minimize} \quad L_\infty(f) =_{\text{def}} - \int_{\Sigma} g(x) \log([\mathcal{K}\mathcal{N}(f)](x)) dx + \int_{\Omega} f(y) dy,$$

subject to $f \geq 0$, and then the convergence of f_S to f^* as $h \rightarrow 0$. There is a crucial inequality which makes all this possible, quite analogous to, for example, the analysis of maximum entropy regularization [Eggermont (1993)]. The inequality describes the behavior of the functional $L_\infty(f)$ in the minimization problem (1.15) around its minimum,

$$(1.16) \quad \begin{aligned} L_\infty(f) - L_\infty(f_S) &\geq \frac{1}{2} \left\{ \int_{\Sigma} \sqrt{r_S} |\mathcal{K}\{ \mathcal{N}(f) - \mathcal{N}(f_S) \}| \right\}^2 \\ &\quad + c_S \|f - f_S\|_{L^1(\Omega)}^2, \end{aligned}$$

for all pdf's f , where c_S is a nonnegative constant and $[r_S](x) = [\mathcal{K}^*(g/\mathcal{K}\mathcal{N}f_S)](x)$. We can prove that $c_S \geq \frac{1}{8}|\Omega| \inf s_h(x, y)$, which is why we require that Ω is bounded and that s_h is bounded away from 0. However, we do not think these assumptions are necessary. We use inequality (1.16) in two ways. On the one hand, if in (1.16) we take $f = f^*$, the true density, then under mild conditions it is possible to show that $f_S \rightarrow f^*$ in $L^1(\Omega)$ as $h \rightarrow 0$. Under suitable regularity conditions, we also get convergence rates. On the other hand, if in (1.16) we take $f = f^n$, then we get that

$$(1.17) \quad c_S \|f^n - f_S\|_{L^1(\Omega)}^2 \leq L_\infty(f^n) - L_\infty(f_S) \leq L_n(f^n, f_S) - L_\infty(f^n, f_S),$$

where

$$(1.18) \quad L_n(f, \varphi) =_{\text{def}} L_n(f) - L_n(\varphi) = \frac{1}{n} \sum_{i=1}^n \log \frac{\mathcal{K}\mathcal{N}\varphi(X_i)}{\mathcal{K}\mathcal{N}f(X_i)},$$

and likewise for $L_\infty(f, \varphi)$. Thus all that is required to establish a.s. rates for the convergence $f^n \rightarrow f^*$ is showing that, under suitable regularity conditions,

$$(1.19) \quad \sup_{f, \varphi \in \mathcal{E}} \frac{|L_n(f, \varphi) - L_\infty(f, \varphi)|}{\|\mathcal{N}f - \mathcal{N}\varphi\|_{L^1(\Omega)}} = \mathcal{O} \left(\left(\frac{\log \log n}{n} \right)^{1/2} \right) \quad \text{a.s.},$$

where

$$(1.20) \quad \mathcal{E} =_{\text{def}} \left\{ f \in L^1(\Omega) : f \text{ is a pdf, } \int_{\Omega} \mathcal{N}f(y) dy \geq \eta \right\},$$

in which η is a fixed positive number. Combining the above results we then get a.s. convergence rates of f^n to f^* as $h \rightarrow 0$ in a way commensurate with $n \rightarrow \infty$. An alternative approach to convergence rates would be that of Cox and O'Sullivan (1990).

We finish the Introduction by noting that problem (1.12) is interesting even for the case that \mathcal{K} is the identity operator, that is, for the usual case of

density estimation. In this case one verifies that the solution to (1.12) is given by

$$(1.21) \quad f^n(y) = \frac{1}{n} \sum_{i=1}^n s_h(X_i, y), \quad y \in \Omega,$$

which is a well-known density estimator. In particular, we get kernel-type density estimators this way, as well as their interpretation as maximum smoothed likelihood density estimators [see Eggermont and LaRiccia (1995)].

2. Summary of assumptions and results. In this section we state the precise conditions and results of this paper. We will comment on the sharpness of these conditions as the occasion arises.

About the kernel k we assume that it is positive, integrable and satisfies

$$(2.1) \quad \int_{\Sigma} k(x, y) dx = 1 \quad \text{for all } y \in \Omega,$$

$$(2.2) \quad \kappa =_{\text{def}} \inf\{k(x, y) : (x, y) \in \Sigma \times \Omega\} > 0,$$

$$(2.3) \quad \omega(\mathcal{X}, h) =_{\text{def}} \text{ess sup}_{x \in \Sigma, y \in \Omega} |k(x + h, y) - k(x, y)| \rightarrow 0 \quad \text{as } |h| \rightarrow 0.$$

The third condition implies that \mathcal{X} is a compact mapping from $L^1(\Omega)$ into $C(\Sigma)$ and also as a mapping from $L^1(\Omega)$ into $L^1(\Sigma)$. We refer to these conditions on k as the *minimal conditions*. In order to get good rates (2.3) must be strengthened to

$$(2.4) \quad k \in C^{\max(2, d)}(\Sigma \times \Omega),$$

that is, if $d > 2$, then k is d times continuously differentiable. In particular, then

$$(2.5) \quad \|\mathcal{D}_x k\|_{C(\Sigma \times \Omega)} =_{\text{def}} \sup_{(x, y) \in \Sigma \times \Omega} |\mathcal{D}_x k(x, y)| < \infty,$$

where $\mathcal{D}_x = \partial_1 \partial_2 \cdots \partial_d$, in which ∂_i denotes differentiation with respect to x_i . If there is no confusion possible, we will just write \mathcal{D} for \mathcal{D}_x .

For our smoothing kernels we choose convolution kernels with boundary corrections. To simplify the presentation, assume that Ω is the unit cube in \mathbb{R}^d . We define our smoothers as

$$(2.6) \quad s_h(x, y) = \frac{h^2}{|\Omega|} + (1 - h^2) \prod_{i=1}^d t_h(x_i, y_i), \quad (x, y) \in \Omega \times \Omega,$$

where $x = (x_1, x_2, \dots, x_d)$, $y = (y_1, y_2, \dots, y_d)$ and, for u and v real numbers,

$$(2.7) \quad t_h(u, v) = \begin{cases} 0, & |u - v| \geq h, \\ (2h)^{-1}, & h < u + v < 2 - h \text{ and } |u - v| < h, \\ h^{-1}, & 0 < u + v < h \\ & \text{or } 2 - h < u + v < 2 \text{ and } |u - v| < h. \end{cases}$$

So, $t_h(u, v) = 1/2h$ when $|u - v| < h$, except in the $(0, 0)$ and $(1, 1)$ corners of $[0, 1] \times [0, 1]$, where it equals $1/h$ on two little triangles. If t_h is deemed to be not smooth enough, then the corresponding smoothing operator T_h defined by (1.7) can be replaced by some fixed power $(T_h)^p$ of T_h . The essential feature of these smoothers is that, for some probability density function A with compact support in \mathbb{R}^d and symmetric about $x = 0$, we have

$$(2.8) \quad s_h(x, y) = (1 - h^2)h^{-d} A(h^{-1}(x - y)) + \frac{h^2}{|\Omega|},$$

for (x, y) away from the boundary of $\Omega \times \Omega$, and with corrections near the boundary. We believe that any reasonable smoother of this kind will work (see Section 7). It is helpful to introduce the functional

$$(2.9) \quad \Lambda_S(f) = h^{-2} \int_{\Omega} \{f(y) - [\mathcal{M}f](y)\} dy.$$

We will show that, for all f with $\log f \in C^3(\Omega)$,

$$(2.10) \quad \Lambda_{\mathcal{S}}(f) =_{\text{def}} \lim_{h \rightarrow 0} \Lambda_S(f) = \frac{1}{6} \int_{\Omega} \frac{|\nabla f|^2}{f} + \int_{\Omega} (f - u) \log\left(\frac{f}{u}\right).$$

Here $u(y) = 1/|\Omega|$ for all $y \in \Omega$, and ∇f denotes the gradient of f . We note that the first term on the right-hand side is the roughness penalty functional of Good and Gaskins (1971).

Before stating our results we introduce the abbreviations

$$(2.11) \quad LL(n) = \left(\frac{\log \log n}{n} \right)^{1/2}, \quad \omega(n) = \omega(\mathcal{K}, (LL(n))^{1/d}).$$

THEOREM 2.1. *Under the minimal conditions (2.1)–(2.3) on k we have that*

$$(2.12) \quad \varepsilon_n =_{\text{def}} \sup_{f, \varphi \in \mathcal{E}} \frac{|L_n(f, \varphi) - L_{\infty}(f, \varphi)|}{\|\mathcal{M}f - \mathcal{M}\varphi\|_{L^1(\Omega)}} \leq \text{const} \cdot \omega(n) \quad a.s.,$$

where \mathcal{E} is defined in (1.20), and

$$(2.13) \quad \|f^n - f_S\|_{L^1(\Omega)} \leq h^{-2} \omega(n) \rightarrow 0 \quad a.s.,$$

provided $\sqrt{\omega(n)}/h \rightarrow 0$. Here S may depend on n .

THEOREM 2.2. *If k satisfies conditions (2.1)–(2.3) and (2.5), then*

$$(2.14) \quad \varepsilon_n \leq \text{const} \cdot \|\mathcal{D}_x k\|_{C(\Sigma \times \Omega)} \cdot LL(n) \quad a.s.$$

If, in addition, f^ satisfies the smoothness assumption $\Lambda_{\mathcal{S}}(f^*) < \infty$, then*

$$(2.15) \quad \|f^n - f_S\|_{L^1(\Omega)} \leq \text{const} \cdot (h^{-2} \cdot LL(n) + (LL(n))^{1/2}) \quad a.s.,$$

where the constants do not depend on n nor h .

In Theorem 2.2 we made the minimal smoothness assumption on f^*

$$(2.16) \quad \Lambda_{\mathcal{S}}(f^*) < \infty.$$

A tacit assumption here is that the solution to the equation $\mathcal{K}f = g$ is unique.

THEOREM 2.3. *If k satisfies assumptions (2.1)–(2.3) and f^* satisfies (2.16), then*

$$\|f_S - f^*\|_{L^1(\Omega)} \rightarrow 0 \quad \text{for } h \rightarrow 0.$$

In order to get convergence *rates*, we strengthen the smoothness condition (2.16) and assume that the following hold:

$$(2.17) \quad f^*, \log f^* \in C^3(\Omega), \quad \partial_n f^* = 0 \quad \text{on } \partial\Omega;$$

$$(2.18) \quad \text{there exists a } \lambda_0 \in L^\infty(\Sigma) \text{ such that } \Lambda'_\mathcal{F}(f^*) = \mathcal{K}^* \lambda_0.$$

Here $\partial_n f$ is the normal derivative of f on the boundary $\partial\Omega$ of Ω . These conditions can be relaxed somewhat, at the expense of slower convergence in the theorems below. We will not explore this further. The assumption on the natural boundary conditions (2.17) reminds one of similar conditions in spline smoothing and regression.

THEOREM 2.4. *Under the assumptions of Theorem 2.3, if k satisfies (2.4) and f^* satisfies (2.17) and (2.18), then*

$$\|f_S - f^*\|_{L^1(\Omega)} \leq \text{const} \cdot h.$$

The above results allow us to prove convergence of f^n to f^* .

THEOREM 2.5. (a) *Under the minimal assumptions on k and f^* , we have*

$$\|f^n - f^*\|_{L^1(\Omega)} \rightarrow 0 \quad \text{a.s.,}$$

provided $\omega(n)/h \rightarrow 0$ and $h \rightarrow 0$.

(b) *If in addition k satisfies condition (2.4), then*

$$\|f^n - f^*\|_{L^1(\Omega)} \leq \text{const} \cdot \left(\frac{\log \log n}{n} \right)^{1/6},$$

provided $h \approx (\log \log n/n)^{1/6}$.

PROOF. The basic ingredient is the triangle inequality

$$\|f^n - f^*\|_{L^1(\Omega)} \leq \|f^n - f_S\|_{L^1(\Omega)} + \|f_S - f^*\|_{L^1(\Omega)}.$$

Part (a) now follows from Theorem 2.1, (2.12) and Theorem 2.3. For part (b) we have, from Theorem 2.2, (2.15), Theorem 2.4 and the triangle inequality, that

$$\|f^n - f^*\|_{L^1(\Omega)} \leq \text{const} \cdot (h_S^{-2} \cdot LL(n) + (LL(n))^{1/2} + h).$$

If we now take $h \approx (LL(n))^{1/3}$ the required estimate emerges. \square

REMARK 2.6. If we had been able to strengthen (2.12) to

$$(2.19) \quad \begin{aligned} & L_n(f, \varphi) - L_\infty(f, \varphi) \\ & \leq \text{const} \cdot \|\mathcal{K}\{\mathcal{N}f - \mathcal{N}\varphi\}\|_{L^1(\Sigma)} \cdot \left(\frac{\log \log n}{n} \right)^{1/2} \quad \text{a.s.}, \end{aligned}$$

for all $f, \varphi \in \mathcal{E}$, then we would have been able to replace the estimate in Theorem 2.5(b) by

$$(2.20) \quad \|f^n - f^*\|_{L^1(\Omega)} \leq \text{const} \cdot \left(\frac{\log \log n}{n} \right)^{1/4} \quad \text{a.s.},$$

provided $h \approx (\log \log n/n)^{1/4}$. We believe that (2.20) would be the best result possible and that the correct order is at least $\mathcal{O}(n^{-1/4+\delta})$, for $\delta > 0$, but otherwise arbitrary. However, *proving* this is another matter.

In Sections 4–6 the results of this section are proved. The properties of the smoothing operators which are necessary for the analysis are collected in Section 7.

3. The EM algorithm. Here we show that the NEMS algorithm (1.11) is an EM algorithm for a modified maximum likelihood problem with missing data, which reduces to problem (1.12) when the samples satisfy certain conditions. According to Vardi, Shepp and Kaufman (1985), inequalities (1.13) and (1.14) then follow from Csiszár and Tusnády (1984).

If we are to apply, in the style of Vardi, Shepp and Kaufman (1985), an EM algorithm to the problem (1.12), then it seems imperative that $k(x, y)[\mathcal{N}f^*](y)$ be a pdf for some random variable (X, Y) . However, it is not. (Note that the arithmetic–geometric mean inequality says that $[\mathcal{N}f](y) \leq [Sf](y)$ for all pdfs f , so that $\int_\Omega [\mathcal{N}f](y) dy \leq 1$, with equality only if f is the uniform distribution on Ω .) However, this problem can be circumvented by replacing the random variable Y with a random variable B taking values in $\Omega \cup e$, where e is an exceptional event with probability $1 - \int_\Omega \mathcal{N}f^*$, and likewise replacing X by a new random variable A . More formally, we introduce a random variable (A, B) taking values in $(\Sigma \cup \varepsilon) \times (\Omega \cup e)$, where $\varepsilon \notin \Sigma$ and $e \notin \Omega$ are *exceptional* events, with pdf given by

$$(3.1) \quad h(a, b) = \begin{cases} k(a, b)[\mathcal{N}f^*](b), & (a, b) \in \Sigma \times \Omega, \\ d(f^*), & (a, b) = (\varepsilon, e), \\ 0, & \text{otherwise.} \end{cases}$$

Here $d(f) = 1 - \int_\Omega [\mathcal{N}f](y) dy$. Direct calculation shows that the marginal distribution of A satisfies

$$h_A(a; f^*) = \begin{cases} [\mathcal{K}\mathcal{N}f^*](a), & a \in \Sigma, \\ d(f^*), & a = \varepsilon. \end{cases}$$

Now assume that we have a random sample (A_i, B_i) , $i = 1, 2, \dots, n$, but that the B_i are missing. The maximum likelihood estimate of f^* is then the solution to

$$(3.2) \quad \text{minimize} \quad -\frac{1}{n} \sum_{i=1}^n \log(h_A(A_i; f)) + \int_{\Omega} f(y) dy,$$

subject to $f \geq 0$. Note that when $A_i \in \Sigma$ for all i then this problem is equivalent to (1.12). In order to derive the EM algorithm for the new problem, we need to consider the negative log-likelihood for the complete data set. It is given by

$$\Lambda(f) = -\frac{1}{n} \sum_{A_i \in \Sigma} \log k(A_i, B_i) - \frac{1}{n} \sum_{A_i \in \Sigma} \log[\mathcal{N}f](B_i) - \frac{I}{n} \log[d(f)],$$

where I denotes the number of samples (A_i, B_i) which equal (ε, e) . The EM algorithm now requires us to compute $\mathbf{E}[\Lambda(f)|A_1, \dots, A_n; f^p]$, where f^p is our current guess for f . Since $\log k(A_i, B_i)$ does not depend on f we may drop it from the calculations, and compute

$$\begin{aligned} Q(f|f^p) &= \mathbf{E} \left[-\frac{1}{n} \sum_{A_i \in \Sigma} \log[\mathcal{N}f](B_i) - \frac{I}{n} \log d(f) \middle| A_1, \dots, A_n; f^p \right] \\ &= -\frac{1}{n} \sum_{A_i \in \Sigma} \mathbf{E}[\log[\mathcal{N}f](B_i)|A_1, \dots, A_n; f^p] \\ &\quad - \frac{\log d(f)}{n} \mathbf{E}[I|A_1, \dots, A_n; f^p] \\ &= -\frac{n-I}{n} \int_{\Omega} \log[\mathcal{N}f](b) \psi^p(b) db - \frac{I}{n} \log[d(f)], \end{aligned}$$

where

$$\psi^p(b) = \frac{1}{n} \sum_{A_i \in \Sigma} \frac{k(A_i, b)[\mathcal{N}f^p](b)}{[\mathcal{N}f^p](A_i)}.$$

This completes the E step of the EM algorithm. For the M step we need to maximize $Q(f|f^p)$ over all pdf's f . Note that in general the answer will depend on I . However, the interest lies in the case $I = 0$ only, so that we need to minimize

$$(3.3) \quad - \int_{\Omega} \psi^p(b) \log[\mathcal{N}f](b) db = - \int_{\Omega} [S\psi^p](b) \log f(b) db.$$

Here we used that $\log \mathcal{N}f = S[\log f]$. Following the argument from Eggermont and LaRiccia (1995), we obtain that the solution is given by $f = S\psi^p$, so that $f^{p+1} = S\psi^p$. This is precisely the iterative step of the NEMS algorithm (1.11). Thus the NEMS algorithm is indeed an EM algorithm.

4. The inequality. In this section we prove inequality (1.16). From Kemperman (1967) we know that

$$(u - v)^2 \leq \left(\frac{2}{3}u + \frac{4}{3}v \right) \left(u \log \frac{u}{v} + u - v \right),$$

for all $u, v \in \mathbb{R}$. Thus with $u = \mathcal{N}f$ and $v = \mathcal{N}f_S$, and multiplying both sides by $r_S = g/\mathcal{N}f_S$ we get from the Cauchy-Schwarz inequality that

$$\begin{aligned} & \left(\int_{\Sigma} \sqrt{[r_S](x)} |[\mathcal{N}(f - f_S)](x)| dx \right)^2 \\ & \leq \left(\int_{\Sigma} \frac{2}{3}[\mathcal{N}f](x) + \frac{4}{3}[\mathcal{N}f_S](x) dx \right) \\ & \quad \times \left(\int_{\Sigma} g(x) \log \frac{[\mathcal{N}f_S](x)}{[\mathcal{N}f](x)} + [r_S](x)([\mathcal{N}f](x) - [\mathcal{N}f_S](x)) dx \right). \end{aligned}$$

Since $[\mathcal{N}f](x) \leq [Sf](x)$ (see Lemma 7.1), it follows that

$$\int_{\Omega} \frac{2}{3}[\mathcal{N}f](x) + \frac{4}{3}[\mathcal{N}f_S](x) dx \leq 2.$$

As for the remaining factor, first note that

$$\int_{\Sigma} g(x) \log \frac{[\mathcal{N}f_S](x)}{[\mathcal{N}f](x)} dx = L_{\infty}(f) - L_{\infty}(f_S),$$

as well as the fact that

$$\int_{\Sigma} [r_S](x)([\mathcal{N}f](x) - [\mathcal{N}f_S](x)) dx = \int_{\Omega} \left(\frac{[\mathcal{N}f](y)}{[\mathcal{N}f_S](y)} - 1 \right) F_S(y) dy,$$

with $F_S = \mathcal{N}(f_S) \mathcal{N}^* r_S$. Since f_S is a fixed point of the NEMS iteration (1.11), it follows that $SF_S = f_S$. Consequently,

$$\int_{\Omega} \left[S \left(\frac{f}{f_S} \right) \right](y) F_S(y) dy = \int_{\Omega} \frac{f(x)}{f_S(x)} [SF_S](x) dx = \int_{\Omega} f(x) dx = 1,$$

and so

$$\int_{\Sigma} [r_S](x)([\mathcal{N}f](x) - [\mathcal{N}f_S](x)) dx = \int_{\Omega} \left\{ \mathcal{N} \left(\frac{f}{f_S} \right) - S \left(\frac{f}{f_S} \right) \right\} F_S.$$

Since $\mathcal{N}(\psi) \leq \{S\sqrt{\psi}\}^2$ (see Lemma 7.1), we obtain, with $\varphi = \sqrt{f/f_S}$,

$$\left[S \left(\frac{f}{f_S} \right) \right](x) - \left[\mathcal{N} \left(\frac{f}{f_S} \right) \right](x) \geq [S(\varphi^2)](x) - ([S\varphi](x))^2.$$

Summarizing, we have shown that

$$\begin{aligned} & \frac{1}{2} \left\{ \int_{\Sigma} \sqrt{[r_S](x)} |[\mathcal{N}(f) - \mathcal{N}(f_S)](x)| dx \right\}^2 \\ (4.1) \quad & \leq L_{\infty}(f) - L_{\infty}(f_S) \\ & \quad - \int_{\Omega} \left\{ [S(\varphi^2)](x) - ([S\varphi](x))^2 \right\} F_S(x) dx. \end{aligned}$$

We now prove that, for all φ ,

$$(4.2) \quad \int_{\Omega} \{S(\varphi^2) - (S\varphi)^2\} F_S \geq h^2 \int_{\Omega} f_S(y) (\varphi(y) - \lambda)^2 dy,$$

for a suitable λ , and with $h^2 = |\Omega| \inf s_h(x, y)$. We then show how, in combination with (4.1), this implies our inequality (1.16).

From the Lagrange identity

$$[S(\varphi^2)](x) - ([S\varphi](x))^2 = \int_{\Omega \times \Omega} s_h(x, y) s_h(x, z) (\varphi(y) - \varphi(z))^2 dy dz$$

we obtain that the left-hand side of (4.2) equals

$$(4.3) \quad \int_{\Omega \times \Omega} (\varphi(y) - \varphi(z))^2 \left\{ \int_{\Omega} F_S(x) S_h(x, y) s_h(x, z) dx \right\} dy dz.$$

Since $s_h(x, z) \geq h^2/|\Omega|$, the expression in curly brackets is bounded below by

$$\frac{h^2}{|\Omega|} \int_{\Omega} F_S(x) s_h(x, y) dx = \frac{h^2}{|\Omega|} f_S(y).$$

It follows that the expression in (4.3) dominates the expression

$$\frac{h^2}{|\Omega|} \int_{\Omega \times \Omega} f_S(y) (\varphi(y) - \varphi(z))^2 dy dz,$$

which in turn dominates

$$\begin{aligned} & \frac{h^2}{|\Omega|} \int_{\Omega} \min_{z \in \Omega} \int_{\Omega} f_S(y) (\varphi(y) - \varphi(z))^2 dy dz \\ & \geq \frac{h^2}{|\Omega|} \int_{\Omega} \min_{\mu} \int_{\Omega} f_S(y) (\varphi(y) - \mu)^2 dy dz \\ & = h^2 \int_{\Omega} f_S(y) (\varphi(y) - \lambda)^2 dy \end{aligned}$$

with $\lambda = \int_{\Omega} \varphi(y) f_S(y) dy$. This establishes inequality (4.2).

We now use (4.2) with $\varphi = \sqrt{f/f_S}$ and $\lambda = \int_{\Omega} \sqrt{ff_S}$ to obtain that

$$\int_{\Omega} \{[S(\varphi^2)](x) - ([S\varphi](x))^2\} F_S(x) dx \geq h^2 \|\sqrt{f} - \lambda \sqrt{f_S}\|_{L^1(\Omega)}^2.$$

Since f and f_S are both pdf's, and interpreting λf_S as the orthogonal projection of f onto the line through 0 and f_S , it is easy to see that

$$\frac{1}{2} \|\sqrt{f} - \sqrt{f_S}\|_{L^2(\Omega)}^2 \leq \|\sqrt{f} - \lambda \sqrt{f_S}\|_{L^2(\Omega)}^2 \leq \|\sqrt{f} - \sqrt{f_S}\|_{L^2(\Omega)}^2,$$

and from Cauchy-Schwarz that

$$\frac{1}{4} \|f - f_S\|_{L^1(\Omega)}^2 \leq \|\sqrt{f} - \sqrt{f_S}\|_{L^2(\Omega)}^2.$$

The inequality now follows, with the constant $\frac{1}{8} |\Omega| \inf s_h(x, y)$. This is (1.16).

5. Almost sure convergence of $L_\infty(f^n) - L_\infty(f_S)$. In this section Theorems 2.1 and 2.2 are considered. For brevity, only Theorem 2.2 is proven.

We start out with a lemma which implies that the $\mathcal{N}f^n$ and $\mathcal{N}f_S$ are bounded away from 0. In particular, then the summands or integrands in $L_n(f^n)$, $L_\infty(f^n)$ and so on are continuous, uniformly in n and h .

LEMMA 5.1. *There exists a positive constant η such that, for all n and h ,*

$$\int_{\Omega} \mathcal{N}f^n(y) dy \geq \eta \quad \text{a.s.} \quad \text{and} \quad \int_{\Omega} \mathcal{N}f_S(y) dy \geq \eta.$$

PROOF. Note that $L_n(f^n) \leq L_n(f^*)$, which is bounded a.s. by a constant, say $\log \alpha$. It follows that $\prod_{i=1}^n [\mathcal{N}f^n](X_i) \geq \alpha^{-n}$, whence it follows that $\max_i [\mathcal{N}f^n](X_i) \geq \alpha^{-1}$. Since $\mathcal{X}: L^1(\Omega) \rightarrow C(\Sigma)$ is bounded, its norm, denoted by $\|\mathcal{X}\|_{1,\infty}$, is finite. Then the inequalities $[\mathcal{N}f^n](X_i) \leq \|\mathcal{N}f^n\|_{C(\Sigma)} \leq \|\mathcal{X}\|_{1,\infty} \|\mathcal{N}f^n\|_{L^1(\Omega)}$ imply the first statement. The second statement follows similarly from the inequality $L_\infty(f_S) \leq L_\infty(f^*)$. \square

COROLLARY 5.2. *There exists a positive constant η such that, for all n and h ,*

$$[\mathcal{N}f^n](x) \geq \kappa \cdot \eta \quad \text{a.s.} \quad \text{and} \quad [\mathcal{N}f_S](x) \geq \kappa \cdot \eta.$$

PROOF. Both statements follow from $[\mathcal{N}f](x) = \int_{\Omega} k(x, y) \mathcal{N}f(y) dy \geq \kappa \int_{\Omega} \mathcal{N}f(y) dy$, by assumption (2.2) on k . \square

PROOF OF THEOREM 2.2. We assume the minimal conditions (2.1)–(2.3) on k , as well as the extra condition (2.5). The first step is to write $L_n(f, \varphi) - L_\infty(f, \varphi)$ as

$$(5.1) \quad L_n(f, \varphi) - L_\infty(f, \varphi) = \int_{\Sigma} \psi(x) [dG_n(x) - dG(x)],$$

where $G_n(x)$ is the empirical distribution function, $G(x)$ is the df of the random variable X and

$$(5.2) \quad \psi(x) = \log \frac{[\mathcal{N}\varphi](x)}{[\mathcal{N}f](x)}.$$

Upon integration by parts (5.1) leads to

$$L_n(f, \varphi) - L_\infty(f, \varphi) = (-1)^d \int_{\Sigma} \mathcal{D}\psi(x) (G_n(x) - G(x)) dx,$$

where \mathcal{D} is defined in (2.5). The inequality

$$(5.3) \quad |L_n(f, \varphi) - L_\infty(f, \varphi)| \leq \|\mathcal{D}\psi\|_{L^1(\Sigma)} \|G_n(x) - G(x)\|_{L^\infty(\Sigma)}$$

follows. Together with assumption (2.5) and the fact that $\mathcal{N}\varphi$ and $\mathcal{N}f$ are bounded and bounded away from zero, it is now easy to show that

$$(5.4) \quad \begin{aligned} \|\mathcal{D}\psi\|_{L^1(\Sigma)} &\leq \text{const}(\|\mathcal{D}\mathcal{X}\{\mathcal{N}\varphi - \mathcal{N}f\}\|_{L^1(\Sigma)} + \|\mathcal{X}\{\mathcal{N}\varphi - \mathcal{N}f\}\|_{L^1(\Sigma)}) \\ &\leq \text{const} \cdot \|\mathcal{N}\varphi - \mathcal{N}f\|_{L^1(\Omega)}. \end{aligned}$$

The estimate (2.14) now follows from the well-known estimate

$$\|G_n - G\|_{L^2(\Sigma)} = \mathcal{O}\left(\left(\frac{\log \log n}{n}\right)^{1/2}\right) \quad \text{a.s.};$$

see Shorack and Wellner (1986).

To prove (2.15) we use (2.14). Since f^n minimizes $L_n(f)$, it follows that $L_n(f_S, f^n) \geq 0$. Hence

$$L_\infty(f^n) - L_\infty(f_S) = -L_\infty(f_S, f^n) \leq L_n(f_S, f^n) - L_\infty(f_S, f^n).$$

and from (2.14) we obtain the inequality

$$(5.5) \quad L_\infty(f^n) - L_\infty(f_S) \leq \text{const} \cdot \|\mathcal{N}f^n - \mathcal{N}f_S\|_{L^1(\Omega)} \cdot LL(n) \quad \text{a.s.}$$

We also need the following property of the smoothing operators (see Lemma 7.3):

$$(5.6) \quad \begin{aligned} \|\mathcal{N}f^n - \mathcal{N}f_S\|_{L^1(\Omega)} &\leq \|\mathcal{A}\{\mathcal{N}f^n - \mathcal{N}f_S\}\|_{L^1(\Sigma)} \\ &\quad + 2\|f^n - f_S\|_{L^1(\Omega)} + 2Mh^2, \end{aligned}$$

where M is finite when the minimal smoothness condition (2.16) on f^* holds. Since $g = \mathcal{K}f^*$ is bounded below and $\mathcal{K}\mathcal{N}f_S$ is bounded, uniformly in h (see Corollary 5.2), it follows that $r_S = g/\mathcal{K}\mathcal{N}f_S$ is bounded below and there is a constant such that

$$(5.7) \quad \|\mathcal{N}f^n - \mathcal{N}f_S\|_{L^1(\Omega)} \leq \text{const} \cdot (E + e + h^2),$$

where $e = \|f^n - f_S\|_{L^1(\Omega)}$, and

$$(5.8) \quad E = \int_{\Sigma} \sqrt{r_S} |\mathcal{A}\{\mathcal{N}(f^n) - \mathcal{N}(f_S)\}|.$$

Now estimating the right-hand side of (1.16) using (5.5)–(5.7), we derive the inequality

$$\frac{1}{2}E^2 + \frac{1}{8}h^2e^2 \leq \text{const} \cdot (E + e + h^2) \cdot LL(n).$$

Moving the E - and e -terms to the left-hand side and completing the squares gives, for suitable constants c ,

$$\begin{aligned} \frac{1}{2}(E - c \cdot LL(n))^2 + \frac{1}{8}(he - c \cdot h^{-1}LL(n))^2 \\ \leq c(h^{-2}(LL(n))^2 + h^2LL(n)). \end{aligned}$$

Now ignore the first term on the left-hand side. The result is inequality (2.15). \square

With the help of Theorem 2.2 one can give a short proof of Theorem 2.1. The basic idea is to approximate \mathcal{K} by an operator which satisfies the smoothness condition (2.5) and then apply Theorem 2.2, and show that the error due to the approximation is small. We omit the details.

6. Convergence of the regularization method. In this section Theorems 2.3 and 2.4 are proved.

PROOF OF THEOREM 2.3. We are interested in the convergence of f_S to the solution f^* of $\mathcal{K}f = g$ as $h \rightarrow 0$. For the moment let f^* be an arbitrary pdf with $\mathcal{K}f^* = g$ and $f^*, \log f^* \in C^3(\Omega)$. Our starting point is inequality (1.16). Let E and e denote the *unknown* quantities

$$(6.1) \quad E = \frac{1}{2} \int_{\Omega} \sqrt{\frac{g}{\mathcal{K}f_S}} |\mathcal{K}\{\mathcal{N}(f^*) - \mathcal{N}(f_S)\}|,$$

$$(6.2) \quad e = \|f_S - f^*\|_{L^1(\Omega)}.$$

Then inequality (1.16) may be written as

$$(6.3) \quad E^2 + \frac{1}{8}h^2e^2 \leq L_{\infty}(f^*) - L_{\infty}(f_S).$$

We need to rewrite this as follows. The right-hand side may be written as

$$(6.4) \quad \begin{aligned} \int_{\Sigma} g \log \frac{\mathcal{K}f_S}{\mathcal{K}f^*} &= \int_{\Sigma} (g - \mathcal{K}f^*) \log \frac{\mathcal{K}f_S}{\mathcal{K}f^*} + \int_{\Sigma} \mathcal{K}f_S - \mathcal{K}f^* \\ &\quad - \int_{\Sigma} \mathcal{K}(\mathcal{N}(f^*)) \log \frac{\mathcal{K}f^*}{\mathcal{K}f_S} + \mathcal{K}f_S - \mathcal{K}f^*. \end{aligned}$$

The expression on the second line is negative, since the integrand is nonpositive. Since \mathcal{K} maps pdfs into pdfs, the second term may be written as $h^2\{\Lambda_S(f) - \Lambda_S(f_S)\}$, with Λ_S given by (2.9). The elementary inequality $|\log t| \leq 2|t - 1|/\sqrt{t}$ now gives

$$(6.5) \quad \left| \log \frac{\mathcal{K}f_S}{\mathcal{K}f^*} \right| \leq 2 \frac{|\mathcal{K}f^* - \mathcal{K}f_S|}{\sqrt{\mathcal{K}f^*} \sqrt{\mathcal{K}f_S}},$$

so that

$$(6.6) \quad \int_{\Sigma} (g - \mathcal{K}f^*) \log \frac{\mathcal{K}f_S}{\mathcal{K}f^*} \leq 4E \left\| \frac{\mathcal{K}(f^* - \mathcal{N}f^*)}{\sqrt{\mathcal{K}f^*} \sqrt{\mathcal{K}f^*}} \right\|_{L^{\infty}(\Omega)},$$

with E as in (6.1). Since $k(x, y)$ is bounded away from 0, the denominator $\sqrt{\mathcal{K}f^*} \sqrt{\mathcal{K}f^*}$ is bounded away from 0, so from Lemma 7.3 it follows that the expression in (6.6) is bounded by $\text{const} \cdot h^2 \cdot E$. Inequality (6.3) may then be written as

$$(6.7) \quad E^2 + \frac{1}{8}h^2e^2 \leq \text{const} \cdot h^2E + h^2\{\Lambda_S(f^*) - \Lambda_S(f_S)\}.$$

Upon ignoring the e^2 term on the left-hand side and the $\Lambda_S(f_S)$ term on the right-hand side, we obtain

$$E^2 \leq \text{const} \cdot h^2E + \text{const} \cdot h^2,$$

which implies that $E \leq \text{const} \cdot h$. Then ignoring the left-hand side of (6.7) yields

$$(6.8) \quad \Lambda_S(f_S) \leq \Lambda_S(f^*) + \text{const} \cdot h \leq M < \infty,$$

where the last bound follows from assumptions (2.10) and (2.19). In Lemma 7.5 we show that (6.8) implies the asymptotic compactness of $\mathcal{N}f_S$ for $h \rightarrow 0$,

that is, there exists a sequence $\{h_n\}_n$ tending to 0, a corresponding sequence of smoothers $\{S_n\}_n$ and an $f_0 \in L^1(\Omega)$ such that for $\varphi_n = \mathcal{N}_{S_n}(f_{S_n})$ we have $\|\varphi_n - f_0\|_{L^1(\Omega)} \rightarrow 0$, and

$$(6.9) \quad \begin{aligned} \Lambda_{\mathcal{S}}(f_0) &\leq \liminf_{h \rightarrow 0} \Lambda_S(f_S) \leq \limsup_{h \rightarrow 0} \Lambda_S(f_S) \\ &\leq \lim_{h \rightarrow 0} \Lambda_S(f^*) = \Lambda_{\mathcal{S}}(f^*). \end{aligned}$$

Since $\mathcal{K}(f_{S_n} - \varphi_n) \rightarrow 0$ in $C(\Sigma)$, [see (7.11)], and $E \rightarrow 0$ we have $\mathcal{K}f_0 = g$. By uniqueness $f_0 = f^*$. So in (6.9) we have equality everywhere. From (6.7) and the aforementioned estimate $E \leq c \cdot h$, then

$$\|f^* - f_S\|_{L^1(\Omega)}^2 \leq \text{const} \cdot h + \Lambda_S(f^*) - \Lambda_S(f_S),$$

and so, by (6.9),

$$(6.10) \quad \lim_{h \rightarrow 0} \|f^* - f_S\|_{L^1(\Omega)}^2 \leq \lim_{h \rightarrow 0} \Lambda_S(f^*) - \Lambda_S(f_S) = 0.$$

We have thus proven the convergence of the regularization method. \square

PROOF OF THEOREM 2.4. It is obvious that in order to get rates we need to be more specific about the term $\Lambda_S(f^*) - \Lambda_S(f_S)$. Since f^* satisfies conditions (2.17) and (2.18), the Gateaux derivatives $\Lambda'_S(f^*)$ and $\Lambda'_{\mathcal{S}}(f^*)$ are well behaved [see (7.4)] and so, by convexity of Λ_S ,

$$\begin{aligned} \Lambda_S(f^*) - \Lambda_S(f_S) &\leq \int_{\Omega} [\Lambda'_S(f^*)](y)(f^*(y) - f_S(y)) dy \\ &\leq \int_{\Omega} [\Lambda'_{\mathcal{S}}(f^*)](y)(f^*(y) - f_S(y)) dy \\ &\quad + \text{const} \cdot h \cdot \|f^* - f_S\|_{L^1(\Omega)}. \end{aligned}$$

Now applying assumption (2.18) results in the estimate

$$\begin{aligned} &\int_{\Omega} [\Lambda'_{\mathcal{S}}(f^*)](y)(f^*(y) - f_S(y)) dy \\ &= \int_{\Omega} [\mathcal{K}^* \lambda_0](y)(f^*(y) - f_S(y)) dy \\ &= \int_{\Sigma} \lambda_0(x)(\mathcal{K}f^*(x) - \mathcal{K}f_S(x)) dx \\ &\leq \|\lambda_0\|_{L^\infty(\Sigma)} \|\mathcal{K}(f^* - f_S)\|_{L^1(\Sigma)}. \end{aligned}$$

Combining the above with inequality (7.13) we obtain that

$$\Lambda_S(f^*) - \Lambda_S(f_S) \leq \text{const} \cdot (E + h^2 + he).$$

Substituting this into inequality (6.7) results in

$$E^2 + \frac{1}{8}h^2e^2 \leq \text{const} \cdot h^2 \cdot (E + h^2 + he).$$

It follows that $e \leq \text{const} \cdot h$, by the familiar reasoning. \square

Assumptions (2.17) and (2.18) appear to be strange, but under mild extra conditions if the rates of Theorem 2.4 hold, then (2.17) and (2.18) must also hold. Compare the similar situation in maximum entropy regularization [Eggermont (1993)].

7. On the smoothing operator. In this section we review the properties of the smoothing operators S and \mathcal{N} . The first lemma is well known and deals with the arithmetic–geometric mean inequality. The remaining results are proved for the smoothing kernels (2.6) and (2.7), but hold in much greater generality, as the proofs make clear.

LEMMA 7.1. *For all nonnegative $f \in L^1(\Omega)$,*

$$[\mathcal{N}f](x) \leq ([S\{f^{1/2}\}](x))^2 \leq [Sf](x), \quad \text{a.e. } x \in \Omega,$$

$$[Sf](x) - 2([Sf^{1/2}](x))^2 + [\mathcal{N}f](x) \geq 0, \quad \text{a.e. } x \in \Omega.$$

PROOF. Fix a nonnegative $f \in L^1(\Omega)$ and $x \in \Omega$. The function $\lambda(r) = ([Sf^r](x))^{1/r}$, for $r > 0$, and $\lambda(0) = [\mathcal{N}f](x)$ is increasing and log-convex [cf. Hardy, Littlewood and Pólya (1952), Section 6.12]. In particular, $\lambda(0) \leq \lambda(\frac{1}{2}) \leq \lambda(1)$. This is the content of the first statement. The fact that $\lambda(r)$ is log-convex and increasing implies that $\lambda(r)$ is convex, and so $\lambda(1) - 2\lambda(\frac{1}{2}) + \lambda(0) \geq 0$. This is the second statement. \square

We now restrict attention to the case where Ω is the unit cube in \mathbb{R}^d , and the smoothing kernels (2.6) and (2.7).

LEMMA 7.2. (*Approximation properties*).

$$(7.1) \quad \|f - Sf\|_{L^\infty(\Omega)} = \mathcal{O}(h), \quad f \in C^2(\Omega),$$

$$(7.2) \quad \|f - Sf\|_{L^1(\Omega)} = \mathcal{O}(h^2), \quad f \in C^3(\Omega),$$

$$(7.3) \quad \Lambda_{\mathcal{S}}(f) = \lim_{h \rightarrow 0} \Lambda_S(f) = \frac{2}{3} \int_{\Omega} |\nabla \sqrt{f}|^2 + \int_{\Omega} (f - u) \log(f/u),$$

$$(7.4) \quad \Lambda_{\mathcal{S}}(f) - \Lambda_S(f) = \mathcal{O}(h).$$

In (7.3), u is the uniform distribution on Ω . In (7.3) and (7.4), we need $f, \log f \in C^3(\Omega)$, and that $\partial_n f = 0$ on $\partial\Omega$.

PROOF. It suffices to consider the one-dimensional case $\Omega = [0, 1]$. Let T be the smoothing operator with smoothing kernel $t_h(x, y)$. It is easily verified that, for $f \in C^3(\Omega)$,

$$(7.5) \quad \begin{aligned} & Tf(x) - f(x) \\ &= \begin{cases} \frac{1}{6}h^2 f''(x) + \mathcal{O}(h^3), & h < x < 1 - h, \\ \frac{1}{2}h(1 - x/h)^2 f'(x) + \mathcal{O}(h^2), & 0 < x < h, \\ -\frac{1}{2}h(1 - (1 - x)/h)^2 f'(x) + \mathcal{O}(h^2), & 1 - h < x < 1. \end{cases} \end{aligned}$$

This shows that $Tf - f = \mathcal{O}(h)$, uniformly on Ω . This is (7.1). Integrating (7.5) yields (7.2), since $Tf - f = \mathcal{O}(h^2)$, except on intervals of length h , where it is $\mathcal{O}(h)$. The modifications for the operator S are trivial.

To prove (7.3) and (7.4), we write

$$(7.6) \quad \begin{aligned} f - \mathcal{N}_T f &= f(1 - \exp[T\{\log f\} - \log f]) \\ &= -f(T\{\log f\} - \log f) + \mathcal{O}(f(T\{\log f\} - \log f)^2), \end{aligned}$$

uniformly on $[0, 1]$, where we used (7.1). Note that upon integration

$$(7.7) \quad \int_{\Omega} f(T\{\log f\} - \log f)^2 \leq \text{const} \cdot h^3.$$

Applying (7.5) to $Tf - f$ as well as to $T\{\log f\} - \log f$, we see that

$$(7.8) \quad Tf - f - f(T\{\log f\} - \log f) = \frac{1}{6}h^2 f'' - \frac{1}{6}h^2 f(\log f)'' + \mathcal{O}(h^3),$$

for $h < x < 1 - h$, and $= \mathcal{O}(h^2)$ for $0 < x < h$ and $1 - h < x < 1$. We now have

$$\Lambda_T(f) = h^{-2} \int_0^1 Tf - f - f(T\{\log f\} - \log f) + \mathcal{O}(h),$$

where we used (7.6) and (7.7). It follows now from (7.8) that

$$(7.9) \quad \Lambda_T(f) = \frac{1}{6} \int_h^{1-h} f'' - f(\log f)'' + \mathcal{O}(h) = \frac{2}{3} \int_0^1 \left| \{\sqrt{f}\}' \right|^2 + \mathcal{O}(h).$$

This is the first part of (7.3). The modification necessary for the operator S are fairly straightforward and are omitted.

To show (7.4), we first note that the boundary conditions $\partial_n f = 0$ on $\partial\Omega$ reduce to $f'(x) = 0$ for $x = 0$ and $x = 1$. Then (7.5) can be extended to

$$(7.10) \quad Tf(x) - f(x) = \frac{1}{6}h^2 f''(x) + \mathcal{O}(h^3) \quad \text{uniformly on } [0, 1].$$

It is fairly obvious that $\Lambda_T(f)$ has a linear Gateaux derivative, which is given by

$$\Lambda_T(f) = 1 - \frac{T\mathcal{N}_T f}{f} = \frac{f - T^2 f}{f} + \frac{T\{Tf - \mathcal{N}_T f\}}{f}.$$

Note that (7.6) together with (7.10) implies that

$$h^{-2}(Tf - \mathcal{N}_T f) = \frac{1}{6}(f'' - f(\log f)'') + \mathcal{O}(h),$$

provided $\log f \in C^3(\Omega)$. It follows that $h^{-2}T(Tf - \mathcal{N}_T f)$ satisfies the same estimate. Likewise, since $f - T^2 f = (\mathcal{J} + T)(f - Tf)$, we get

$$h^{-2}(f - T^2 f) = -\frac{1}{3}f'' + \mathcal{O}(h).$$

Putting the above together yields

$$\begin{aligned} \Lambda_T(f) &= -\frac{1}{3} \frac{f''}{f} + \frac{1}{6} \left(\frac{f''}{f} - (\log f)'' \right) + \mathcal{O}(h) \\ &= \frac{1}{6} \left| \frac{f'}{f} \right|^2 - \frac{1}{3} \frac{f''}{f} + \mathcal{O}(h). \end{aligned}$$

The right-hand side, minus the $\mathcal{O}(h)$ term, is the Gateaux derivative of $\Lambda_{\mathcal{J}}(f)$ (for $f' = 0$ at 0 and 1). This is (7.4). The modifications necessary for the operator S are omitted. \square

LEMMA 7.3. *Suppose that f^* satisfies the minimal smoothness assumption (2.16) and that \mathcal{K} satisfies (2.4). Let f_S denote the solution of (1.15). Then*

$$(7.11) \quad \|\mathcal{K}(f^* - \mathcal{N}f^*)\|_{L^\infty(\Sigma)} \leq c \cdot h^2, \quad \|\mathcal{K}(f_S - \mathcal{N}f_S)\|_{L^\infty(\Sigma)} \leq c \cdot h^2.$$

$$(7.12) \quad \|\mathcal{N}f - \mathcal{N}f_S\|_{L^1(\Omega)} \leq \|\mathcal{K}(\mathcal{N}f - \mathcal{N}f_S)\|_{L^1(\Sigma)} + 2\|f - f_S\|_{L^1(\Omega)} + 2Mh^2,$$

$$(7.13) \quad \|\mathcal{K}(f_S - f^*)\|_{L^1(\Sigma)} \leq \|\mathcal{K}(\mathcal{N}f_S - \mathcal{N}f^*)\|_{L^1(\Sigma)} + c(h^2 + h\|f_S - f^*\|_{L^1(\Omega)}),$$

where $M = \sup_h \Lambda_T(f_S) < \infty$. Inequality (7.12) holds for all nonnegative $f \in L^1(\Omega)$.

PROOF OF (7.11). We recall that $\mathcal{K}: L^1(\Omega) \rightarrow C(\Sigma)$ is bounded, so its norm $\|\mathcal{K}\|_{1,\infty}$ is finite. We write $\mathcal{K}(f^* - \mathcal{N}f^*) = \mathcal{K}(f^* - Sf^*) + \mathcal{K}(Sf^* - \mathcal{N}f^*)$ and note that

$$\|\mathcal{K}(\mathcal{J} - S)f^*\|_{C(\Sigma)} \leq \|\mathcal{K}\|_{1,\infty} \|(\mathcal{J} - S)f^*\|_{L^1(\Omega)} \leq c \cdot h^2,$$

since $f^* \in C^3(\Omega)$ [see assumption (2.20), and (7.2)]. We also note that $Sf^* - \mathcal{N}f^* \geq 0$, so

$$\begin{aligned} \|\mathcal{K}(Sf^* - \mathcal{N}f^*)\|_{C(\Sigma)} &\leq \|\mathcal{K}\|_{1,\infty} \|Sf^* - \mathcal{N}f^*\|_{L^1(\Omega)} \\ &\leq c \cdot h^2 \cdot \Lambda_S(f^*) \leq c \cdot h^2. \end{aligned}$$

The estimate (7.11) for f^* follows. The proof for f_S is similar.

PROOF OF (7.12). Let $\Omega^+ = \{y \in \Omega: \mathcal{N}f(y) \geq \mathcal{N}f_S(y)\}$ and $\Omega^- = \Omega \setminus \Omega^+$. Then

$$0 \leq \int_{\Omega^+} \{\mathcal{N}f - \mathcal{N}f_S\} = \int_{\Omega^+} \{\mathcal{N}f - Sf\} + \int_{\Omega^+} S(f - f_S) + \int_{\Omega^+} \{Sf_S - \mathcal{N}f_S\}.$$

The first integral is negative since the integrand is negative by Lemma 7.1. For the same reason the integrand of the last integral is positive. It follows that

$$\begin{aligned} (7.14) \quad 0 &\leq \int_{\Omega^+} \{\mathcal{N}f - \mathcal{N}f_S\} \\ &\leq \|S(f - f_S)\|_{L^1(\Omega)} + h^2 \Lambda_S(f_S) \\ &\leq \|f - f_S\|_{L^1(\Omega)} + Mh^2. \end{aligned}$$

Likewise,

$$\begin{aligned} (7.15) \quad 0 &\leq \int_{\Omega^-} \{\mathcal{N}f_S - \mathcal{N}f\} \\ &= \int_{\Omega^-} \{\mathcal{N}f_S - \mathcal{N}f\} + \int_{\Omega^+} \{\mathcal{N}f - \mathcal{N}f_S\} \\ &= \int_{\Sigma} \mathcal{K}\{\mathcal{N}f_S - \mathcal{N}f\} + \int_{\Omega^+} \{\mathcal{N}f - \mathcal{N}f_S\} \\ &\leq \|\mathcal{K}\{\mathcal{N}f_S - \mathcal{N}f\}\|_{L^1(\Sigma)} + \|f - f_S\|_{L^1(\Omega)} + Mh^2, \end{aligned}$$

where in the last step we used (7.14). The inequality now follows from (7.14) and (7.15).

The proof of (7.13) follows similarly. \square

In the following we let $H^1(\Omega)$, resp. $H^2(\Omega)$, denote the Sobolev space of all square integrable functions on Ω with square integrable first-order, resp. second-order, derivatives. We note that closed, bounded subsets of $H^1(\Omega)$, resp. $H^2(\Omega)$, are compact in $L^2(\Omega)$ [see Adams (1975)].

LEMMA 7.4. *Let \mathcal{L} be the operator with domain*

$$(7.16) \quad \mathcal{D}(\mathcal{L}) = \{f \in H^2(\Omega) : \partial_n f = 0 \text{ on } \partial\Omega\}$$

and defined by

$$(7.17) \quad \mathcal{L}f = -\Delta f, \quad f \in \mathcal{D}(\mathcal{L}),$$

where $\Delta = \partial_1^2 + \dots + \partial_d^2$ is the Laplacian. Then the operators $h^{-2}(\mathcal{L} - T^2) - T\mathcal{L}T$ are semi-positive-definite.

PROOF. We prove this for the one-dimensional case $\Omega = [0, 1]$. The extension to cubes in \mathbb{R}^d is straightforward.

We note that the eigenfunctions of T are known. If we let $u_n(x) = \cos(\pi nx)$, then $Tu_n = \lambda_n u_n$, where $\lambda_n = \sin(\pi nh)/(\pi nh)$. This is most elegantly seen as follows. Let $f \in L^2(0, 1)$. Extend f to an even function f_e on $[-1, 1]$ and then to a periodic function on the line with period 2. Then we may write

$$(7.18) \quad Tf(x) = T_e f_e(x) =_{\text{def}} \frac{1}{2h} \int_{x-h}^{x+h} f_e(y) dy, \quad x \in [0, 1],$$

which is the (2-periodic) convolution of f_e with a box function. It follows that the eigenfunctions of T_e are precisely $\cos(\pi nx)$ and $\sin(\pi nx)$. The even eigenfunctions are inherited by T , whence the result follows. The eigenvalues are then easily calculated from (7.18). Note that $\cos(\pi nx)$, $n = 0, 1, \dots$, is a basis for $L^2(0, 1)$, so we do indeed have all the eigenfunctions.

Now consider the linear operator \mathcal{M} defined by $\mathcal{M}u_n = \mu_n u_n$, for $n \geq 0$, with μ_n still to be determined. The eigenfunctions of the operator $\mathcal{Q} = h^{-2}(\mathcal{L} - T^2) - T\mathcal{M}T$ are again the u_n , and the eigenvalues are $\nu_n = h^{-2}(1 - \lambda_n^2) - \lambda_n^2 \mu_n$, which we may write as

$$\nu_n = \lambda_n^2 \left[(\pi n)^2 \frac{(\pi nh)^2 - (\sin(\pi nh))^2}{(\pi nh)^2 \sin^2(\pi nh)} - \mu_n \right].$$

It is easily verified that

$$\frac{x^2 - (\sin x)^2}{(x \sin x)^2} \geq \frac{1}{3} \quad \text{for all } x \in \mathbb{R},$$

and so $\nu_n \geq \lambda_n^2(\frac{1}{3}(\pi n)^2 - \mu_n)$. Thus, taking $\mu_n = \frac{1}{3}(\pi n)^2$ makes $\nu_n \geq 0$ and makes \mathcal{Q} semi-positive-definite. With this choice \mathcal{M} is defined by $\mathcal{M}f = -\frac{1}{3}f''$, and $\langle \varphi, \mathcal{M}\psi \rangle = \frac{1}{3} \int_{\Omega} \nabla \varphi \cdot \nabla \psi$. Identifying \mathcal{L} with \mathcal{M} proves the lemma. \square

LEMMA 7.5. *Let $\{f_h: h > 0\}$ be a bounded subset of $L^1(\Omega)$ satisfying for all $h > 0$*

$$\Lambda_S(f_h) \leq M \quad \text{and} \quad f_h \geq h^2.$$

Then $\{\mathcal{N}(f_h): h > 0\}$ is asymptotically compact, that is, every sequence $\{h_n\}_n$ converging to 0, has a subsequence, also denoted by $\{h_n\}_n$, for which $\{\mathcal{N}_{h_n}(f_{h_n})\}_n$ converges strongly in $L^1(\Omega)$, to some element φ_0 . Moreover,

$$\Lambda_{\mathcal{S}}(\varphi_0) \leq \liminf_{h \rightarrow 0} \Lambda_S(f_h).$$

Here \mathcal{N}_{h_n} is the nonlinear smoothing operator corresponding to S_{h_n} .

PROOF. We first prove it for the operators T . We denote the nonlinear smoothing operator associated with T by \mathcal{N}_T . Lemma 7.1 says that $T\varphi_h - \mathcal{N}_T(\varphi_h) \geq 2(T\varphi_h - \{T\varphi_h^{1/2}\}^2)$. Then

$$\begin{aligned} \Lambda_T(\varphi_h) &\geq 2h^{-2} \int_{\Omega} T\varphi_h - \{T\varphi_h^{1/2}\}^2 \\ (7.19) \quad &= 2h^{-2} \int_{\Omega} \varphi_h^{1/2} [(\mathcal{S} - T^2)\varphi_h^{1/2}] \\ &\geq 2 \int_{\Omega} [T\varphi_h^{1/2}] [\mathcal{S}T\varphi_h^{1/2}] = \frac{2}{3} \|\nabla T\varphi_h^{1/2}\|_{L^2(\Omega)}^2, \end{aligned}$$

where ∇ denotes the gradient. Since $\Lambda_T(\varphi_h) \leq M$, it follows that

$$\frac{2}{3} \|\nabla T\varphi_h^{1/2}\|_{L^2(\Omega)}^2 \leq M,$$

and hence the $T\varphi_h^{1/2}$ lie in a closed and bounded, hence weakly compact, subset of $H^1(\Omega)$. This implies that there is a weakly convergent [in $H^1(\Omega)$] sequence of $T\varphi_h^{1/2}$ with weak limit $\varphi_0^{1/2} \in H^1(\Omega)$. It follows that this subsequence of $T\varphi_h^{1/2}$ converges strongly in $L^2(\Omega)$, and hence $(T\varphi_h^{1/2})^2$ converges strongly in $L^1(\Omega)$ with limit φ_0 . Since $\|(T\varphi_h^{1/2})^2 - \mathcal{N}_T(\varphi_h)\|_{L^1(\Omega)} \leq h^2 \Lambda_T(\varphi_h)$ (cf. Lemma 7.1), the strong convergence of $\mathcal{N}_T(\varphi_h)$ follows as well. Now the weak convergence in $H^1(\Omega)$ of $T\varphi_h^{1/2}$ and the weak lower semicontinuity of the norm yield that

$$\frac{2}{3} \|\nabla \varphi_0^{1/2}\|_{L^2(\Omega)}^2 \leq \frac{2}{3} \liminf_{h \rightarrow 0} \|\nabla T\varphi_h^{1/2}\|_{L^2(\Omega)}^2 \leq \liminf_{h \rightarrow 0} \Lambda_T(\varphi_h).$$

The left-hand side is precisely $\Lambda_{\mathcal{S}}(\varphi_0)$. This concludes the case for the operator T .

Now consider the smoothing operators S . We let \mathcal{E} denote the operator defined as

$$\mathcal{E}f(y) = \frac{1}{|\Omega|} \int_{\Omega} f, \quad y \in \Omega,$$

so that in operator notation $S = (1 - h^2)T + h^2\mathcal{E}$. We let \mathcal{N} denote the nonlinear smoothing operator associated with S . As before,

$$\Lambda_S(\varphi_h) = \Lambda_T(\varphi_h) + h^{-2} \int_{\Omega} \mathcal{N}_T \varphi_h - \mathcal{N} \varphi_h.$$

We abbreviate $\mathcal{N}\varphi_h$ as Φ_S and write

$$\begin{aligned}
 h^{-2} \int_{\Omega} \mathcal{N}_T \varphi_h - \mathcal{N}\varphi_h &= h^{-2} \int_{\Omega} \Phi_S(\exp[(T - S)\log \varphi_h] - 1) \\
 &\geq h^{-2} \int_{\Omega} \Phi_S\{(T - S)\log \varphi_h\} \\
 &= \int_{\Omega} \Phi_S(T\{\log \varphi_h\} - \mathcal{E}\{\log \varphi_h\}) \\
 &= (1 - h^2)^{-1} \int_{\Omega} \Phi_S\{(1 - h^2)T\{\log \varphi_h\} \\
 &\quad + h^2\mathcal{E}\{\log \varphi_h\} - \mathcal{E}\log \varphi_h\} \\
 &= (1 - h^2)^{-1} \int_{\Omega} \Phi_S[(\mathcal{I} - \mathcal{E})S\{\log \varphi_h\}] \\
 &= (1 - h^2)^{-1} \int_{\Omega} (\Phi_S - \mathcal{E}\Phi_S)S\{\log \varphi_h\} \\
 &= (1 - h^2)^{-1} \int_{\Omega} (\Phi_S - 1)S\{\log \varphi_h\} \\
 &\quad + h^2(1 - h^2)^{-1} \Lambda_S(\varphi_h)\mathcal{E}S\{\log \varphi_h\} \\
 &\geq (1 - h^2)^{-1} \int_{\Omega} (\mathcal{N}(\varphi_h) - 1)\log \mathcal{N}\varphi_h \\
 &\quad + h^2(1 - h^2)^{-1} \Lambda_S(\varphi_h)\log h^2,
 \end{aligned}$$

where in the last step we used that $\varphi_h \geq h^2$. To summarize, we have shown that

$$\begin{aligned}
 \Lambda_S(\varphi_h) &\geq \Lambda_T(\varphi_h) + \frac{1}{1 - h^2} \int_{\Omega} (\mathcal{N}(\varphi_h) - 1)\log \mathcal{N}\varphi_h \\
 (7.20) \quad &+ \frac{h^2}{1 - h^2} \Lambda_S(\varphi_h)\log h^2.
 \end{aligned}$$

Since $(t - 1)\log t \geq 0$ and $h^2 \log h^2 \rightarrow 0$, the boundedness of $\Lambda_S(\varphi_h)$ implies that $\Lambda_T(\varphi_h)$ is bounded as $h \rightarrow 0$, and so by the first part of the proof, a subsequence of $\mathcal{N}_T(\varphi_h)$ converges in $L^1(\Omega)$, say, to some f_0 . Since

$$\begin{aligned}
 \|\mathcal{N}_T(\varphi_h) - \mathcal{N}(\varphi_h)\|_{L^1(\Omega)} &\leq h^2(\Lambda_T(\varphi_h) + \Lambda_S(\varphi_h)) + \|(T - S)\varphi_h\|_{L^1(\Omega)} \\
 &\leq \text{const} \cdot h^2,
 \end{aligned}$$

$\mathcal{N}\varphi_h$ also converges weakly in $L^1(\Omega)$ to f_0 . Then it follows from (7.19) that $\lim_h \Lambda_T(f_0) \leq \liminf_h \Lambda_T(\varphi_h)$, and by Fatou's lemma that

$$\int_{\Omega} (f_0 - 1)\log f_0 \leq \liminf_{h \rightarrow 0} (1 - h^2)^{-1} \int_{\Omega} (\mathcal{N}\varphi_h - 1)\log \mathcal{N}\varphi_h.$$

Consequently,

$$\Lambda_{\mathcal{F}}(f_o) = \liminf_{h \rightarrow 0} \Lambda_S(f_o) \leq \liminf_{h \rightarrow 0} \Lambda_S(\varphi_h).$$

This concludes the treatment of the operators S . \square

REFERENCES

- ADAMS, R. A. (1975). *Sobolev Spaces*. Academic Press, New York.
- CARROLL, R. J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.
- COX, D. D. and O'SULLIVAN, F. (1990). Asymptotic analysis of penalized likelihood and related estimators. *Ann. Statist.* **18** 1676–1695.
- CSISZÁR, I. and TUSNÁDY, G. (1984). Information geometry and alternating minimization procedures. *Statist. Decisions Suppl.* **1** 205–237.
- EGGERMONT, P. P. B. (1992). Nonlinear smoothing and the EM algorithm for positive integral equations of the first kind. Unpublished manuscript.
- EGGERMONT, P. P. B. (1993). Maximum entropy regularization for Fredholm integral equations of the first kind. *SIAM J. Math. Anal.* **24** 1557–1576.
- EGGERMONT, P. P. B. and LARICCIA, V. N. (1995). Maximum smoothed likelihood density estimation. *Journal of Nonparametric Statistics*. To appear.
- GOOD, I. J. and GASKINS, R. A. (1971). Nonparametric roughness penalties for probability densities. *Biometrika* **58** 255–277.
- GREEN, P. J. (1990). On the use of the EM algorithm for penalized likelihood estimation. *J. Roy. Statist. Soc. Ser. B* **52** 443–452.
- HARDY, G. H., LITTLEWOOD, J. E. and PÓLYA, G. (1952). *Inequalities*. Cambridge Univ. Press.
- KAY, J. (1992). On the convergence of the EMS algorithm. Unpublished manuscript.
- KEMPERMAN, J. H. B. (1967). *On the Optimum Rate of Transmitting Information. Lecture Notes in Math.* **23** 126–169. Springer, New York.
- LATHAM, G. A. and ANDERSSON, R. S. (1993). On the stabilization inherent in the EMS algorithm. Research Report CMA-MR07-93, Australian National Univ.
- MENDELSON, J. and RICE, J. (1982). Deconvolution of microfluorometric histograms with B -splines. *J. Amer. Statist. Assoc.* **77** 748–753.
- SHORACK, G. R. and WELLNER, J. A. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- SILVERMAN, B. W. (1982). On the estimation of a probability density function by the maximum penalized likelihood method. *Ann. Statist.* **10** 795–810.
- SILVERMAN, B. W., JONES, M. C., WILSON, J. D. and NYCHKA, D. W. (1990). A smoothed EM algorithm approach to indirect estimation problems, with particular reference to stereology and emission tomography. *J. Roy. Statist. Soc. Ser. B* **52** 271–324.
- VARDI, Y. and LEE, D. (1993). From image deblurring to optimal investments: maximum likelihood solutions for positive linear inverse problems (with discussion). *J. Roy. Statist. Soc. Ser. B* **55** 569–612.
- VARDI, Y., SHEPP, L. A. and KAUFMAN, L. (1985). A statistical model for positron emission tomography. *J. Amer. Statist. Assoc.* **80** 8–38.
- WILSON, J. D. (1989). A smoothed EM algorithm for the solution of Wicksell's corpuscle problem. *J. Statist. Comput. Simul.* **31** 195–221.

DEPARTMENT OF MATHEMATICAL SCIENCES
UNIVERSITY OF DELAWARE
NEWARK, DELAWARE 19716