CONVERGENCE RATES FOR REGULARIZED SOLUTIONS OF INTEGRAL EQUATIONS FROM DISCRETE NOISY DATA

Running head: Regularization of Integral Equations.

DOUGLAS W. NYCHKA

Department of Statistics, North Carolina State University

and

DENNIS D. COX[1]

Department of Statistics, University of Wisconsin, Madison

[1]This author's research was supported by the National Science Foundation under Grant No. MCS-820-2560.

AMS 1980 subject classifications. Primary, 62-G05, Secondary, 62J05, 41-A35, 41-A25, 47-A53, 45-L10, 45-M05.

Key words and phrases. Method of regularization, approximate solution of integral equations, rates of convergence, nonparametric regression, Abel's equation, particle size distribution.

SUMMARY

Given data $y_i = (Kg)(u_i) + \varepsilon_i$ where the ε's are random errors, the u's are known, g is an unknown function in a reproducing kernel space with kernel r , and K is a known integral kernel operator, it is shown how to calculate convergence rates for the regularized solution of the equation as the evaluation points $\{u_i\}$ become dense in the interval of interest. These depend on the eigenvalue asymptotics of KRK^* , where R is the integral operator with kernel r . The theory is applied to periodic deconvolution and Abel's equation. A particular example of the latter is the estimation of particle size densities in stereology. Rates of convergence of regularized histogram estimates of the particle size density are given.

1. Introduction.

Integral equations often provide a crucial link between observations on a system and a function g which characterizes the state of the system. We consider an observational model of the form

$$(1.1) \qquad y_{in} = \int_0^1 k(u_{in},v)g(v)dv + \varepsilon_{in} \quad , \qquad i=1,2,\ldots,n.$$

Here, the kernel function $k : [0,1]\times[0,1] \to \mathbb{R}$ is assumed known, the u's are known points in [0,1], the ε's are random errors, and $g : [0,1] \to \mathbb{R}$ is an unknown function. g is thought to lie in a Hilbert space $\mathcal{X}$, which is a reproducing kernel Hilbert space (RKHS) (i.e., the evaluation functional $g \mapsto g(v)$ is continuous on $\mathcal{X}$ for each $v\in[0,1]$) with reproducing kernel $r : [0,1] \times [0,1] \to \mathbb{R}$. See Aronsajn (1950) for RKHS theory. Given the observation vector

$$\underset{\sim}{y}_n = (y_{i1},\ldots,y_{in})' \quad ,$$

the statistical problem is to obtain an estimate of g . If $\mathcal{X}$ is finite dimensional, then this can be treated by standard parametric regression techniques. However, there is frequently no sound basis for assuming a parametric form for g , in which case it is appropriate to apply nonparametric methods, i.e., use an infinite dimensional parameter space $\mathcal{X}$. This makes the estimation problem somewhat more difficult.

The type of estimator which is analyzed here is the so-called method of regularization (MOR) estimator, which was first proposed in the integral equation context by Tikhonov (1963) (see also Nashed and Wahba, 1974). Statistical justifications for such estimators have been given by Li (1982) (see also Speckman, 1979), and the arguments in Kimeldorf and Wahba (1970) can be adapted to show that linear Bayes estimators are MOR estimators. The estimator is obtained by minimization over $h \in \mathcal{X}$ of

$$L_{n\lambda}(h) = \lambda \langle h, Ph \rangle_{\mathcal{X}} + \frac{1}{n} \sum_{i=1}^{n} (y_{in} - h(u_{in}))^2 \qquad , \tag{1.2}$$

where $P \in \mathcal{L}(\mathcal{X},\mathcal{X})$ = the collection of all bounded linear operators $\mathcal{X} \to \mathcal{X}$, and $\lambda \in [0,\infty]$ is the smoothing parameter. It is assumed the quadratic form $h \mapsto \langle h, Ph \rangle_{\mathcal{X}}$ is positive semidefinite ($\langle \cdot , \cdot \rangle_{\mathcal{X}}$ denotes the inner product on $\mathcal{X}$), and that $\langle h, Ph \rangle_{\mathcal{X}} = 0$ and $h(u_{in}) = 0$, $1 \le i \le n$, imply $h \equiv 0$, which guarantees that the resulting estimate (denoted $\hat{g}_{n\lambda}$) is uniquely defined. Determination of λ is a problem of some difficulty about which we will only make a few remarks. MOR estimators are widely used and appear to work well in practice. Some pertinent references are Wahba (1977, 1980) and Lukas (1980, 1981).

One example that motivated this research is derived from the kernel

$$k(u,v) = I_{[u,1]}(v)(v-u)^{-\frac{1}{2}} \qquad , \tag{1.3}$$

where I_A denotes the indicator function of A. This gives the weakly singular Volterra equation known as Abel's equation

$$\eta(u) = \int_u^1 \frac{g(v)}{\sqrt{v-u}}\, dv \qquad .$$

This particular kernel has a diverse range of applications in the physical sciences (Kosarev, 1980; Bullen, 1963), and in stereological microscopy (Anderssen and Jakeman, 1975a,b; Anderssen, 1976; Nychka, 1983). In this latter field, one is interested in estimating the probability distribution (or its density) of the radii of spherical bodies embedded in a medium using observations on cross sections in a planar slice of the medium. See Watson (1971) and Franklin (1981) for a more traditional statistical approach to this problem. We propose and analyze here a MOR estimator based on a histogram (see Section 5).

The main interest here is to study asymptotic properties of $\hat{g}_{n\lambda}$ as $n \to \infty$ and the u_{in} become dense in [0,1]. There will also be some restrictions on λ (that it is not too small). In particular, we will show that for appropriate sequences of λ values, $\hat{g}_{n\lambda}$ is a consistent estimator, and we can obtain upper bounds on the rates of convergence. These results are useful for indicating how the features of the problem (i.e., choice of λ , $\mathcal{X}$, P , the true g , and the design points $\{u_{in}\}$) affect the estimation error.

The following asymptotic notations will be used.

$$\alpha(n) \asymp \beta(n) \qquad \text{as} \quad n\to\infty$$

means $\qquad A\beta(n) < |\alpha(n)| < B\beta(n)$ for all n sufficiently large, where A,B are positive constants. We will also need

$$\alpha(n,\xi) \asymp \beta(n,\xi) \quad \text{as} \quad n\to\infty \text{ uniformly in } \xi\in\Xi_n$$

if

$$A\beta(n,\xi) \le |\alpha(n,\xi)| \le B\beta(n,\xi)$$

for all $\xi\in\Xi_n$, for all n sufficiently large.

We also use the one-sided relation

$$\alpha(n) \lesssim \beta(n) \qquad \text{if} \qquad |\alpha(n)| \le B\beta(n)$$

for all n sufficiently large, and asymptotic negligibility

$$\alpha(n) << \beta(n) \qquad \text{if} \qquad \lim_{n\to\infty} \frac{\alpha(n)}{\beta(n)} = 0 \qquad .$$

Certain function spaces are now defined. The Sobolev spaces on [0,1] are

$$W_2^k = \{h:\ h, h^{(1)}, \ldots, h^{(k-1)} \text{ are absolutely continuous, and } \int_0^1 (h^{(k)}(u))^2 du < \infty\} \quad ,$$

and their norms are given by

$$\|h\|^2_{W_2^k} = \|h\|^2_{L_{2[0,1]}} + \|h^{(k)}\|^2_{L_{2[0,1]}} \quad .$$

We also need the Sobolev spaces with certain boundary conditions, which we denote by

$$W^k_{2,BC} = \{h \in W_2^k : h(1) = h^{(1)}(1) = \ldots = h^{(k-1)}(1) = 0\} \quad ,$$

equipped with W_2^k norm ($W^k_{2,BC}$ is a closed subspace of W_2^k). Note that $W_2^0 = W^0_{2,BC} = L_2$, and L_2 norm is equivalent to W_2^0 norm.

It is convenient to define

(1.4) $$\mathcal{Y}_n = R^n \quad , \qquad \langle \underset{\sim}{y}, \underset{\sim}{\eta} \rangle_{\mathcal{Y}_n} = \frac{1}{n} \underset{\sim}{y}' \underset{\sim}{\eta} \quad .$$

Let $K_n \in \mathcal{L}(\mathcal{X},\mathcal{Y}_n)$ denote the linear operator given by

$$K_n h = \underset{\sim}{\eta} \quad \text{iff} \quad \eta_j = \int_0^1 k(u_{jn},v)h(v)dv \ , \quad 1\le j\le n \quad . \tag{1.5}$$

Let $\underset{\sim}{\varepsilon}_n = (\varepsilon_{1n},\ldots,\varepsilon_{nn})'$ denote the error vector, so that the observational model (1.1) may be written

$$\underset{\sim}{y}_n = K_n g + \underset{\sim}{\varepsilon}_n \quad .$$

In the following theorem, we take for parameter space

$$\mathcal{X} = W^2_{2,BC} \quad . \tag{1.6}$$

The estimate $\hat{g}_{n\lambda}$ of g is obtained by minimizing

$$\lambda\|h^{(2)}\|^2_{L_2} + \|\underset{\sim}{y}_n - K_n h\|^2_{\mathcal{Y}_n} = \lambda\int_0^1 [h^{(2)}(u)]^2 du + \frac{1}{n}\sum_{i=1}^n [y_{in} - (Kh)(u_{in})]^2 \quad .$$

The following assumptions are required for the theorem and the remainder of the paper, except Section 2.

$\underset{\sim}{\text{I.}}$ $E\underset{\sim}{\varepsilon}_n = \underset{\sim}{0}$, *and there are constants* $\{s_n\}\subseteq(0,\infty)$ *such that*

$$E\langle\underset{\sim}{\varepsilon}_n,\underset{\sim}{\eta}\rangle^2_{\mathcal{Y}_n} \asymp s_n\|\underset{\sim}{y}\|^2_{\mathcal{Y}_n}$$

uniformly in $\underset{\sim}{\eta}\in\mathcal{Y}_n$, *as* $n\to\infty$.

II. *The design sequence* $\{u_{jn}:1\le j\le n\}\subseteq[0,1]$ *has empirical distribution*

$$F_n(u)=\frac{1}{n}\sum_{j=1}^{n} I_{[-\infty,u]}(u_{jn})$$

such that there is a distribution function F *with density* f *bounded away from* 0 *and* ∞ *on* $[0,1]$ *satisfying*

$$d_n \equiv \sup_{0\le u\le 1}|F(u)-F_n(u)| \to 0$$

as $n\to\infty$.

Theorem 1.1. *An upper bound on the rate of convergence of* $\hat{g}_{n\lambda}$ *to* g *is*

$$E\|K(\hat{g}_{n\lambda}-g)\|^2_{W_2^{2\rho}} \lesssim \min\{1,\lambda^{(t-\rho)}\}\|g\|^2_{W_2^{2t}} + s_n\lambda^{-(\rho+\frac{1}{4})} \tag{1.7}$$

uniformly in $g\in W^{2t}_{2,BC}$ *and* $\lambda\in[\lambda_n,\infty]$. *Here,* ρ *is allowed to take on the values* 0 , ½ , 1 , *and* $t\ge\rho$ *the values* ½ , 1 . $\{\lambda_n\}$ *is any sequence such that as* $n\to\infty$,

$$\lambda_n \to 0 \qquad\qquad d_n\lambda_n^{-(\rho+\frac{1}{4})/6} \to 0 \quad . \tag{1.8}$$

The proof is given in Section 5. The following result gives the optimal upper bound in (1.7) over the range of $\lambda\in[\lambda_n,\infty]$ where λ_n satisfies (1.8).

Corollary 1.2. *If* $g \in W_{2,BC}^{2t}$, $\lambda_n^* = s_n^{4/(4t+1)}$, *and* $d_n \le s_n \to 0$ *as* $n \to \infty$, *then*

$$E\|K(\hat{g}_{n\lambda}-g)\|_{L_2}^2 \lesssim s_n^{4(t-\rho)/(4t+1)} \quad .$$

Remarks. There are several possible improvements in Theorem 1.1. We believe that all mention of boundary conditions can be removed and the theorem still holds. The assumption of boundary conditions greatly simplifies the argument in Section 5. For the application to stereology (see Section 5), this assumption is not unreasonable, as it corresponds to a probability density function, and its first derivative vanishing at an upper truncation point.

The bounds on rate of convergence can probably be improved on. We conjecture that it is possible to obtain

$$E\|K(\hat{g}_{n\lambda}-g)\|_{W_2^{2\rho}}^2 \lesssim \min\{1,\lambda^{(t-\rho+\frac{1}{4})}\}\|g\|_{W_2^{2t}}^2 + s_n\lambda^{-(\rho+1/5)}$$

uniformly in $g \in W_2^{2t}$ and $\lambda \in [\lambda_n,\infty]$ where $\lambda_n \to 0$ and $d_n\lambda_n^{-5/4k} \to 0$.

Note that λ plays a critical role in the convergence rate, and s_n does also, but to a lesser extent. The smoothness properties of g (i.e., value of t) and choice of loss function (i.e., value of ρ) are also important. The design $\{u_{in}, 1 \le i \le n\}$ is only important in determining the cutoff λ_n .

Sections 2 and 3 contain general results on the rate of convergence problem. In Section 2, we give an extension of the work in Cox (1983) (1983) on abstract MOR estimators. It is shown in Section 3 how to fit the particulars of an integral equation problem into the framework of Section 2. There are three important steps in the process. One is the determination of the asymptotic behavior of the eigenvalues of the integral operator Q with kernel

$$q(u,v) = \int_0^1\int_0^1 k(u,x)r(x,y)k(v,y)dxdy$$

where r(x,y) is the reproducing kernel for $\mathcal{X}$ (see Assumption VI and the discussion prior to it in Section 3). The relation between these eigenvalues and the convergence rates of $\hat{g}_{n\lambda}$ to g has already been conjectured by Wahba (1973) and Lukas (1981). A second key step is in showing that $\mathcal{R}(K)$, the range of K , is continuously embedded in a Sobolev space W_2^p for p≥1 (see VII in Section 3). The final important step is the determination of a family of norms defined abstractly in (2.5) and (2.6). It is necessary to obtain estimates on these norms in terms of more readily computible and interpretable norms.

Applications are given in the final two sections. The case of periodic deconvolution is especially simple and is treated at some length in Section 4, primarily for tutorial purposes. Abel's kernel, including the application to stereology, is studied in Section 5.

2. Convergence rates for MOR estimators. This section will enlarge the convergence and approximation theorem in Cox (1983). The extension is useful when we apply those results to situations where the eigenvalues of Q are not known to decay as an exact power, which will be the case for Abel's equation (Section 5). First, we formulate an abstract version of our problem, state the relevant assumptions, prove some important lemmas, and then get to the main result, Theorem 2.1. The remainder of the section assembles the proof, with a couple of other useful results proved along the way.

Assumption A. *Let* $\mathcal{Y}_n$ *and* $\mathcal{X}$ *be separable Hilbert spaces with* $K_n \in \mathcal{L}(\mathcal{X},\mathcal{Y}_n)$ *for all* n . *Suppose*

$$y_n = K_n g + \varepsilon_n$$

where ε_n *is a random element in* $\mathcal{Y}_n$ *satisfying* $E(\varepsilon_n) = 0$ *and* $E\langle\eta,\varepsilon_n\rangle^2_{\mathcal{Y}_n} \asymp s_n\|\eta\|^2_{\mathcal{Y}_n}$ *uniformly in* $\eta\in\mathcal{Y}_n$ *for some sequence* $\{s_n\} \subseteq (0,\infty)$.

Assumption B. $P\in\mathcal{L}(\mathcal{X},\mathcal{X})$ *is such that*

(i) $\langle h,Ph\rangle_{\mathcal{X}} \geq 0$ *for all* $h\in\mathcal{X}$;

(ii) $\dim \mathcal{N}(P) = m < \infty$, *where* $\mathcal{N}(P) = \{\mathcal{X} : P_{\mathcal{X}} = 0\}$;

(iii) *There is an* N *such that for all* $n \geq N$ *there exists* $\bar{\lambda},c$ *in* $(0,\infty)$ *such that for all* $h\in\mathcal{R}(K_n^*) = K_n^*(\mathcal{Y}_n)$,

$$\langle h,(\bar{\lambda}P+K_n^*K_n)^{-1}h\rangle_{\mathcal{X}} \geq c\|h\|^2_{\mathcal{X}} \quad .$$

Assumption C. *There is a self-adjoint operator*

$U \in \mathcal{L}(\mathcal{X},\mathcal{X})$ *such that*

(i) $\mathcal{B}(h_1,h_2) = \langle h_1, Uh_2\rangle_{\mathcal{X}}$ *is positive definite* (i.e., $\mathcal{B}(h,h)>0$ *if* $h\neq 0$) ;

(ii) $\mathcal{B}$ *is completely continuous with respect to* $\mathcal{A}(h_1,h_2) = \langle h_1, Ph_2\rangle_{\mathcal{X}} + \langle h_1, Uh_2\rangle_{\mathcal{X}}$;

(iii) *For some constant* $c\in(0,\infty)$ *and* $\forall h\in\mathcal{X}$, $\mathcal{A}(h,h) \geq c\|h\|^2_{\mathcal{X}}$.

(iv) *The eigenvalues*

$$1 = \alpha_1 = \alpha_1 = \ldots = \alpha_m > \alpha_{m+1} \geq \alpha_{m+2} \geq \ldots > 0 \tag{2.1}$$

of the Rayleigh quotient $\mathcal{B}(h,h)/\mathcal{A}(h,h)$ *satisfy*

$$j^{-q} \lesssim \alpha_j \leq j^{-r} \tag{2.2}$$

for some constants q,r $(0<r\leq q<\infty)$.

Assumption B(iii) guarantees that $(\lambda P+K_n^*K_n)$ has a continuous inverse on $\mathcal{R}(K_n^*)$, the range of K_n^* , and hence that the MOR estimator

$$\hat{g}_{n\lambda} = (\lambda P+K_n^*K_n)^{-1}K_n^*y_n$$

is uniquely defined for $\lambda\in(0,\infty)$. We also put

$$\hat{g}_{no} = (K_n^*K_n)^{\dagger}K_n^*y_n$$

provided the generalized inverse $(K_n^* K_n)^\dagger$ is defined on $\mathcal{R}(K_n^*)$ (see Groetsch, 1977), which is always the case if dim $\mathcal{Y}_n < \infty$. Let $\bar{K}_n \in \mathcal{L}(\mathcal{N}(P), \mathcal{Y}_n)$ denote the restriction of K_n to $\mathcal{N}(P)$, then we put

$$\hat{g}_{n\infty} = (\bar{K}_n^* \bar{K}_n)^\dagger \bar{K}_n^* y_n \quad ,$$

which is well defined as dim $\mathcal{N}(P) < \infty$.

See pages 38-39 and 50-53 of Weinberger (1974) for relevant definitions in Assumption C. It will be convenient to work with a different set of eigenvalues and associated eigenvectors. Let $\{\tilde{\phi}_j\}$ denote the eigenvectors corresponding to $\{\alpha_j\}$, so that

$$\mathcal{A}(\tilde{\phi}_i, \tilde{\phi}_j) = \delta_{ij} \quad , \qquad \mathcal{B}(\tilde{\phi}_i, \tilde{\phi}_j) = \alpha_j \delta_{ij} \quad ,$$

where δ_{ij} denotes Kronecker's delta. It is more convenient to use

$$\phi_j = \alpha_j^{-\frac{1}{2}} \tilde{\phi}_j \quad , \qquad \gamma_j = \alpha_j^{-1} - 1 \quad ,$$

for which we have

(2.3) $$j^r \lesssim \gamma_j \lesssim j^q \quad , \qquad j \to \infty ,$$

(2.4) $$\langle \phi_i, U\phi_j \rangle_{\mathcal{X}} = \delta_{ij} \quad , \qquad \langle \phi_i, P\phi_j \rangle_{\mathcal{X}} = \gamma_j \delta_{ij} \quad .$$

Before stating the final assumption of this section, which relates how U approximates

$$U_n = K_n^* K_n \in \mathcal{L}(\mathcal{X}, \mathcal{X}) \quad ,$$

we introduce a parameterized family of Hilbert spaces. For $\rho\in(0,\infty)$, put

$$\mathcal{X}_\rho^o = \{h\in \mathcal{X}: \sum_{j>m} \gamma_j^\rho \mathcal{E}(h,\phi_j)^2 < \infty\}$$

and for $h_1,h_2\in\mathcal{X}_\rho^o$, let

$$\langle h_1,h_2\rangle_\rho = \sum_{j\le m} \mathcal{E}(h_1,\phi_j)\,\mathcal{E}(h_2,\phi_j) + \sum_{j>m} \gamma_j^\rho\,\mathcal{E}(h_1,\phi_j)\,\mathcal{E}(h_2,\phi_j) \tag{2.5}$$

$$\|h_1\|_\rho^2 = \langle h_1,h_1\rangle_\rho \quad . \tag{2.6}$$

Let $\mathcal{X}_\rho$ be the completion of $\mathcal{X}_\rho^o$, for which we continue to use the same notations for the inner product and norm. Note that an equivalent norm is given by

$$|||h|||_\rho^2 = \mathcal{E}(h,h) + \sum_{j>m} \gamma_j^\rho\,\mathcal{E}(h,\phi_j)^2 \equiv \mathcal{A}_\rho(h,h) \quad . \tag{2.7}$$

Assumption D. *There is a* $p\ge1$ *and a sequence* $d_n\to0$ *such that for all* n *and all* $h_1,h_2\in\mathcal{X}$,

$$|\langle h_1,Uh_2\rangle_{\mathcal{X}} - \langle K_nh_1,K_nh_2\rangle_{\mathcal{Y}_n}| \le d_n\|h_1\|_{1/p}\|h_2\|_{1/p} \quad .$$

Lemma 2.1

(i) *If* $\rho<\rho'$, *then* $\mathcal{X}_{\rho'} \subseteq \mathcal{X}_\rho$, *and* $\mathcal{X}_{\rho'}$ *norm is stronger than* $\mathcal{X}_\rho$ *norm.*

(ii) $\mathcal{X}_1 = \mathcal{X}$ *as sets, and* $\mathcal{X}_1$ *norm is equivalent to* $\mathcal{X}$ *norm.*

(iii) $\|\cdot\|_0$ *norm is equivalent to* $\mathcal{B}(h,h)^{\frac{1}{2}}$.

(iv) $\left|\|h\|_0^2 - \|K_n h\|_{\mathcal{Y}_n}^2\right| \lesssim d_n \|h\|_{1/p}^2$ *for all* $h \in \mathcal{X}_{1/p}$.

(v) *If* $h \in \mathcal{X}_\rho$, *then*

$$h = \sum_{j=1}^{\infty} \mathcal{B}(h,\phi_j)\phi_j$$

where the series converges in $\mathcal{X}_\rho$.

Proof. Part (i) follows from (2.3) which in particular implies $\gamma_j \to \infty$ as $j \to \infty$. Part (ii) follows from C as $\mathcal{A}(h,h) = \||h\||_1^2 \geq c\|h\|_{\mathcal{X}}^2$ by C(iii) , and $\|(P+U)\|^{-1}_{\mathcal{L}(\mathcal{X},\mathcal{X})} \||h\||_1^2 \leq \|h\|_{\mathcal{X}}^2$. Part (iii) is elementary, and (iv) is immediate from Assumption D. For $\rho=1$, (v) follows from the discussion on pages 50-53 of Weinberger (1974), and the argument for general ρ is obtained in a similar fashion by consideration of the quadratic form $\mathcal{A}_\rho(h_1,h_2)$ above.

Q.E.D.

We briefly indicate another interpretation of the spaces $\mathcal{X}_\rho$, following the development on pages 9-10 of Lions and Magenes (1972). Define an unbounded operator in $\mathcal{X}_0$ by

$$Lh = \sum_{j\leq m} \mathcal{B}(h,\phi_j)\phi_j + \sum_{j>m} \gamma_j^{\frac{1}{2}} \mathcal{B}(h,\phi_j)\phi_j .$$

Then the domain of L is $\mathcal{D}(L) = \mathcal{X}_1$, and $\|h\|_1 = \|Lh\|_0$, $\forall h \in \mathcal{X}_1$. This extends to arbitrary $\rho \in (0,\infty)$, i.e., L^ρ may be defined in the usual way as an unbounded self-adjoint operator in $\mathcal{X}_0$, and $\mathcal{D}(L^\rho) = \mathcal{X}_\rho$ with $\|h\|_\rho = \|L^\rho h\|_0$.

The following quantity figures prominently in our development.

$$C(\lambda,\rho) = \sum_{j>m} \gamma_j^\rho (1+\lambda\gamma_j)^{-2} \quad . \tag{2.8}$$

Lemma 2.2. *Fix* $\rho>0$ *, then* $C(\lambda,\rho) < \infty$ *for* $\rho < 2-1/r$. *For all such* ρ *, we have*

$$\lambda^{-(r\rho+1)/q} \lesssim C(\lambda,\rho) \lesssim \lambda^{-(\rho+1/r)} \tag{2.9}$$

as $\lambda \downarrow 0$ *, and*

$$C(\lambda,\rho) \asymp \lambda^{-2} \tag{2.10}$$

as $\lambda \uparrow \infty$.

Proof. The case q=r was already treated in Lemma 2.1 of Cox (1983), so we only sketch the proof. Define the function

$$\omega(u) = \omega(u;\rho) = u^\rho (1+u)^{-2} \quad .$$

Then,

$$C(\lambda,\rho) = \lambda^{-\rho} \sum_{j>m} \omega(\lambda\gamma_j) \quad .$$

Note that ω has its maximum at $u_0 = \rho/(2-\rho)$. To make (2.3) explicit, suppose

$$\gamma_j \geq M_0 j^r \quad , \quad \forall j > m$$

and let

$$n(\lambda) = \max\{m,[(u_0/\lambda M_0)^{1/r}]+1\} \quad .$$

Then, since $\omega(u)$ is decreasing for $u>u_0$,

$$C(\lambda,\rho) \leq \lambda^{-\rho}[\omega(u_0)n(\lambda) + \sum_{j\geq n(\lambda)} \omega(\lambda M_0 j^r)]$$

$$\leq \lambda^{-\rho}[\omega(u_0)n(\lambda) + (M_0\lambda)^{-1/r} \int_{u_0^{1/r}}^{\infty} x^{\rho r}(1+x^r)^{-2}dx] \quad .$$

Note that the integral is finite if $\rho<2-1/r$. If this is the case, then both terms in brackets in the last expression are $\mathcal{O}(\lambda^{-1/r})$ so the upper bound in (2.9) follows. For the lower bound, suppose $\gamma_j \leq M_1 j^q$, $j>m$, and use

$$\lambda^{-\rho}\omega(\lambda\gamma_j) \geq M_0^{\rho} j^{r\rho}/(1+\lambda M_1 j^q)^2 = (M_0^{\rho}/M_1^{r\rho/q})\lambda^{-r\rho/q}\omega(M_1\lambda j^q;r\rho/q) \quad ,$$

valid for $j>m$. Hence,

$$C(\lambda,\rho) \geq M\lambda^{-r\rho/q} \sum_{j>m} \omega(M_1\lambda_j{}^q;\ r\rho/q) \quad .$$

Now as $\lambda \downarrow 0$,

$$(M_1\lambda)^{1/q} \sum_{j>m} \omega(M_1\lambda_j^{\,q};r\rho/q) \to \int_0^\infty x^{r\rho}(1+x^q)^{-2}dx \quad .$$

If this is used in the previous expression, the lower bound in (2.9) is obtained. Finally, the estimates in (2.10) follow just as in Cox (1983).

Q.E.D.

The next lemma provides the estimates that are needed to extend the main theorem of Cox (1983) to the case when $q>r$. The proof is lengthy, but straightforward.

Lemma 2.3. *For* $\rho\in[0,2-1/r-1/p)$, *define*

$$\text{(2.11)} \qquad \mu(\rho) = 1/2\{(1-r/q)\rho+2/p+2/r-1/q\} \quad .$$

If $\{\lambda_n\} \subseteq (0,\infty)$ *is a sequence for which* $\lambda_n \to 0$ *as* $n\to\infty$, *then each of the quantities*

$$\text{(2.12a)} \qquad C(\lambda,\rho+1/p)/C(\lambda,\rho) \quad ,$$

$$\text{(2.12b)} \qquad C(\lambda,\rho+1/p)C(\lambda,1/p)/C(\lambda,\rho) \quad ,$$

$$\text{(2.13)} \qquad C(\lambda,\rho+1/p)\lambda^{(\rho-1/p)}$$

are $\lesssim \lambda_n^{-2\mu(\rho)}$, *uniformly in* $\lambda\in[\lambda_n,\infty]$.

In this general context, the MOR estimator $\hat{g}_{n\lambda}$ is obtained by minimization of

$$\lambda\langle h,Ph\rangle_{\mathcal{X}} + \langle K_n h,K_n h\rangle_{\mathcal{Y}_n} - 2\langle K_n h,y_n\rangle_{\mathcal{Y}_n} \quad ,$$

or, if $\dim \mathcal{Y}_n < \infty$, of

$$\lambda\langle h,Ph\rangle_{\mathcal{X}} + \|K_n h-y_n\|^2_{\mathcal{Y}_n} \quad .$$

One can show by a standard argument that the resulting estimate is given by

$$\hat{g}_{n\lambda} = (\lambda P+U_n)^{-1}K_n^* y_n = (\lambda P+U_n)^{-1}U_n g + (\lambda P+U_n)^{-1}K_n^*\epsilon_n \quad . \tag{2.14}$$

We now state the main theorem of this section.

Theorem 2.1. *Assume A, B, C, and D. Fix* ρ *such that*

$$0 \le \rho < 2 - 1/r - 1/p \quad . \tag{2.15}$$

Let the sequence $\{\lambda_n\} \subseteq (0,\infty)$ *satisfy*

$$\lambda_n \to 0 \quad , \qquad d_n\lambda_n^{-\mu(\rho)} \to 0 \tag{2.16}$$

as $n\to\infty$. *Then the following asymptotic relations hold uniformly in* $\lambda\in[\lambda_n,\infty]$ *as* $n\to\infty$.

(a) *If* $Pg=0$, *then*

$$E\|\hat{g}_{n\lambda}-g\|_\rho^2 \asymp s_n\{C(\lambda,\rho)+m\} \tag{2.17}$$

(b) *If* $g\in\mathcal{X}_t$ *for some* t *satisfying*

$$1/p \le t \le 2+1/p \quad , \qquad \rho < t \le \rho+2 \quad , \tag{2.18}$$

then

$$E\|\hat{g}_{n\lambda}-g\|_\rho^2 \lesssim \min\{1,\lambda^{(t-\rho)}\}\|g\|_t^2 + s_n\{m+C(\lambda,\rho)\}$$

uniformly in $g\in\mathcal{X}_t$.

(c) *If* $g\in\mathcal{X}_\rho$ *and* $\rho \ge 1/p$, *then*

$$E\|\hat{g}_{n\lambda}-g\|_\rho^2 \lesssim \min\{1,o(1)\} + s_n\{m+C(\lambda,\rho)\} \tag{2.19}$$

where $o(1)$ *is as* $\lambda\to 0$, *and the relation is uniform in* $\lambda\in[\lambda_n,\infty]$.

(d) *If also* $d_n\lambda_n^{-(1/p+1/2r+1)}\to 0$, *and if* t *satisfies* (2.18), *then*

$$\min\{1,\lambda^2\}\sum_{j>m}\gamma_j^t\mathcal{E}^2(g,\phi_j) + s_n\{m+C(\lambda,\rho)\} \lesssim E\|\hat{g}_{n\lambda}-g\|_\rho^2$$

uniformly in $g\in\mathcal{X}_t$.

To prove the theorem, we introduce the asymptotic bias and variability operators

(2.20a) $$B_\lambda = (\lambda P+U)^{-1}U-I \in \mathcal{L}(\mathcal{X},\mathcal{X}) \quad ,$$

(2.20b) $$V_{n\lambda} = (\lambda P+U)^{-1}K_n^* \in \mathcal{L}(\mathcal{Y}_n,\mathcal{X}) \quad ,$$

where I denotes an identity operator. The proof consists of showing that $\hat{g}_{n\lambda}$ is well approximated by

(2.21) $$\tilde{g}_{n\lambda} = (\lambda P+U)^{-1}Ug + V_{n\lambda}\,\varepsilon_n \quad .$$

It will be proven that each of the asymptotic relations (2.17) and (2.19) hold with $\hat{g}_{n\lambda}$ replaced by $\tilde{g}_{n\lambda}$, and then that $E\|\hat{g}_{n\lambda}-\tilde{g}_{n\lambda}\|_\rho^2$ is asymptotically negligible in comparison with the bound on $E\|\tilde{g}_{n\lambda}-g\|_\rho^2$. Note that $E\varepsilon_n=0$ implies

(2.22) $$E\|\tilde{g}_{n\lambda}-g\|_\rho^2 = \|B_\lambda g\|_\rho^2 + E\|V_{n\lambda}\varepsilon_n\|_\rho^2 \quad ,$$

which is the familiar bias squared plus variance formula. The next theorem establishes the convergence rate for the bias, and Theorem 2.3 deals with the variance.

Theorem 2.2. *Let* $\rho\geq 0$ *be given.*

(a) $\|B_\lambda g\|_\rho << \lambda$ *as* $\lambda\downarrow 0$ *if and only if* $g\in\mathcal{N}(P)$. *In this case,* $B_\lambda g=0$, $\forall\lambda$.

(b) $\|B_\lambda g\|_\rho \lesssim \lambda$ *as* $\lambda \downarrow 0$ *if and only if* $g \in \mathcal{X}_{\rho+2}$.

(c) *Given any* $\alpha \in [0,1]$, $\|B_\lambda g\|_\rho \leq \lambda^\alpha \|g\|_{\rho+2\alpha}$ *as* $\lambda \downarrow 0$, *uniformly in* $g \in \mathcal{X}_{\rho+2\alpha}$.

(d) *If* $g \in \mathcal{X}_0$ *and* $\rho > 0$ *are such that* $\|B_\lambda g\|_\rho \lesssim \lambda^\alpha$ *for some* $\alpha \in (0,1)$, *then for every* $\delta > 0$, $g \in \mathcal{X}_{\rho+2\alpha-\delta}$.

(e) *If* $g \in \mathcal{X}_\rho$, *then* $\|B_\lambda g\|_\rho \to 0$ *as* $\lambda \downarrow 0$, *and* $\|B_\lambda g\|_\rho \leq \|g\|_\rho$ *for all* $\lambda \geq 0$.

(f) *For* $0 \leq \alpha \leq 1$, $\|B_\lambda g\|_\rho \geq \lambda^2 \sum_{j>m} \gamma_j^{\rho+2\alpha} \mathcal{E}^2(g,\phi_j)$ *uniformly in* $g \in \mathcal{X}_{\rho+2\alpha}$, *as* $\lambda \downarrow 0$.

Proof. Parts (a) and (b) follow from Lemma 3.5 of Cox (1983). Part (c) follows from Lemma 3.4(i) of that paper. Part (d) follows from Lemma 3.5 of that paper, and the remark there just prior to Assumption D. For part (e), note that if $g \in \mathcal{X}_\rho$, then $g = \Sigma a_j \phi_j$ where the scalars a_j satisfy $\Sigma \gamma_j^\rho a_j^2 < \infty$. Since

$$B_\lambda \phi_j = [\lambda\gamma_j/(1+\lambda\gamma_j)]\phi_j \qquad ,$$

we have

$$\|B_\lambda g\|_\rho^2 = \sum_{j>m} \lambda^2 \gamma_j^2 (1+\lambda\gamma_j)^{-2} \gamma_j^\rho a_j^2 \qquad .$$

As $\lambda \downarrow 0$, each term in the summation decreases to 0 .

Since the j*th* summand is bounded by $\gamma_j^\rho a_j^2$ (which implies $\|B_\lambda g\|_\rho \le \|g\|_\rho$), the dominated convergence theorem yields that $\|B_\lambda g\|_\rho^2 \downarrow 0$ as $\lambda \downarrow 0$.

For (f), we have

$$\|B_\lambda g\|_\rho^2 = \lambda^2 \sum_{j>m} (1+\lambda\gamma_j)^{-2} \gamma_j^{2(1-\alpha)} \gamma_j^{\rho+2\alpha} a_j^2$$

$$\ge \lambda^2 \ \gamma_{m+1}^{2(1-\alpha)} \sum_{j>m} (1+\lambda\gamma_j)^{-2} \gamma_j^{\rho+2\alpha} a_j^2 \quad .$$

Now as $\lambda \downarrow 0$,

$$\sum_{j>m} (1+\lambda\gamma_j)^{-2} \gamma_j^{\rho+2\alpha} a_j^2 \uparrow \sum_{j>m} \gamma_j^{\rho+2\alpha} a^2 \qquad ,$$

by dominated convergence, which completes the proof.

Q.E.D.

Remark. One can give a sharper characterization of g in part (d) [see Lemma 3.5 of Cox (1983)]. Unfortunately, the result is difficult to use. The result in (c) is best possible, and one can obtain an asymptotic equivalence when $\alpha=1$ by (f), but apart from this case, one cannot replace the one-sided relation $\lesssim$ by $\asymp$, and in fact, a bound of the form $\gtrsim \lambda^u$ for some $u<2$ appears impossible, even if uniformity in g is dropped.

Theorem 2.3. *Suppose* $\{\lambda_n\} \subseteq (0,\infty)$ *satisfies* (2.16). *Then if* ρ *satisfies* (2.15),

$$E\|V_{n\lambda}\varepsilon_n\|_\rho^2 \asymp s_n[m+C(\lambda,\rho)]$$

as $n\to\infty$, *uniformly in* $\lambda\in[\lambda_n,\infty]$.

Proof. From the definition of $\mathcal{X}_\rho$ norm, and self-adjointness of $(\lambda P+U)^{-1}$,

$$E\|V_{n\lambda}\varepsilon_n\|_\rho^2 = E[\sum_{j\le m} |\langle\varepsilon_n, K_n(\lambda P+U)^{-1}U\phi_j\rangle_{\mathcal{Y}_n}|^2 + \sum_{j>m} \gamma_j^\rho |\langle\varepsilon_n, K_n(\lambda P+U)^{-1}U\phi_j\rangle_{\mathcal{Y}_n}|^2]..$$

Using the fact that

$$(\lambda P+U)^{-1}U\phi_j = (1+\lambda\gamma_j)^{-1}\phi_j$$

[see equation (3.12) of Cox (1983)], along with the bounds in A, it follows that

$$E\|V_{n\lambda}\varepsilon_n\|_\rho^2 \le Ms_n[\sum_{j\le m} \langle K_n\phi_j, K_n\phi_j\rangle_{\mathcal{Y}_n} + \sum_{j>m} \gamma_j^\rho(1+\lambda\gamma_j)^{-2}\langle K_n\phi_j, K_n\phi_j\rangle_{\mathcal{Y}_n}]$$

for some constant $M>0$. Now add and subtract $\gamma_j^\rho(1+\lambda\gamma_j)^{-2}\langle\phi_j, U\phi_j\rangle_{\mathcal{X}}$ in each summand and use the fact that $\|\phi_j\|_{1/p}^2 = 1$ for $j\le m$ and $= \gamma_j^{1/p}$ for $j>m$ along with D to obtain

$$E\|V_{n\lambda}\varepsilon_n\|_\rho^2 \le Ms_n[(1+d_n)m+d_nC(\lambda,\rho+1/p) + C(\lambda,\rho)] \quad .$$

By the hypothesis (2.16) and Lemma 2.3,

$$d_n C(\lambda,\rho+1/p) << C(\lambda,\rho) \quad .$$

Thus,

$$E\|V_{n\lambda}\varepsilon_n\|_\rho^2 \lesssim s_n[m+C(\lambda,\rho)] \quad .$$

A similar argument provides the opposite relation of $\gtrsim$ in the above, and hence the theorem.

Q.E.D.

Proof of Theorem 2.1. From the previous two theorems, we see that the convergence rates in Theorem 2.1 hold if $\hat{g}_{n\lambda}$ is replaced by $\tilde{g}_{n\lambda}$. Since

$$|E\|\tilde{g}_{n\lambda}-g\|_\rho^2-E\|\hat{g}_{n\lambda}-g\|_\rho^2| \lesssim E\|\hat{g}_{n\lambda}-\tilde{g}_{n\lambda}\|_\rho^2 + 2(E\|\hat{g}_{n\lambda}-\tilde{g}_{n\lambda}\|_\rho^2 E\|\tilde{g}_{n\lambda}-g\|_\rho^2)^{\frac{1}{2}} \quad ,$$

it suffices to prove

$$E\|\hat{g}_{n\lambda}-\tilde{g}_{n\lambda}\|_\rho^2 << s_n[m+C(\lambda,\rho)]$$

under the hypotheses of (a), and

$$E\|\hat{g}_{n\lambda}-\tilde{g}_{n\lambda}\|_\rho^2 << \min\{1,\lambda^{2(t-\rho)}\} + s_n[m+C(\lambda,\rho)]$$

under (b). By Lemma 3.3 of Cox (1983),

$$E\|\hat{g}_{n\lambda}-\tilde{g}_{n\lambda}\|_\rho^2 \le 4d_n^2[m+C(\lambda,\rho+1/p)][\|B_\lambda g\|_{1/p}^2 + E\|V_{n\lambda}\varepsilon_n\|_{1/p}^2]$$

provided that

$$d_n^2[2m+C(\lambda,1/p) + C(\lambda,2/p)] \le 1/9 \quad .$$

Using Lemma 2.2 and the hypothesis (2.16), one can verify that this inequality does hold for all n sufficiently large, uniformly for $\lambda\varepsilon[\lambda_n, \infty]$. Using (2.12a) and (2.12b) from Lemma 2.3,

$$d_n^2[m+C(\lambda,\rho+1/p)]E\|V_{n\lambda}\varepsilon_n\|_{1/p}^2 << s_n[m+C(\lambda,\rho)] \quad .$$

In view of Theorems 2.1 and 2.2(a), part (a) is now evident, so we proceed to treat the bias term under the hypotheses of (b). By (2.13) in Lemma 2.3,

$$d_n^2[m+C(\lambda,\rho+1/p)]\|B_\lambda g\|_{1/p}^2 \lesssim d_n^2\lambda_n^{-2\mu(\rho)}\min\{1,\lambda^{(t-\rho)}\}\|g\|_t^2 << \min\{1,\lambda^{(t-\rho)}\}\|g\|_t^2$$

where the last relation follows from (2.16) again. This same result also proves part (c). Part (d) follows from Theorem 2.2(f) and a similar calculation.

Q.E.D.

Corollary 2.4. *If* $1/p<t\leq 2$ *, then*

$$\left|E\|K_n(\hat{g}_{n\lambda}-g)\|^2_{\mathcal{Y}_n} - E\|\hat{g}_{n\lambda}-g\|^2_0\right| << \min\{1,\lambda^t\}\|g\|^2_t + s_n[m+C(\lambda,0)]$$

uniformly for $\lambda\varepsilon[\lambda_n,\infty]$ *and* $g\varepsilon\chi_t$*, provided* $\lambda_n\to 0$, $d_n\lambda_n^{-\mu(0)}\to 0$. *In particular,*

$$E\|K_n(\hat{g}_{n\lambda}-g)\|^2_{\mathcal{Y}_n} \lesssim \min\{1,\lambda^t\}\|g\|^2_t + s_n[m+C(\lambda,0)] \quad .$$

If $g\varepsilon\eta(P)$*, then the term* $\min\{1,\lambda^t\}\|g\|^2_t$ *may be dropped from each upper bound.*

Proof. By Lemma 2.1(iv),

$$\left|\|K_n(\hat{g}_{n\lambda}-g)\|^2_{\mathcal{Y}_n} - \|\hat{g}_{n\lambda}-g\|^2_0\right| \lesssim d_n\|\hat{g}_{n\lambda}-g\|^2_{1/p}$$

and by Theorem 2.1(b),

$$E\|\hat{g}_{n\lambda}-g\|^2_{1/p} \lesssim \min\{1,\lambda^{t-1/p}\}\|g\|^2_t + s_n[m+C(\lambda,1/p)] \quad .$$

The first bound follows from this and Lemma 2.3, while the second follows from the first. The statements about $g\epsilon\eta(P)$ follow in a similar way from Theorem 2.1(a).

Q.E.D.

3. Application to Integral Equations. A collection of hypotheses will be stated here which will allow application of Theorem 2.1 to integral equations. The assumptions stated below, along with I and II from the Introduction, will imply A - D of Section 2.

III. *$\mathcal{X} \subseteq L_2[0,1]$ is an RKHS with continuous kernel $r:[0,1] \times [0,1] \to \mathbb{R}$. Let $R \in \mathcal{L}(L_2[0,1], L_2[0,1])$ denote the corresponding integral operator*

$$(3.1) \qquad (Rh)(u) = \int_0^1 r(u,v)h(v)dv \quad ,$$

with null space $\mathcal{N}(R) = \{0\}$.

IV. *$P \in \mathcal{L}(\mathcal{X},\mathcal{X})$ is such that the quadratic form $h \mapsto \langle h, Ph\rangle_{\mathcal{X}}$ is positive semidefinite and satisfies*

$$\dim \mathcal{N}(P) = m < \infty$$

and for all n sufficiently large,

$$(3.2) \qquad \mathcal{N}(P) \cap \mathcal{N}(K_n) = \{0\} \quad .$$

V. *Suppose $K \in \mathcal{L}(L_2, L_2)$ is the integral operator with kernel $k:[0,1] \times [0,1] \to \mathbb{R}$ satisfying*

(i) $\mathcal{N}(K) = \{0\}$ *and* $\mathcal{R}(K)$ *is dense in* L_2

(ii) $\sup_{u \in [0,1]} \int_0^1 |k(u,v)|dv < \infty$

(iii) *If* $\tilde{K}\epsilon\mathcal{L}(\mathcal{X},L_2(F))$ *denotes the restriction of* K *to* $\mathcal{X}$ *, then* $(P+\tilde{K}^*\tilde{K})^{-1}\ \epsilon\mathcal{L}(\mathcal{X},\mathcal{X})$.

Before stating the next assumption, note that R is a compact operator by Mercer's theorem, so

$$Q = KRK^*\epsilon\mathcal{L}(L_2,L_2) \tag{3.3}$$

is also compact, and hence has a spectral decomposition with eigenvalues

$$\mu_1\geq\mu_2\geq\mu_3\geq\ \cdots\qquad > 0 \tag{3.4}$$

where $\mu_j\to 0$ as $j\to\infty$. Positivity of the eigenvalues follows as $\langle h,Qh\rangle_{L_2} = \langle(K^*h),R(K^*h)\rangle_{L_2} > 0$, $\forall h\neq 0$, by III and V.

VI. *The eigenvalues* (3.4) *of* Q *satisfy*

$$j^{-q} \lesssim \mu_j \lesssim j^{-r}$$

for some positive $r\leq q$.

VII. *There is a* $p\geq 1$ *and a constant* M $(0<M<\infty)$ *such that for all* $h\epsilon\mathcal{X}$,

$$\|Kh\|_{W_2^p} \leq M\|h\|_{\mathcal{X}} \quad .$$

In the following discussion, it is shown that I - VII imply A - D. For A, it is necessary to show K_n as given in (1.5) is in $\mathcal{L}(\mathcal{X},\mathcal{Y}_n)$, for which is suffices to show that there exists $M\in(0,\infty)$ such that for $h\in\mathcal{X}$,

$$\left|\int k(u_{jn},v)h(v)dv\right| \le M\|h\|_{\mathcal{X}} \quad , \qquad 1\le j\le n \quad .$$

Applying V(ii) gives

$$\left|\int k(u,v)h(v)dv\right| \le [\sup_u \int |k(u,v)|dv][\sup_v |h(v)|]$$

$$\le [\sup_u \int |k(u,v)|dv][\sup_u \sqrt{r(u,u)}]\|h\|_{\mathcal{X}} \quad ,$$

where the last line follows from continuity of r in III and Proposition 1, page 116, of Aubin (1979). The remaining part of A is immediate from I. In view of IV, for B it is only necessary to establish that $(P+K_n^*K_n)^{-1}$ exists as a bounded linear operator on $\mathcal{R}(K_n^*)$ to $\mathcal{X}$ which follows from (3.2) and the fact that $\dim \mathcal{R}(K_n^*) \le n < \infty$.

The Assumptions C and D are more involved. We first motivate the choice of the correct operator U. Referring to D, we wish to approximate (for $h_1,h_2\in\mathcal{X}$),

$$(3.5) \qquad \langle K_nh_1,K_nh_2\rangle_{\mathcal{Y}_n} = \frac{1}{n}\sum_{i=1}^{n}\left(\int k(u_i,v)h_1(v)dv\right)\left(\int k(u_i,v)h_2(v)dv\right)$$

$$= \iint h_1(v)w_n(v,v')h_2(v')dvdv'$$

where the kernel w_n is given by

$$w_n(v,v') = \int k(u,v)k(u,v')dF_n(u) \quad .$$

If we replace w_n by its continuous analog

$$w(v,v') = \int k(u,v')k(u,v')dF(u) \ ,$$

then the last expression in (3.5) becomes

$$\iint h_1(v)w(v,v')h_2(v')dvdv' = \langle \tilde{K}h_1, \tilde{K}h_2 \rangle_{L_2(F)} = \langle h_1, \tilde{K}^*\tilde{K}h_2 \rangle_{\mathcal{X}}$$

where $\tilde{K}$ is given in V(iii). Hence,

$$U = \tilde{K}^*\tilde{K} \tag{3.6}$$

is the correct choice. We now check that C holds for this operator. Clearly, $U \in \mathcal{L}(\mathcal{X},\mathcal{X})$, and C(i) holds by V(i). Condition VII says that K is a bounded linear operator $\mathcal{X} \to W_2^p$, some $p \geq 1$, and since the unit ball of W_2^p is relatively compact in L_2 (Remark 4.10.2.4, page 353 of Triebel, 1978), it follows that K is a compact operator $\mathcal{X} \to L_2$, i.e., $\tilde{K}$ is compact. Therefore, U is compact. Hence, if we show C(iii) (so that $\mathcal{Q}(h,h)^{\frac{1}{2}}$ is equivalent to $\|\cdot\|$), then C(ii) follows as well (see page 53 of Weinberger, 1974), but C(iii) is immediate from V(iii). The remaining assumptions from Section 2 [C(iv) and D] follow from the next two theorems in conjunction with VI and VII. First, we need the following technical result, which should be well known, although we cannot find it stated anywhere.

Lemma 3.1. *The operator* $R^{\frac{1}{2}}$ *is a Hilbert space isomorphism* $L_2 \to \mathcal{X}$.

Proof. By III and Mercer's theorem (Riesz and Sz. Nagy 1955), there is an orthonormal bias $\{\zeta_j: j=1,2,\ldots\}$ of L_2 and a sequence of corresponding eigenvalues $\{\nu_j: j=1,2,\ldots\}$ with $\nu_j>0$, $\forall_j$, such that

$$r(u,v) = \sum_{j=1}^{\infty} \nu_j \zeta_j(u)\zeta_j(v) \tag{3.7}$$

where the convergence is uniform on $[0,1] \times [0,1]$. Now let

$$h(v) = \sum_{i=1}^{N} a_i r(u_i,v)$$

be a finite linear combination of $r(u_1,\cdot),\ldots,r(u_n,\cdot)$. Note that the collection $\mathcal{M}$ of all such h's (obtained by letting N and $u_1,\ldots,u_N$ vary) is dense in $\mathcal{X}$ (Proposition 1, page 116 of Aubin, 1979). Thus,

$$\begin{aligned} \langle h,h\rangle_{\mathcal{X}} &= \sum_i \sum_k a_i \langle r(u_i,\cdot),r(u_k,\cdot)\rangle_{\mathcal{X}} a_k \\ &= \sum_i \sum_k a_i r(u_i,u_k) a_k \end{aligned}$$

where the latter follows from the "reproducing" property of r . Substituting (3.7) in this last expression gives

$$\langle h,h\rangle_{\mathcal{X}} = \sum_j \nu_j \Big(\sum_i a_i \zeta_j(u_i)\Big)^2 < \infty \quad ,$$

which implies

$$\theta = \sum_j \nu_j^{\frac{1}{2}} \Big(\sum_i a_i \zeta_j(u_i)\Big)\zeta_j \in L_2 \quad .$$

One checks that $\langle\theta,\theta\rangle_{L_2} = \langle h,h\rangle_{\mathcal{X}}$, and also

$$\text{(3.8)} \qquad R^{\frac{1}{2}}\theta = \sum_j \nu_j\Big(\sum_i a_i\zeta_j(u_i)\Big)\zeta_j$$

$$= \sum_i a_i \sum_j \nu_j\zeta_j(u_i)\zeta_j$$

$$= \sum_i a_i r(u_i,\cdot) = h \quad .$$

We have shown that there is an operator $T_o : \mathcal{M} \to L_2$ given by

$$T_o\Big(\sum_i a_i r(u_i,\cdot)\Big) = \sum_j \nu_j\Big(\sum_i a_i\zeta_j(u_i)\Big)\zeta_j$$

which satisfies $\langle h,h\rangle_{\mathcal{X}} = \langle T_o h, T_o h\rangle_{L_2}$, for all $h\in\mathcal{M}$. By the density of $\mathcal{M}$ in $\mathcal{X}$, T_o extends to $T\in\mathcal{L}(\mathcal{X},L_2)$, satisfying $\langle h,h\rangle_{\mathcal{X}} = \langle Th,Th\rangle_{L_2}$ (see Theorem 3, page 14 of Aubin, 1979, for the existence of T, and apply continuity to prove the relation). The calculation (3.8) shows $R^{\frac{1}{2}}T$ is the identity on $\mathcal{X}$, so it only remains to show $\mathcal{R}(T) = L_2$. Suppose $\theta\in L_2$ is such that $\theta\perp\mathcal{R}(T)$, so that for all $h = \sum_i a_i r(u_i,\ )\in\mathcal{M}$,

$$0 = \langle\theta,Th\rangle_{L_2} = \Big\langle\theta,\sum_j \nu_j\Big(\sum_i a_i\zeta_j(u_i)\Big)\zeta_j\Big\rangle_{L_2}$$

$$= \sum_i a_i \sum_j \nu_j\langle\theta,\zeta_j\rangle_{L_2}\zeta_j(u_i)$$

$$= \sum_i a_i(R\theta)(u_i) \quad .$$

As the $\{a_i\}$ and $\{u_i\}$ are arbitrary, we must have $R\theta = 0$ (note $R\theta$ is a continuous function by continuity of r), which implies $\theta = 0$ as $\mathcal{N}(R) = \{0\}$ (see III).

Q.E.D.

Theorem 3.1. *The eigenvalues* $\{\mu_j\}$ *of* Q *and the eigenvalues* $\{\alpha_j\}$ *of* $\mathcal{B}/\mathcal{A}$ *in* (2.1) *with* U *given in* (3.6) *satisfy*

$$\mu_j \asymp \alpha_j \quad , \qquad j \to \infty \quad .$$

Proof. We first show that it is possible to assume P is the identity on $\mathcal{X}$ and that $F(x) = x$ on $[0,1]$ (i.e., the design points $\{u_{in}\}$ are asymptotically uniform). Boundedness of F' away from 0 and ∞ by II, and continuity of P+U and $(P+U)^{-1}$ on $\mathcal{X}$ by IV and V(iii) yields the existence of constants $c_1, c_2 \in (0,\infty)$ such that for all $h \in \mathcal{X}$,

$$c_1\tilde{\mathcal{B}}(h,h) \le \mathcal{B}(h,h) \le c_2\tilde{\mathcal{B}}(h,h) \quad ,$$

$$c_1\tilde{\mathcal{A}}(h,h) \le \mathcal{A}(h,h) \le c_2\tilde{\mathcal{A}}(h,h) \ ,$$

where

$$\tilde{\mathcal{B}}(h_1,h_2) = \langle Kh_1, Kh_2\rangle_{L_2}$$

$$= \int\left(\int k(u,v)h_1(v)dv\right)\left(\int k(u,v)h_2(v)dv\right)du$$

and

$$\tilde{\mathcal{A}}(h_1,h_2) = \langle h_1,h_2\rangle_{\mathcal{X}} + \tilde{\mathcal{E}}(h_1,h_2) \quad .$$

It then follows that for $h \in \mathcal{X}$, $h \neq 0$,

$$\tilde{\mathcal{B}}(h,h) \geq \left(\frac{c_1}{c_2}\right)\left(\frac{\mathcal{B}(h,h)}{\mathcal{A}(h,h)}\right)\tilde{\mathcal{A}}(h,h)$$

and

$$\tilde{\mathcal{A}}(h,h) \geq c_2^{-1}\,\mathcal{A}(h,h) \quad .$$

Now let $\{\tilde{\alpha}_j\}$ denote the eigenvalues of $\tilde{\mathcal{B}}/\tilde{\mathcal{A}}$. By the last two displayed inequalities and the Mapping Principle on page 57 of Weinberger (1974),

$$\tilde{\alpha}_j \geq \left(\frac{c_1}{c_2}\right)\alpha_j \quad , \quad \forall_j \quad .$$

A similar argument shows

$$\tilde{\alpha}_j \leq \frac{c_2}{c_1}\,\alpha_j \quad , \quad \forall_j \quad .$$

Hence, it suffices to show $\mu_j \asymp \tilde{\alpha}_j$, i.e., we may assume P=I and $F'=1$, since the operators K and R (and hence Q) do not depend on P and F.

Within this context, the eigenvalues $\{\gamma_j\}$ and eigenfunctions $\{\phi_j\}$ of (2.4) are defined by the relations

$$(3.9) \qquad \langle \phi_i , \phi_j \rangle_{\mathcal{X}} = \gamma_i \delta_{ij} \quad , \qquad \langle K\phi_i , K\phi_j \rangle_{L_2} = \delta_{ij} \quad .$$

Now let $T = R^{-\frac{1}{2}}$ denote the isomorphism $\mathcal{X} \rightarrow L_2$ given in the previous lemma, and put

$$\psi_j = \gamma_j^{-\frac{1}{2}} T\phi_j \quad , \qquad \forall_j \quad .$$

Noting that $T^{-1} = R^{\frac{1}{2}}$ as a point function, the relations (3.9) become

$$\langle \psi_i , \psi_j \rangle_{L_2} = \delta_{ij}$$

$$\langle KT^{-1}\psi_i , KT^{-1}\psi_j \rangle_{L_2} = \langle KR^{\frac{1}{2}}\psi_i , KR^{\frac{1}{2}}\psi_j \rangle_{L_2}$$

$$= \langle \psi_i , R^{\frac{1}{2}}K^* KR^{\frac{1}{2}}\psi_j \rangle_{L_2} = \gamma_i^{-1}\delta_{ij}$$

[Note that $R^{\frac{1}{2}} \in \mathcal{L}(L_2,L_2)$ is self-adjoint], so we see that $\{\gamma_j^{-1}\}$ and $\{\psi_j\}$ are the usual eigenvalues and eigenvectors of the operator

$$\tilde{Q} = R^{\frac{1}{2}}K^* KR^{\frac{1}{2}} = (KR^{\frac{1}{2}})^* KR^{\frac{1}{2}} \in \mathcal{L}\,(L_2,L_2) \quad .$$

Since $\gamma_j^{-1} = (\alpha_j^{-1}-1)^{-1} \asymp \alpha_j$ as $j \to \infty$, it suffices to show $\mu_j = \gamma_j^{-1}$, $\forall_j$.

Writing $A = KR^{\frac{1}{2}}$, then $Q = KRK^* = AA^*$ and $\tilde{Q} = A^*A$. Put $\eta_j = A\psi_j$, then

$$Q\eta_j = AA^*A\psi_j = A(\gamma_j^{-1}\psi_j) = \gamma^{-1}\eta_j \quad .$$

This shows that each η_j is an eigenvector for Q and has eigenvalue γ_j^{-1} . Further, as $\mathcal{N}(A) = \{0\}$ by III and V(i), the η_j's corresponding to the same eigenvalue are distinct, and hence the multiplicity of γ_j^{-1} as an eigenvalue of Q is at least equal to its multiplicity as an eigenvalue of $\tilde{Q}$. Further, by reversing the roles of A^* and A , and of η_j and ψ_j , one sees that the multiplicities are equal. Note that $\mathcal{N}(A^*) = \{0\}$ as $\mathcal{N}(R^{\frac{1}{2}}) = \{0\}$ by III and $\mathcal{N}(K^*) = \mathcal{R}(K)^\perp = \{0\}$ as $\mathcal{R}(K)$ is dense by V(i). As neither Q nor $\tilde{Q}$ has zero in its spectrum, they have exactly the same spectrum, and hence $\{\gamma_j^{-1}\}$ is the collection of eigenvalues of Q, completing the proof the theorem.

Q.E.D.

Theorem 3.2. II *and* VII *imply* D , *with the same values of* d_n *and* p .

Proof. The proof is given in Lemmas 4.1 and 4.2 of Cox (1983), so we only sketch it here. Firstly, for $f_1, f_2 \in W_2^1$, an integration by parts followed by Cauchy-Schwarz shows

$$\left|\int f_1 f_2 d(F_n - F)\right| \le d_n\{[\int (f_1')^2 \int f_2^2]^{\frac{1}{2}} + [\int f_1^2 \int (f_2')^2]\}^{\frac{1}{2}}$$

$$\le 2d_n \|f_1\|_{W_2^1} \|f_2\|_{W_2^1}$$

So, if $f_i = Kh_i$ for $h_i \in \mathcal{X}$, $i=1,2$, then by VII, $f_i \in W_2^p$, $p \geq 1$, which implies $f_i \in W_2^1$, and hence

$$\begin{aligned} &|\langle h_1, Uh_2\rangle_{\mathcal{X}} - \langle K_n h_1, K_n h_2\rangle_{\mathcal{H}_n}| \\ &= \left|\int (Kh_1)(Kh_2)\,d(F-F_n)\right| \\ &\leq 2\,d_n \|Kh_1\|_{W_2^1} \|Kh_2\|_{W_2^1} \quad . \end{aligned}$$

To complete the proof, we show that

$$\text{(3.10)} \qquad \|Kh\|_{W_2^1} \leq M\|h\|_{1/p} \quad , \qquad \forall h \in \mathcal{X}$$

for some constant $M \in (0,\infty)$. The proof is standard interpolation theory. One may regard $K \in \mathcal{L}(\mathcal{X}_1, W_2^p)$ and $K \in \mathcal{L}(\mathcal{X}_0, L_2(F))$, and since $\mathcal{X}_{1/p}$ and W_2^1 are obtained by the same interpolation functor applied to $(\mathcal{X}_0, \mathcal{X}_1)$ and $(L_2(F), W_2^p)$, respectively (namely the K-method interpolation functor $(\mathcal{X}_0, \mathcal{X}_1)_{1/p,2}$), then $K \in \mathcal{L}(\mathcal{X}_{1/p}, W_2^1)$, which proves (3.10). The details are given in Lemma 4.2 of Cox (1983). See also the introductions to interpolation theory in Triebel (1978), pages 18-27, or Butzer and Behrens (1967).

Q.E.D.

One problem we have not dealt with so far in this section is the explicit determination of the spaces $\mathcal{X}_\rho$, $\rho \geq 0$. Only for $\rho=1$ and $\rho=0$ is there an easily obtainable equivalent norm, namely

(3.11) $$\|h\|_1 \asymp \|h\|_{\mathcal{X}} \quad , \quad \|h\|_0 \asymp \|Kh\|_{L_2} \quad .$$

These equivalences follow from Lemma 2.1, the definition of U in (3.6), and equivalence of $L_2(F)$ and L_2 norms (Assumption II). Furthermore, as P and F are not involved in $\|h\|_{\mathcal{X}}$ and $\|Kh\|_{L_2}$, we have that different choices of P and F [subject to II, IV, and V(iii)], lead to equivalent $\mathcal{X}_\rho$ norms for $0\leq\rho\leq1$. For $\rho=0,1$, this follows from the above, while for $0<\rho<1$, this follows from the fact that $\mathcal{X}_\rho$ norm is equivalent to the norm on $(\mathcal{X}_0,\mathcal{X}_1)_{\rho,2}$ given by the K-method of interpolation, which is invariant up to norm equivalence. See the discussion after Lemma 2.3 of Cox (1983).

4. Periodic Deconvolution. In this section, we consider a particularly simple example of the theory presented in Section 3. Although periodic deconvolution is seldom used in practice, it is easy to see the main features of the theory because of the common Fourier basis:

$$\tilde{\psi}_1 \equiv 1$$

$$\left.\begin{aligned} \tilde{\psi}_{2j}(u) &= 2\cos(2\pi j u) \\ \tilde{\psi}_{2j+1}(u) &= 2\sin(2\pi j u) \end{aligned}\right] \qquad j=1,2,\ldots \quad .$$

As it is easier to work with complex exponentials, we use

$$\psi_j(u) = \exp[2\pi i j u] \quad , \quad -\infty<j<\infty$$

and require in all Fourier expansions

$$h = \sum_{j=-\infty}^{\infty} \langle h,\psi_j\rangle_{L_2}\psi_j$$

that $\langle h,\psi_j\rangle_{L_2} = \overline{\langle h,\psi_{-j}\rangle_{L_2}}$, so that the functions are real valued.

Suppose that the kernel of interest is

$$k(u,v) = \sum_{j=-\infty}^{\infty} w_j\psi_j(u-v) \equiv \kappa(u-v) \qquad , \tag{4.1}$$

so the integral operator K corresponds to periodic convolution with κ . We assume that for some positive constants a, b, c_1, c_2, with $b>1$,

$$c_1|j|^{-a} \le |w_j| \le c_2|j|^{-b} \quad , \qquad \forall j \in \mathbb{Z} \qquad . \tag{4.2}$$

Now suppose we observe data as in (1.1) with the errors mean 0, uncorrelated, and bounded variance uniformly in n . Then I holds with $s_n = 1/n$. Assume for now that the set of design points $\{u_{jn} = 1 \le j \le n\}$ is the regular grid

$$u_{jn} = (j-1)/n \qquad , \qquad 1 \le j \le n \tag{4.3}$$

so that the asymptotic design distribution in II is uniform

$$F' \equiv 1 \quad ,$$

and the convergence rate of F_n to F is

$$d_n = \frac{1}{n} \quad . \tag{4.4}$$

For q=0,1,2,... , consider the periodic Sobolev spaces

$$\mathscr{S}^{\nu} = \{h \in W_2^{\nu} : h^{(j)}(0) = h^{(j)}(1) \quad , \quad 0 \le j < \nu\} \quad ,$$

equipped with inner product

$$\langle h_1,h_2\rangle_{\mathscr{S}^\nu} = \left(\int h_1(u)du\right)\left(\int h_2(u)du\right) + \langle h_1^{(\nu)},h_2^{(\nu)}\rangle_{L_2} \quad .$$

Equivalently,

$$\mathscr{S}^\nu = \{h\in L_2 : \sum_{j=-\infty}^{\infty} (2\pi j)^{2\nu}|\langle h,\psi_j\rangle_{L_2}|^2<\infty\} \quad ,$$

(4.5)
$$\langle h_1,h_2\rangle_{\mathscr{S}^\nu} = \langle h_1,\psi_0\rangle_{L_2}\langle h_2,\psi_0\rangle_{L_2} + \sum_{j=-\infty}^{\infty} (2\pi j)^{2\nu}\langle h_1,\psi_j\rangle_{L_2}\overline{\langle h_2,\psi_j\rangle}_{L_2} \quad ,$$

and these definitions extend to arbitrary real $\nu\geq 0$. It is easy to check that $\mathscr{S}^\nu$ is an RKHS with kernel

$$r(u,v) = 1 + \sum_{\substack{j=-\infty \\ j\neq 0}}^{\infty} (2\pi j)^{-2\nu} \psi_j(u-v) \quad ,$$

which satisfies III provided $\nu>\frac{1}{2}$. Put

(4.6)
$$\mathscr{X} = \mathscr{S}^\nu \quad , \qquad \text{some integer} \quad \nu\geq 1 \quad .$$

For the operator P in the definition of the MOR estimator, we utilize

$$Ph = h - \int_0^1 h(u)dn$$

which is the projection on the subspace complementary to the constants. This gives

$$<h,Ph>_{\mathcal{X}} = \|h^{(\nu)}\|^2_{L_2}$$

and so $\mathcal{N}(P)$ is the constants, i.e., $m=\dim \mathcal{N}(P)=1$. Since (4.2) implies $w_0 \neq 0$, if h is constant and $K_n h = (w_0 h,\ldots,w_0 h)'=0$, then $h=0$, and thus IV holds. The assumed relation (4.2) also implies $w_j \neq 0$, $\forall_j$, and hence that $\mathcal{N}(K)=0$. As $\phi_j \in \mathcal{R}(K)$, $\forall_j$, assumption V(i) holds. Also,

$$\int_0^1 |k(u,v)|dv = \int_0^1 |\kappa(v)|dv \leq \sum_{j=-\infty}^{\infty} |w_j| < \infty$$

as b in (4.2) is assumed > 1 , and thus V(ii) holds. For V(iii), note that

$$\tilde{K}h = Kh = \sum_{j=-\infty}^{\infty} w_j <h,\psi_j>_{L_2} \psi_j$$

$$\tilde{K}^* h = \sum_j [1+(2\pi j)^{2\nu}]^{-1} \bar{w}_j <h,\psi_j>_{L_2} \psi_j$$

and hence

$$(P+\tilde{K}^*\tilde{K})h = |w_0|^2 <h,\psi_0>_{L_2} + \sum_{j\neq 0} \{1+|w_j|^2[1+(2\pi j)^{2\nu}]^{-1}\} <h,\psi_j>_{L_2} \psi_j$$

and hence

$$\|(P+\tilde{K}^*\tilde{K})^{-1}\|_{\mathcal{L}(\mathcal{X},\mathcal{X})} = \max\{|w_1|^{-2} \quad , \quad \sup_{j\neq 0}[|w_j|^2 + 1+(2\pi j)^{2\nu}]^{-1}\} < \infty$$

so that V(iii) holds.

One can check that the operator Q of (3.3) is given by

$$Qh = |w_0|^2 \langle h,\psi_0\rangle_{L_2} + \sum_{j\neq 0} |w_j|^2 (2\pi j)^{-2\nu} \langle h,\psi_j\rangle_{L_2} \psi_j$$

so the eigenvalues of (3.4) are given by

$$\mu_j = \begin{cases} |w_0|^2 & \text{if} \quad j=0 \\ |w_j|^2 (2\pi j)^{-2\nu} & \text{if} \quad j\neq 0 \end{cases}$$

and by (4.2),

$$|j|^{-2(\nu+a)} \lesssim \mu_j \lesssim |j|^{-2(\nu+b)} \tag{4.7}$$

and so VI holds with

$$r=2(\nu+b) \quad , \quad q=2(\nu+a) \quad . \tag{4.8}$$

If

$$p = \nu+b \tag{4.9}$$

then for $h\in\mathcal{X}$

$$\|Kh\|^2_{\mathcal{A}^p} = w_0^2 \langle h,\psi_0\rangle^2_{L_2} + \sum_{j\neq 0} (2\pi j)^{2p} |w_j|^2 \langle h,\psi_j\rangle^2_{L_2} \leq M^2 \|h\|^2_{\mathcal{A}^\nu}$$

where by (4.2)

$$M^2 = \max\{|w_0|^2 \quad , \quad \sup_{j\neq 0}(2\pi j)^{2(p-\nu)}|w_j|^2\} < \infty$$

and VII holds.

Before theorem 2.1 can be fruitfully applied to this problem, it is necessary to investigate the $\mathcal{X}_\rho$ norms. The eigenvectors and eigenvalues of (2.4) are

$$\phi_j = |w_j|^{-2}\,\psi_j$$

$$\gamma_j = (2\pi j)^{2\nu}|w_j|^{-2}$$

and so

$$\|h\|^2_{\mathcal{X}_\rho} = |w_0|^{-2}\langle h,\psi_0\rangle^2_{L_2} + \sum_{j\neq 0}(2\pi j)^{2\nu\rho}|w_j|^{-2\rho}\langle h,\psi_j\rangle^2_{L_2} \quad .$$

From (4.2) and (4.5), we see that

$$(4.10) \qquad c_1\|h\|_{\mathcal{J}_{\rho(\nu+b)}} \leq \|h\|_{\mathcal{X}_\rho} \leq c_2\|h\|_{\mathcal{J}_{\rho(\nu+a)}}$$

for some finite, positive constants c_1, c_2 .

By (4.8), the bounds (2.9) are

$$(4.11) \qquad \lambda^{-(\nu+b)\rho/(\nu+a)-1/[2(\nu+a)]}$$

$$\lesssim C(\lambda,\rho) \lesssim \lambda^{-\rho+1/[2(\nu+b)]} \quad , \quad \text{as} \quad \lambda\downarrow 0 \quad ,$$

valid for $\rho<2-1/[2(\nu+b)]$.

For this case, (2.11) gives

$$(4.12) \qquad \mu(\rho) = \frac{5}{4(\nu+b)} + \frac{(a-b)}{2(\nu+a)} \left\{\rho + \frac{1}{2(\nu+b)}\right\}$$

$$= \frac{5\nu+6a-b+2(a-b)(\nu+b)\rho}{4(\nu+a)(\nu+b)} \quad .$$

If we put

$$(4.13) \qquad \lambda_n = n^{1/\mu(\rho)} \log n \quad ,$$

then (2.16) holds.

We are now in a position to detail the conclusions of theorem 2.1 within the context of the foregoing discussion. For conclusions (a), (b), and (c), fix u such that

$$0\le u<2(\nu+b)-3/2 \quad .$$

(a) If g is a constant, then

$$(4.14) \qquad n^{-1}\{1+\lambda^{-[2u(\nu+b)+1]/[2(\nu+a)]}\}$$

$$\lesssim E\|\hat{g}_{n\lambda}-g\|^2_{\mathscr{A}_u} \lesssim n^{-1}\{1+\lambda^{-(2u-1)/[2(\nu+b)]}\} \quad .$$

(b) If $g\epsilon\mathscr{A}_w$ for some w satisfying

$$\left(\frac{\nu+a}{\nu+b}\right)u < w \leq 2(\nu+a) + \left(\frac{\nu+a}{\nu+b}\right)u \quad ,$$

$$\frac{\nu+a}{\nu+b} \leq w \leq 2(\nu+a) + \frac{\nu+a}{\nu+b} \quad ,$$

then

$$(4.15) \qquad E\|\hat{g}_{n\lambda}-g\|^2_{\mathscr{A}_u} \lesssim \min\{1,\lambda^{[w/(\nu+a)-u/(\nu+b)]}\}\|g\|^2_{\mathscr{A}_w}$$

$$n^{-1}\{1+\lambda^{-(2u-1)/[2(\nu+b)]}\} \quad .$$

Where both relations (4.14) and (4.15) hold uniformly in $\lambda\epsilon[\lambda_n,\infty]$, and (4.15) holds uniformly in $g\epsilon\mathscr{A}_w$.

(c) If we drop the assumption (4.3) on the evaluation points but merely retain the general Assumption II and the condition $d_n \approx 1/n$, then the above conclusions in (b) hold but under the additional restrictions that $u\leq(\nu+b)/(\nu+a)$ and $w\leq(\nu+b)/(\nu+a)$. This follows from the concluding remark of Section 3.

(d) If $g \in \mathcal{J}_u$ and $u \geq 1$, then

$$E\|\hat{g}_{n\lambda} - g\|^2_{\mathcal{J}_w} << 1$$

where

$$w = (\nu+a)u/(\nu+b) \quad .$$

(e) According to Corollary 2.4, if $g \in \mathcal{J}_w$ for some w satisfying

$$\frac{\nu+a}{\nu+b} \leq w \leq 2 \ ,$$

then

$$\frac{1}{n} E \sum_{j=1}^{n} \left(\int k((j-1)/n,u)[\hat{g}_{n\lambda}(u) - g(u)]du\right)^2$$

$$\lesssim \min\{1, \lambda^{w/(\nu+a)}\}\|g\|^2_w + s_n[m + C(\lambda,0)] \quad .$$

The expression on the left of this latter relation is appropriately called predictive integrated mean squared error, as it measures how well $\hat{g}_{n\lambda}$ can predict the values $\int k(u_{ni},u)g(u)du$.

5. Abel's Kernel. Theorem 1.1 is proved here by verifying the conditions III - VII of Section 3 and then applying Theorem 2.1. We also discuss the regularization of histograms to estimate particle size distribution in stereology.

For some $k\geq 2$, put

$$\mathcal{X} = W^k_{2,BC}$$

with inner product

$$<f,h>_{\mathcal{X}} = <f^{(k)},h^{(k)}>_{L_2} \quad .$$

A reproducing kernel is

$$r(u,v) = \frac{1}{(k!)^2} \int_0^1 (w-u)_+^{k-1} (v-w)_+^{k-1} dw$$

where $(x)_+ = [\max\{x,0\}]$. Now r is continuous on $[0,1]^2$. It will be shown $Rh = 0 \Rightarrow h = 0$ to show III holds. R is the Green's operator for the operator D^{2k} with domain $\{h\epsilon W_2^{2k}\cap\mathcal{X} : h^{(j)}(0) = 0 , j=k, k=1,\ldots,2k-1\}$. As the only solution of $D^{2k}h=0$, satisfying the 2k boundary conditions, is $h\equiv 0$, it follows from Theorem 2, page 31 of Naimark (1967) that $\mathcal{N}(R)=0$.

We take P=I , the identity on $\mathcal{X}$, and Condition IV holds.

Turning to V, some integral calculus shows

$$\int_0^1 (Kh)^2(u)du = 2\int_0^1\int_0^w h(v) \log[(\sqrt{w}+\sqrt{v})/(\sqrt{w}-\sqrt{v}) \ h(w)dvdw \quad ,$$

and hence $K \in \mathcal{L}(L_2, L_2)$ as

$$\|K\|^2_{\mathcal{L}(L_2,L_2)} \le 2\int_0^1 \int_0^w \log^2[(\sqrt{w}+\sqrt{v})/(\sqrt{w}-\sqrt{v})]dvdw < \infty \quad .$$

Also, one can derive the inversion formula

$$Kh=g \Leftrightarrow Kg \text{ absolutely continuous and } -\frac{1}{\pi} DKg = h \text{ a.e.}$$

(see e.g., page 7 of Cochran, 1972). Thus,

$$Kh=0 \Rightarrow h=0 \text{ a.e.} \Rightarrow \mathcal{N}(K) = \{0\} \quad .$$

It also follows from the inversion formula that

$$h_n(v) \equiv v^n \in \mathcal{R}(K) \qquad \text{for } n=1,2,\ldots \quad .$$

The uniform closure of $\{h_1, h_2, \ldots\}$ is $\{h \in C[0,1] : h(0) = 0\}$, which is dense in L_2, so V(i) holds. Also, V(ii) follows from a straightforward calculation, and V(iii) is immediate from our choice of P .

Before proceeding further, it will be useful to cite the relationship of K and R to fractional integration. For $\alpha>0$, put

$$(I_\alpha h)(u) = \frac{1}{\Gamma(\alpha)} \int_u^1 (v-u)^{\alpha-1} h(v)\,dv \quad .$$

Then $K = \pi^{\frac{1}{2}} I_{\frac{1}{2}}$, while $R = I_k\, I_k^*$, where $I_k \in \mathcal{L}(L_2, L_2)$. I_α extends α-times iterated integration to fractional orders (Ross, 1974). Further, one can show

$$(5.1) \qquad I_\alpha I_\beta = I_{\alpha+\beta} \quad , \qquad \alpha,\beta>0$$

and hence

$$(5.2) \qquad Q = KRK^* = \pi I_{k+\frac{1}{2}}\, I_{k+\frac{1}{2}}^* \quad .$$

This latter formula is now used to estimate the eigenvalues of Q .

Lemma 5.1. *The eigenvalues* $\{\mu_j\}$ *of* $I_{k+\frac{1}{2}}\, I_{k+\frac{1}{2}}^*$ *satisfy*

$$j^{-2(k+1)} \lesssim \mu_j \lesssim j^{-2k} \quad ,$$

and hence VI *holds with* $q=2(k+1)$ *and* $r=2k$.

Proof. Putting $M = \|I_{\frac{1}{2}}\|_{\mathcal{L}(L_2,L_2)}$ and using (5.1), one can show

$$(5.3) \qquad M^{-2}\|I_{k+1}h\|_{L_2}^2 \le \langle I_{k+\frac{1}{2}} I_{k+\frac{1}{2}}^*\, h,h\rangle_{L_2} \le M^2\|I_k h\|_{L_2}^2 \quad .$$

Let $\{\nu_j^{(i)}\}$ denote the eigenvalues of $I_i I_i^*$, then by the Mapping Principle (Theorem 3.6.1 of Weinberger, 1974), and the above inequalities,

$$M^{-2}\nu_j^{(k+1)} \le \mu_j \le M^2\nu_j^{(k)} \quad .$$

Now as already argued, $I_i I_i^*$ is the Green's operator for the differential operator D^{2i} with boundary conditions f

$$f^{(0)}(1) = f^{(1)}(1) = \ldots = f^{(i-1)}(1) = 0$$

and

$$f^{(i)}(0) = \ldots = f^{(2i-1)}(0) = 0 \quad .$$

The standard theory for the asymptotics of the eigenvalues of such operators (e.g., Section 4 of Naimark, 1967, or page 392 of Triebel, 1978) yields

$$\nu_j^{(i)} \approx j^{-2i} \quad .$$

The lemma is now evident.

Q.E.D.

Finally, we show VII holds with $p=k$. A k-fold integration by parts shows that for $h\in\mathcal{X}$,

$$D^k Kh = (-1)^k K(D^k h) \quad ,$$

and hence,

$$\|Kh\|_{W_2^k} = \|D^k Kh\|_{L_2} \leq M\|D^k h\|_{L_2} = M\|h\|_{\mathcal{X}} \quad , \tag{5.4}$$

where $M = \|K\|_{\mathcal{L}(L_2,L_2)}$.

Before applying Theorem 2.1 in order to prove Theorem 1.1, it is necessary to work out relations between the $\mathcal{X}_\rho$ norms and more readily computable norms. Thinking of K as a point map, we have the following picture:

$$\mathcal{X}_1 \xrightarrow{K} W_2^k$$

$$\mathcal{X}_0 \xrightarrow{K} L_2$$

where the top line follows from (5.4), and the bottom line is automatic, as in the proof of Theorem 3.2. Applying the K-method of interpolation, we obtain for $\rho\in[0,1]$

$$\mathcal{X}_\rho \xrightarrow{K} W_2^{k\rho} \quad ,$$

where $W_2^{k\rho}$ is a Besov space if $k\rho$ is not an integer. Consult Section 4.3 of Triebel (1978) for definitions and the relevant interpolation theorems. One may restrict attention to $\rho=i/k$ for $i=0,1,\ldots,k$ for simplicity. Hence, we have the bounds

$$(5.5) \qquad \|Kh\|_{W_2^{\rho k}} \lesssim \|h\|_{\mathcal{X}_\rho} \quad , \quad \forall h \in \mathcal{X}_\rho \quad ,$$

if $0 \le \rho \le 1$. Let $Jh=h$, then

$$W^{k}_{2,BC} \xrightarrow{J} \mathcal{X}_1$$

$$L_2 = W^0_{2,BC} \xrightarrow{J} \mathcal{X}_0$$

where the top line is immediate and the bottom line follows from the facts that $K : \mathcal{X}_0 \to L_2$ is an isomorphism and that $K : L_2 \to L_2$, so that

$$\|Jh\|_{\mathcal{X}_0} = \|h\|_{\mathcal{X}_0} = \|Kh\|_{L_2} \lesssim \|h\|_{L_2} \quad , \quad \forall h \in L_2 \quad .$$

Hence, an application of the K-method of interpolation gives

$$W^{\rho k}_{2,BC} \xrightarrow{J} \mathcal{X}_\rho \quad , \quad 0 \le \rho \le 1 \quad .$$

The fact that the K-method applied to the couple $(W^k_{2,BC}, L_2)$ gives $W^{k\rho}_{2,BC}$ is given in Section 4.3.3 of Triebel (1978). From this we have

$$(5.6) \qquad \|h\|_{\mathcal{X}_\rho} \lesssim \|h\|_{W_2^{\rho k}} \quad , \qquad \forall h \in W_{2,BC}^{\rho k} \quad ,$$

provided $0 \le \rho \le 1$.

Now we can apply Theorem 2.1 to obtain an extended version of Theorem 1.1. Firstly,

$$\mu(\rho) = \frac{1}{2(k+1)} (\rho+5/2+3/k)$$

which gives $(\rho+4)/6$ when $k=2$. Also, $m=0$, and Lemma 2.2 yields

$$C(\lambda,\rho) \lesssim \lambda^{-(\rho+1/2k)}$$

for $\rho \in [0,1]$, uniformly in $\lambda \in [0,\infty]$. Assuming

$$(5.7) \qquad \lambda_n \to 0 \quad , \quad d_n \lambda_n^{-(\rho+5/2+3/k)/2(k+1)} \to 0 \quad ,$$

we have by Theorem 2.1(b), (5.5), and (5.6) that if $0 \le \rho \le 1$ and $1/k \le t \le 1$, then

$$(5.8) \qquad E\|K(\hat{g}_{n\lambda}-g)\|^2_{W_2^{\rho k}} \lesssim E\|\hat{g}_{n\lambda}-g\|^2_{\mathcal{X}_\rho}$$

$$\lesssim \min\{1,\lambda^{(t-\rho)}\}\|g\|^2_{\mathcal{X}_t} + s_n C(\lambda,\rho)$$

$$\lesssim \min\{1,\lambda^{(t-\rho)}\}\|g\|^2_{W_2^{tk}} + s_n \lambda^{-(\rho+1/2k)}$$

uniformly in $g \in W_{2,BC}^{tk}$ and $\lambda \in [\lambda_n,\infty]$. Specializing to $k=2$ gives Theorem 1.1.

Theorem 2.1 also gives the following results. If g=0 , then by part (a) of the theorem,

$$E\|K\hat{g}_n\|^2_{W_2^{\rho k}} \leq s_n \lambda^{-(\rho+1/2k)}$$

and

$$E\|\hat{g}_{n\lambda}\|^2_{W_2^{\rho k}} \gtrsim s_n \min\{\lambda^{-2}, \lambda^{-(2k\rho+1)/2(k+1)}\} \quad ,$$

where the latter relation follows from Lemma 2.2 and (5.6), and both of the above are uniform in $\lambda\epsilon[\lambda_n,\infty]$. Part (c) of the theorem allows the following sharper conclusion when $\rho=t$ in (5.8):

$$E\|K(\hat{g}_{n\lambda}-g)\|^2_{W_2^{\rho k}} \lesssim \min\{1,o(1)\} + s_n\lambda^{-(\rho+1/2k)} \quad .$$

Finally, the lower bound in part (d) requires

$$\lambda_n \to 0 \quad , \quad d_n\lambda_n^{-(1+5/4k)} \to 0 \quad ,$$

which is always stronger than (5.7), and we obtain

$$E\|\hat{g}_{n\lambda}-g\|^2_{W_2^{\rho k}} \gtrsim \min\{1,\lambda^2\}\|Kg\|^2_{W_2^{tk}} + s_n \min\{\lambda^{-2}, \lambda^{-(2k\rho+1)/2(k+1)}\} \quad ,$$

uniformly in $\lambda\epsilon[\lambda_n,\infty]$ and g such that $Kg\epsilon W^{tk}_{2,BC}$ (e.g., $g\epsilon W^{tk-1}_{2,BC}$) .

We now describe our proposed estimator for the stereology problem and give upper bounds on its convergence rate. We suppose that the centers of the spheres are given by a three-dimensional Poisson process of rate Λ_3 per unit volume, and that the radius R of a sphere is a random variable independent of the centers of the spheres and the radii of the other spheres. Letting $A_3 = \pi R^2$ denote the equatorial area, we assume A_3 has Lebesque density $f_3(a)$, $0<a<1$ (any other upper bound can be accommodated by rescaling). Let

$$h_3(a) = \Lambda_3 f_3(a) \quad .$$

Note that the number of spheres with equitorial area between a_0 and a_1 $(0\le a_0<a_1\le 1)$ centered in a unit volume is a Poisson random variable with mean

$$\int_{a_0}^{a_1} h_3(a)\, da \quad .$$

Now suppose a random planar region of unit area is selected independently of the sphere process. Then the number of circular cross sections of spheres with area between a_0 and a_1 $(0\le a_0<a_1\le 1)$ is a Poisson random variable with mean

$$\int_{a_0}^{a_1} h_2(a)\, da \quad ,$$

where

$$h_2(a) = 2(Kh_3)(a) \quad ,$$

K being Abel's kernel operator in (1.3). This latter equation follows from a little variable changing in equation (5) of Watson (1971). Now put

$$g = (2h_3) \quad .$$

Suppose that a cross sectional planar slice of area b is taken and we observe spherical cross sections of areas $A_1, A_2, \ldots, A_N$. Let $0 = u_{0n} < u_{1n} < \ldots < u_{nn} = 1$ be a partition of [0,1] and put

$$y_{in} = \left[b(u_{in}-u_{i-1,n})\right]^{-1} \sum_{j=1}^{N} I\left[(u_{i-1,n},u_{in})\right.(A_j) \quad ,$$

for $1 \le i \le n$. By the mean value theorem for integrals,

$$Ey_{in} = (u_{in}-u_{i-1,n})^{-1} \int_{u_{i-1,n}}^{u_{in}} h_2(a)\, da$$

$$= h_2(u_{in}^*)$$

$$= (Kg)(u_{in}^*)$$

where

$$u_{i-1,n} \le u_{in}^{*} \le u_{in} \quad .$$

These latter inequalities guarantee that if F_n and F_n^* are the empirical c.d.f.'s of $\{u_{in} : 0 \le i \le n\}$ and $\{u_{in}^* : 1 \le i \le n\}$, respectively, then $\sup|F_n - F_n^*| \lesssim n^{-1}$, so if $F_n \to F$ where F is absolutely continuous, then $F_n^* \to F$ and $d_n^* = \sup|F_n^* - F| \asymp d_n$ since $d_n \gtrsim 1/n$ always. If we assume that F_n satisfies the requirements of II, then so also will F_n^* (with the same d_n) . Now let $\varepsilon_{in} = y_{in} - Ey_{in}$, and then

$$E|\frac{1}{n}\underset{\sim}{\eta}'\underset{\sim}{\varepsilon}_n|^2 = \frac{1}{n^2}\sum_{j=1} \eta_j^2 \operatorname{Var}(y_{in})$$

$$= \frac{1}{n^2}\sum_{j=1}^{n} \eta_j^2 \left[b(u_{in}-u_{i-1,n})\right]^{-2} b\int_{u_{i-1,n}}^{u_{i,n}} h_2(a)da$$

$$\le \left[nb \min_{1\le i\le n} (u_{in}-u_{i-1,n})\right]^{-1} \sup_{0\le u\le 1} |(Kg)(u)| \left(\frac{1}{n}\sum_{j=1}^{n} \eta_j^2\right) \quad .$$

As already noted above, $Kg \in W_2^1$ if $g \in W_{2,BC}^1$, and hence $\sup|(Kg)| < \infty$ in this case. Thus,

$$E|\frac{1}{n}\underset{\sim}{\eta}'\underset{\sim}{\varepsilon}_n|^2 \le s_n\left(\frac{1}{n}\sum_{j=1}^{n} \eta_j^2\right) \quad , \qquad \forall \eta \in \mathbb{R}^n \tag{5.9}$$

where

(5.10) $$s_n = \left[nb \min_{1\le i\le n} (u_{in}-u_{i-1,n})\right]^{-1} \quad .$$

We have then that I and II hold, except that we do not obtain a lower bound on variance in I. This is only needed to get the lower bounds on $E\|\hat{g}_{n\lambda}-g\|_\rho^2$ in Theorem 2.1, parts (a) and (d). In particular, the conclusions of Theorem 1.1 hold. If

$$\max_{1\le i\le n}\left(u_{i,n}-u_{i-1,n}\right) \lesssim \min_{1\le i\le n}\left(u_{i,n}-u_{i-1,n}\right) \quad , \quad n\to\infty$$

then

$$d_n \asymp n^{-1} \qquad , \qquad s_n \asymp b^{-1} = b_n^{-1}$$

and so if $g\in W_{2,BC}^t$, then

$$E\|K(\hat{g}_{n\lambda}-g)\|_{W_2^{\rho k}}^2 \le \lambda^{(t-\rho)} + b_n^{-1}\lambda^{-(\rho+1/2k)}$$

as $n\to\infty$, $b_n\to\infty$, and $\lambda\to 0$, provided that $n^{-1}\lambda^{-(\rho+5/2+3/k)/2(k+1)} \to 0$ and $0\le\rho\le t\le 1$.

Typically, one is interested in the density f_3 of spherical radii. If g is normalized to integrate to 1 , then an estimate of f_3 can be obtained from a simple change of variables.

REFERENCES

ANDERSSEN, R. S. (1976). Stable procedures for the inversion of Abel's equation. *Inst. Maths. Applics.* 17 329-342.

ANDERSSEN, R., and JAKEMAN, A. (1975a). Product integration for functionals of particle size distribution. *Utilitas Mathematica* 8.

ANDERSSEN, R., and JAKEMAN, A. (1975b). Abel type integral equations in stereology, II. Computational methods of solution and the random spheres approximation. *J. Microscopy* 105 135-153.

ARONSZAJN, N. (1950). Theory of reproducing kernels. *Trans. Am. Math. Soc.* 68 337-404.

AUBIN, J. (1979). *Applied Functional Analysis.* John Wiley and Sons, New York.

BULLEN, H. E. (1963). *An Introduction to Seismology.* Cambridge University Press, Cambridge.

BUTZER, P., and BEHRENS, H. (1967). *Semigroups of Operators and Approximation.* Springer-Verlag, New York.

COCHRAN, J. A. (1972). *The Analysis of Linear Integral Equations.* McGraw-Hill, New York.

COX, D. (1983). Approximation of method of regularization estimators. Tech. Report 723, Dept. of Statistics, University of Wisconsin, Madison. Submitted to *Ann. Statist.*

FRANKLIN, J. (1981). Confidence limits for stereological estimators with infinite variance. *SIAM J. Applied Math.* 40 179-190.

GROETSCH, C. (1971). *Generalized Inverses of Linear Operators.* Marcel Dekker, New York.

KIMELDORF, G., and WAHBA, G. (1970). A correspondence between Bayesian estimation on stochastic processes and smoothing by splines. *Ann. Math. Statist.* 41 492-502.

KOSAREV, E. L. (1980). Applications of integral equations of the first kind in experimental physics. *Computer Physics Communications* 20 69-75.

LI, K. C. (1982). Minimaxity of the method of regularization on stochastic processes. *Ann. Statist.* 10 937-942.

LIONS, J., and MAGENES, E. (1973). *Nonhomogeneous Boundary Value Problems and Applications.* Springer-Verlag, New York.

LUKAS, M. (1980). Regularization. *Numerical Solution of Integral Equations.* 152-182 (R. Anderssen, ed.). Sijthoff and Noordoff.

LUKAS, M. (1981). Regularization of linear operator equations. Unpublished Ph.D. thesis, Australian National University.

NAIMARK, M. A. (1967). *Linear Differential Operators, Part I.* Frederick Ungar, New York.

NASHED, M., and WAHBA, G. (1974). Generalized inverses in reproducing kernel spaces: an approach to regularization of linear operator equations. *SIAM J. Math. Anal.* 5 974-987.

NYCHKA, D. (1983). The solution of Abel type integral equations with an application in stereology. Unpublished Ph.D. dissertation, University of Wisconsin, Madison.

RIESZ, F., and SZ.-NAGY, B. (1955). *Functional Analysis.* Frederick Ungar, New York.

ROSS, B. (1974). Fractional calculus and its applications. *Proc. of the International Conference at University of New Haven*. Springer-Verlag, New York.

SPECKMAN, P. (1979). Minimax estimates of linear functionals in a Hilbert space. Unpublished.

TIKHONOV, A. N. (1963). Solution of incorrectly posed problems and the method of regularization. *Soviet Math. Dokl.* 4 1035-1038.

TRIEBEL, H. (1978). *Interpolation Theory, Function Spaces, Differential Operators*. North Holland, New York.

WAHBA, G. (1973). Convergence rates of certain approximate solutions to Fredholm integral equations of the first kind. *J. Approx. Theory* 7 167-185.

WAHBA, G. (1977). Practical approximate solutions to linear operator equations when the data are noisy. *SIAM J. Numer. Anal.* 14 651-667.

WAHBA, G. (1980). Ill posed problems: numerical and statistical methods for mildly, moderately, and severely ill posed problems with noisy data. Tech. Report 595, Dept. of Statistics, University of Wisconsin, Madison.

WATSON, G. S. (1971). Estimating functionals of particle size distributions. *Biometrika* 58 483-490.

WEINBERGER, H. (1974). *Variational Methods for Eigenvalue Approximation*. J. W. Arrowsmith Ltd., Bristol.