The Annals of Statistics
1988, Vol. 16, No. 2, 694–712

APPROXIMATION OF METHOD OF REGULARIZATION ESTIMATORS[1]

By Dennis D. Cox

University of Illinois

The Tikhonov method of regularization (MOR) estimator provides a general method for estimation of a nonparametric regression parameter in an abstract linear model with discrete noisy data. An asymptotic analysis is given in which the discrete estimation problem is approximated by a continuous one. Rates of convergence are calculated in a family of norms natural to the problem. The general theory is applied to the estimation of functions from noisy evaluations of the function and one of its derivatives.

0. Introduction. Consider a sequence of statistical experiments with observation given by an "abstract" linear model

$$y_n = T_n\xi + \varepsilon_n. \tag{0.1}$$

Here, y_n nominally lies in a real separable Hilbert space $\mathscr{Y}_n$, the unknown regression parameter ξ nominally lies in a real separable Hilbert space $\mathscr{X}$ and the design operator T_n belongs to the space of bounded linear operators from $\mathscr{X}$ to $\mathscr{Y}_n$, denoted $\mathscr{B}(\mathscr{X}, \mathscr{Y}_n)$. The "noise" or "error" term ε_n in (0.1) is nominally a $\mathscr{Y}_n$-valued random vector. (The three appearances of the word "nominally" will be explained shortly.) In this paper, we investigate a class of widely used linear estimators for ξ called method of regularization (MOR) estimators. Our goal is to obtain useful approximations to the first and second moments of the norm of the estimation error. We consider some examples.

Example 0.1 (Finite-dimensional parameter). Suppose $\mathscr{Y}_n = R^n$, $\mathscr{X} = R^p$, T_n is represented by an $n \times p$ matrix and ε_n is mean 0 with covariance $\sigma_n^2 I$, where I always denotes an identity operator. The MOR estimator $\xi_{n\lambda}$ is obtained by minimization over $x \in \mathscr{X}$ of

$$L_{n\lambda}(x) = \lambda\langle x, Wx\rangle_{\mathscr{X}} + \|y_n - T_n x\|^2_{\mathscr{Y}_n}, \tag{0.2}$$

where $W \in \mathscr{B}(\mathscr{X}) = \mathscr{B}(\mathscr{X}, \mathscr{X})$ is a positive operator (called the regularization operator), and $\lambda \in [0, \infty]$ is the regularization parameter. "W is a positive operator" means: (i) W is self-adjoint (i.e., represented by a symmetric matrix); and (ii) $\langle x, Wx\rangle_{\mathscr{X}} \geq 0$ for all $x \in \mathscr{X}$. Put $U_n = T_n^* T_n$, $G_{n\lambda} = (\lambda W + U_n)^{-1}$, where

Received May 1984; revised September 1987.
[1]Research supported by the National Science Foundation under Grants MCS-82-02560 and DMS-86-03083.
AMS 1980 *subject classifications.* Primary 62G05; secondary 41A25, 47A50, 41–00.
Key words and phrases. Nonparametric regression, method of regularization, smoothing splines, rates of convergence.

T_n^*: $\mathscr{Y}_n \to \mathscr{X}$ is the adjoint of T_n: $\mathscr{X} \to \mathscr{Y}_n$ [18] (i.e., transpose).

Assuming $G_{n\lambda}^{-1}$ exists,

$$\xi_{n\lambda} = G_{n\lambda}^{-1} T_n^* y_n, \qquad 0 < \lambda < \infty. \tag{0.3}$$

Also, $\xi_{n0} = T_n^+ y_n$ is the ordinary least-squares estimator of ξ of minimal norm [T_n^+ is the Moore–Penrose generalized inverse of T_n ([16] and [9])], and $\xi_{n\infty}$ is the minimal norm least-squares estimator of ξ restricted to $\mathscr{N}(W) = \{\zeta \in \mathscr{X}: W\zeta = 0\}$, the null space of W. With these conventions, $\xi_{n\lambda}$ is defined continuously in λ. One may be led to such an estimator via Bayesian methods under a mean zero prior on ξ with covariance $(\sigma_n^2 \lambda W)^{-1}$. Alternatively, such estimators arise in ridge regression [10] and as the solution of certain minimax problem [20]. The bias of $\xi_{n\lambda}$ is given by

$$B_{n\lambda}\xi = \left(G_{n\lambda}^{-1} U_n - I\right)\xi. \tag{0.4}$$

The covariance is $\sigma_n^2 V_{n\lambda}$, where

$$V_{n\lambda} = G_{n\lambda}^{-1} U_n G_{n\lambda}^{-1}. \tag{0.5}$$

In the following discussion we indicate a simple approach for obtaining useful approximations to $B_{n\lambda}$ and $V_{n\lambda}$ as $n \to \infty$.

EXAMPLE 0.2 (Smoothing splines). Let m be a positive integer and let $\mathscr{X}$ be the Sobolev space $W_2^m = W_2^m[0,1]$ of functions on $[0,1]$ with m derivatives in L_2 equipped with the inner product

$$\langle g, h\rangle_{W_2^m} = \langle g^{(m)}, h^{(m)}\rangle_{L_2} + \langle g, h\rangle_{L_2}.$$

It can be shown [11] that the evaluation functional mapping $x \to x(t)$ is continuous for each $t \in [0,1]$. Thus, there is a q: $[0,1] \to \mathscr{X}$ such that

$$\left\langle q(t), x\right\rangle_{\mathscr{X}} = x(t), \qquad \forall\, x \in \mathscr{X}, \forall\, t \in [0,1].$$

This is the "reproducing kernel Hilbert space" property of W_2^m. For each n, let $t_{ni} = i/n$, $1 \le i \le n$. Then the operator T_n given by $T_n\xi = (\xi(t_{n1}), \dots, \xi(t_{nn}))$ is continuous from $\mathscr{X}$ to R^n. Take $\mathscr{Y}_n = R^n$ but with inner product $\langle y, \eta\rangle_{\mathscr{Y}_n} = n^{-1}\langle y, \eta\rangle_{R^n} = n^{-1}\sum_{i=1}^n y_i \eta_i$.

Let $\varepsilon_n = (\varepsilon_{n1}, \dots, \varepsilon_{nn})$ have mean 0 and covariance $\sigma^2 I$, and suppose we observe y_n with components $y_{ni} = \xi(t_{ni}) + \varepsilon_{ni}$, $1 \le i \le n$, where ξ is thought to lie in $\mathscr{X} = W_2^m$. The smoothing spline estimate of ξ is given by minimization over $x \in \mathscr{X}$ of

$$L_{n\lambda}(x) = \lambda \int_0^1 [x^{(m)}(t)]^2\, dt + n^{-1} \sum_{i=1} [y_{ni} - x(t_{ni})]^2. \tag{0.6}$$

Let W be defined by $\langle x, Wx\rangle_{\mathscr{X}} = \|x^{(m)}\|_{L_2}^2$; then with our choice of $\langle\cdot,\cdot\rangle_{\mathscr{Y}_n}$, (0.6) is the same as (0.2) and smoothing spline is given by (0.3).

One new phenomenon, which arises when $\dim \mathscr{X} = \infty$, is that $\xi_{n\lambda}$ may be consistent for ξ even if $\xi \notin \mathscr{X}$. For instance, if $\mathscr{X} = W_2^2$ but ξ is only in W_2^1, then we may still have $E\|\xi_{n\lambda} - \xi\|_{L_2}^2 \to 0$ as $n \to \infty$. This is the reason we say ξ "nominally" is in $\mathscr{X}$.

EXAMPLE 0.3 (Continuous time smoothing spline). Let $\mathscr{X}$ be as in Example 0.2, and let $\varepsilon_n(t)$ be a Gaussian white-noise process of intensity σ^2/n for some $\sigma^2 > 0$. This is obtained formally by differentiation of $\sigma n^{-1/2}B(t)$, where $B(t)$ denotes a standard Wiener process on $[0,1]$. Technically speaking, ε_n does not "live" in $L_2[0,1]$, but a "stochastic inner product" [12] like $\langle g, \varepsilon_n\rangle_{L_2} = \int_0^1 g(t)\varepsilon_n(t)\,dt$, $g \in L_2$, can be defined as the stochastic integral $\sigma n^{-1/2}\int_0^1 g(t)\,dB(t)$. For our purposes, it is sufficient to know that

$$E\langle g, \varepsilon_n\rangle_{L_2} = 0, \qquad \forall\, g \in L_2,$$

$$E\langle g, \varepsilon_n\rangle_{L_2}\langle \varepsilon_n, h\rangle_{L_2} = n^{-1}\sigma^2\langle g, h\rangle_{L_2}, \qquad \forall\, g, h \in L_2.$$

In effect, ε_n has mean 0 and covariance $\sigma^2 n^{-1}I$, where I is the identity on L_2. Suppose we observe $y_n(t) = \xi(t) + \varepsilon_n(t)$, $0 \le t \le 1$. Let $\mathscr{Y}_n = \mathscr{Y} = L_2[0,1]$ for all n. (This is the "nominal" observation space as ε_n and y_n are almost surely not in $\mathscr{Y}_n$.) Let W be the same as Example 0.2, and let $T_n = T$ for all n be the imbedding of W_2^m into L_2 (i.e., $Tx = x \in L_2$ for all $x \in W_2^m$). Consider the objective

$$L_{n\lambda}(x) = \lambda\langle x, Wx\rangle_{\mathscr{X}} + \|Tx\|^2_{\mathscr{Y}_n} - 2\langle y_n, T_n x\rangle_{\mathscr{Y}_n}, \tag{0.7}$$

which would have the same minimizer as (0.2) in the previous examples. However, (0.2) is not defined here.

Estimators obtained by minimization of (0.7) are solutions of certain minimaxity problems [13]. One can show that $\xi_{n\lambda}$ is given again by (0.3). The results of this article show that the first and second moments of the estimation error in Examples 0.2 and 0.3 behave similarly as $n \to \infty$ and $\lambda \to 0$ not too fast.

EXAMPLE 0.4 (Smoothing splines with derivative data). Let p be a positive integer and $\mathscr{X} = W_2^m$ for some $m > p$. Let $(t_{n1}, \dots, t_{nn}) \in [0,1]^n$ and suppose we observe

$$y_{n1}(i) = \xi(t_{ni}) + \varepsilon_{n1}(i),$$
$$y_{n2}(i) = \xi^{(p)}(t_{ni}) + \varepsilon_{n2}(i), \qquad 1 \le i \le n,$$

where $\varepsilon_{n1}(1), \dots, \varepsilon_{n1}(n), \varepsilon_{n2}(1), \dots, \varepsilon_{n2}(n)$ are mean zero uncorrelated random variables with $\operatorname{Var}\varepsilon_{n1}(i) = \sigma^2$, $\operatorname{Var}\varepsilon_{n2}(i) = b^{-1}\sigma^2$, $1 \le i \le n$, for some positive b and σ^2. See Schwarz [19], page 413, where a higher-dimensional version of this example arises in a geodetic context. We take $\mathscr{Y}_n = R^n \times R^n$ equipped with the inner product $\langle(\eta_1, \eta_2), (\xi_1, \xi_2)\rangle_{\mathscr{Y}_n} = n^{-1}[\langle \eta_1, \xi_1\rangle_{R^n} + b\langle \eta_2, \xi_2\rangle_{R^n}]$. One can show that the map T_n given by $T_n\xi = (\xi(t_{n1}), \dots, \xi(t_{nn}), \xi^{(p)}(t_{n1}), \dots, \xi^{(p)}(t_{nn}))$ is in $\mathscr{B}(\mathscr{X}, \mathscr{Y}_n)$ as $m > p$. With these specifications of $\mathscr{X}$, $\mathscr{Y}_n$ and T_n, an MOR estimate of ξ can be obtained by minimization of (0.2) as before.

EXAMPLE 0.5 (Regularized solutions of integral equations). Let $\mathscr{X}, \mathscr{Y}_n$, $\{t_{ni}\}$ and ε_n be as in Example 0.2. Suppose $K(s,t)$ is continuous on $[0,1]^2$ and we observe $y_{ni} = \int_0^1 K(t_{ni}, s)\xi(s)\,ds + \varepsilon_{ni}$, $1 \le i \le n$. The inversion of the integral equation (i.e., estimation of ξ) from such data is known as an "ill-posed" problem. The terminology "method of regularization" was first coined in this setting [23].

We now give an approach to asymptotics for Example 0.1. There are two fundamental difficulties with the analysis as $n \to \infty$. One is specifying the dependence of λ on n. That is avoided here by giving approximations that are uniform in λ. (It is assumed that λ is not random.) The second is specifying the dependence of T_n on n. We assume there is a positive sequence $\{a_n\}$ and a fixed operator $U \in \mathscr{B}(\mathscr{X})$ such that $a_n T_n^* T_n \to U$, where the limit may be taken in any of the usual senses when $\dim \mathscr{X} < \infty$.

By rescaling the model (0.1), we may assume $a_n \equiv 1$. Assume U is nonsingular and hence also $U_n = T_n^* T_n$ for all n sufficiently large. Then we show $B_{n\lambda}$ and $V_{n\lambda}$ of (0.4) and (0.5) can be approximated by

$$B_\lambda = G_\lambda^{-1}U - I, \qquad V_\lambda = G_\lambda^{-1}UG_\lambda^{-1}, \tag{0.8}$$

where $G_\lambda = (\lambda W + U)$, B_λ and V_λ are the "continuous" analogs of the "discrete" operators $B_{n\lambda}$ and $V_{n\lambda}$.

For this setup (with $\dim \mathscr{X} < \infty$), the continuous operators are easy to analyze. Put $A = U^{-1}W$; then

$$G_\lambda^{-1}U = (\lambda W + U)^{-1}U = I - \lambda A + (\lambda A)^2 + \cdots . \tag{0.9}$$

Hence, as $\lambda \to 0$, $\|B_\lambda\|_{\mathscr{B}(\mathscr{X})} = O(\lambda)$, where $\|\cdot\|_{\mathscr{B}(\mathscr{X})}$ denotes the usual uniform operator norm on $\mathscr{B}(\mathscr{X})$. Also, trace $V_\lambda = (\operatorname{trace} U^{-1})(1 + o(1))$, since $G_\lambda^{-1} \to U^{-1}$ as $\lambda \to 0$.

Now define the perturbation operator

$$R_{n\lambda} = G_\lambda^{-1}(U - U_n). \tag{0.10}$$

Note that

$$\|G_\lambda^{-1}\|_{\mathscr{B}(\mathscr{X})} \le \left\|(I + \lambda A)^{-1}\right\|_{\mathscr{B}(\mathscr{X})}\|U^{-1}\|_{\mathscr{B}(\mathscr{X})} \le \|U^{-1}\|_{\mathscr{B}(\mathscr{X})},$$

so $\|R_{n\lambda}\|_{\mathscr{B}(\mathscr{X})} \to 0$ as $n \to \infty$, uniformly in $\lambda \in [0, \infty]$. One can show

$$B_{n\lambda} = B_\lambda + R_{n\lambda}B_{n\lambda}. \tag{0.11}$$

From this it follows that as $n \to \infty$, $\|(B_{n\lambda} - B_\lambda)\|_{\mathscr{B}(\mathscr{X})} = o(1)\|B_{n\lambda}\|_{\mathscr{B}(\mathscr{X})}$, where the $o(1)$ quantity is independent of λ. It also follows from (0.11) that $\|B_{n\lambda}\|_{\mathscr{B}(\mathscr{X})} = O(\lambda)$ as $n \to \infty$ and $\lambda \to \infty$. Similarly, trace $V_{n\lambda} = (\operatorname{trace} V_\lambda)(1 + o(1))$, $n \to \infty$, uniformly in λ. Combining the preceding results gives

$$\begin{aligned} E\|\xi_{n\lambda} - \xi\|_{\mathscr{X}}^2 &= \|B_{n\lambda}\xi\|_{\mathscr{X}}^2 + \sigma_n^2 \operatorname{trace} V_{n\lambda} \\ &= O(1 \vee \lambda^2)\|\xi\|_{\mathscr{X}}^2 + \sigma_n^2(\operatorname{trace} U^{-1})(1 + o(1)), \end{aligned}$$

as $n \to \infty$, uniformly in λ and ξ. Here $a \vee b = \max\{a, b\}$.

This sort of analysis becomes more tricky when $\dim \mathscr{X} = \infty$, for reasons discussed at greater length in Section 1. In that section, it is also indicated how one obtains a continuous analog for the discrete problem of Example 0.4. Just as in the $\dim \mathscr{X} < \infty$ case, the continuous problem is analyzed first, in Section 2. One of the problems when $\dim \mathscr{X} = \infty$ is the multiplicity of norms (e.g., W_2^r for different r). In the abstract setting of Section 2, there is a natural family of norms for which it is easy to calculate the mean squared error. Unfortunately, it is rather difficult to make sense of these norms in a concrete setting. Some techniques for doing this are discussed in Section 3. Section 4 contains the main contribution of the paper: The general convergence theory developed for the continuous case is carried over to the discrete problem. In Section 2–4, the abstract theory is developed and then specialized to Example 0.4 and its continuous analog given in Section 1. Application to examples of the type 0.5 will be given in a subsequent paper [5]. Further applications of the mathematical techniques will appear in [4] and [6]. Related results appear in [21], [17] and [14].

1. The limiting continuous problem. Let us now return to the setup of Example 0.4 and seek a corresponding continuous problem. We look for a limiting version U of the operators U_n. To identify T_n^*, let $x \in \mathscr{X}$ and $\eta = (\eta_1, \eta_2) \in \mathscr{Y}_n$, and then

$$\begin{aligned}\langle T_n x, (\eta_1, \eta_2)\rangle_{\mathscr{Y}_n} &= n^{-1}\left[\sum_{i=1}^{n} \langle x, q(t_{ni})\rangle_{\mathscr{X}} \eta_1(i) + b \sum_{i=1}^{n} \langle x, q^{(p)}(t_{ni})\rangle_{\mathscr{X}} \eta_2(i)\right] \\ &= \left\langle x, \left\{n^{-1} \sum_{i=1}^{n} \left[q(t_{ni})\eta_1(i) + bq^{(p)}(t_{ni})\eta_2(i)\right]\right\}\right\rangle_{\mathscr{X}} \\ &= \langle x, T_n^* \eta\rangle_{\mathscr{X}}.\end{aligned}$$

Thus, $T_n^* \eta$ is the quantity in braces, and

$$\langle x, U_n \zeta\rangle_{\mathscr{X}} = \int \left[x(t)\zeta(t) + bx^{(p)}(t)\zeta^{(p)}(t)\right] dF_n(t), \tag{1.1}$$

where $F_n(t) = n^{-1}\#\{i\colon\ t_{ni} \le t\}$. Let us suppose that the $F_n \to F$ uniformly, where F has a smooth density $f \approx 1$. (The notation $\approx$ means the l.h.s. can be bounded above and below by positive, finite constant multiples of the r.h.s.) Let $\mathscr{Y} = L_2(F) \oplus L_2(bF)$ and $T\colon \mathscr{X} \to \mathscr{Y}$ the map $Tx = (x, x^{(p)}) \in \mathscr{Y}$. Then $U = T^*T$ is given by

$$Ux = \int \left[x(t)q(t) + bx^{(p)}(t)q^{(p)}(t)\right] dF(t). \tag{1.2}$$

This is an $\mathscr{X}$-valued integral as $q\colon [0,1] \to \mathscr{X}$ (see Example 0.2). We now relate this operator to a statistical problem as in equation (0.1).

EXAMPLE 1.1 (Continuous analog of Example 0.4). Let $y_n = (y_{n1}, y_{n2})$ be given by

$$\begin{aligned} y_{n1}(t) &= \xi(t) + \varepsilon_{n1}(t), \\ y_{n2}(t) &= \xi^{(p)}(t) + \varepsilon_{n2}(t), \qquad 0 \le t \le 1, \end{aligned}$$

where the error vector $\varepsilon_n = (\varepsilon_{n1}, \varepsilon_{n2})$ satisfies

$$\varepsilon_{n1}(t) = \sigma n^{-1/2} f^{-1/2}(t)\omega_1(t), \qquad \varepsilon_{n2}(t) = \sigma b^{-1/2} n^{-1/2} f^{-1/2}(t)\omega_2(t).$$

Here, ω_1 and ω_2 are independent unit intensity Gaussian white noises. One may check that $E\langle \eta, \varepsilon_n\rangle_{\mathscr{Y}} = 0$ and $E\langle \eta, \varepsilon_n\rangle_{\mathscr{Y}}\langle \zeta, \varepsilon_n\rangle_{\mathscr{Y}} = n^{-1}\sigma^2\langle \eta, \zeta\rangle_{\mathscr{Y}}$, for $\eta, \zeta \in \mathscr{Y}$, i.e., ε_n has mean 0 and covariance $\sigma_n^2 I$ on $\mathscr{Y}$ with $\sigma_n^2 = \sigma^2 n^{-1}$. Also, the design operator T is as given previously.

The asymptotic analysis for Example 0.1 ($\dim \mathscr{X} < \infty$) relied heavily on boundedness of U^{-1}. One can show that U in (1.2) has range $\mathscr{R}(U) = \{x \in W_2^{2m-p}\colon D^j x(t) = 0 \text{ at } t = 0 \text{ and } 1,\ m \le j \le 2m - p\}$, which is a proper dense subspace of $\mathscr{X}$.

It follows from the open mapping theorem that U^{-1} is not a bounded operator. Thus the asymptotics for $\dim \mathscr{X} < \infty$ do not extend to $\dim \mathscr{X} = \infty$. The unboundedness of U^{-1} is the reason for the instability or "ill-posedness" of our estimation problem. See Engl [8] for further discussion.

2. Spectral analysis of continuous regularization. In this section we consider the "continuous" case when $\mathscr{Y}_n = \mathscr{Y}$ and $T_n = T$ for all n. Referring back to Example 0.1, a more thorough analysis of B_λ and V_λ can be accomplished via "simultaneous" diagonalization of the operators U and W. Using the theory in Rao [16], page 41, there is a basis $\{\phi_\nu\}$ for $\mathscr{X} = R^p$ such that for all pairs of indices (ν, μ),

$$\langle \phi_\nu, U\phi_\mu \rangle_{\mathscr{X}} = \delta_{\nu\mu}, \qquad \langle \phi_\nu, W\phi_\mu \rangle_{\mathscr{X}} = \gamma_\nu \delta_{\nu\mu}, \tag{2.1}$$

where $\{\gamma_\nu\}$ are the associated eigenvalues of W w.r.t. U, and $\delta_{\nu\mu}$ denotes Kronecker's delta. It follows that $\langle(\lambda W + U)x, \phi_\nu\rangle_{\mathscr{X}} = (1 + \lambda\gamma_\nu)\langle x, U\phi_\nu\rangle$, and hence that $G_\lambda^{-1}U\phi_\nu = (1 + \lambda\gamma_\nu)^{-1}\phi_\nu$. Thus,

$$\begin{aligned} G_\lambda^{-1}x &= \sum_\nu \langle G_\lambda^{-1}x, U\phi_\nu\rangle_{\mathscr{X}}\phi_\nu \\ &= \sum \langle x, G_\lambda^{-1}U\phi_\nu\rangle_{\mathscr{X}}\phi_\nu \\ &= \sum (1 + \lambda\gamma_\nu)^{-1}\langle x, U\phi_\nu\rangle_{\mathscr{X}}\phi_\nu. \end{aligned}$$

This implies in particular that

$$B_\lambda \xi = -\sum_\nu [\lambda\gamma_\nu/(1 + \lambda\gamma_\nu)]\langle \xi, U\phi_\nu\rangle_{\mathscr{X}}\phi_\nu. \tag{2.2}$$

In general, calculations involving the continuous operators are much easier when this eigensystem is used.

In order to accomplish this "simultaneous diagonalization" in general we need the following postulates.

ASSUMPTION 2.1. (a) For each n, the stochastic linear functional $\eta \mapsto \langle \varepsilon_n, \eta\rangle_{\mathscr{Y}}$ is defined for all $\eta \in \mathscr{Y}$ a.s. and has finite second moment. Furthermore,

$$E\langle \varepsilon_n, \eta\rangle_{\mathscr{Y}} = 0, \qquad \forall\, \eta \in \mathscr{Y}, \tag{2.3}$$

and there is a sequence of positive constants σ_n^2 such that

$$E\langle \varepsilon_n, \eta\rangle_{\mathscr{Y}}\langle \varepsilon_n, \zeta\rangle_{\mathscr{Y}} = \sigma_n^2\langle \eta, \zeta\rangle, \qquad \forall\, \eta \in \mathscr{Y}, \forall\, \zeta \in \mathscr{Y}. \tag{2.4}$$

(b) $T \in \mathscr{B}(\mathscr{X}, \mathscr{Y})$ is a compact operator, and $\mathscr{N}(T) = \{0\}$.

(c) $W \in \mathscr{B}(\mathscr{X})$ is a positive operator with $\dim \mathscr{N}(W) < \infty$.

Assumption (a) holds when $\varepsilon_n = \sigma_n\omega$ for a mean zero Gaussian process ω, and $\mathscr{Y}$ is the "reproducing kernel Hilbert space" for ω [12], not to be confused with the concept of the same name mentioned in Example 0.2. The assumptions on T

hold in many instances. Compactness of T in Example 0.3 follows from compactness of the imbedding J_{kp}: $W_2^k \to W_2^p$ for $k > p$ ([1], Theorem 6.2, equation (3)). For Example 1.1, decompose T into $J_{m0} \times (J_{p0} \circ D^p)$ to see that it is compact. [Note that $J_{p0} \circ D^p$ is compact by compactness of J_{p0} and the proof of Theorem 4.18(f) of [18].] $\mathcal{N}(T) = \{0\}$ implies $\mathcal{R}(T^*) = \mathcal{R}(U)$ is dense in $\mathscr{X}$ (Theorem 4.12 of [18]).

PROPOSITION 2.2. *There are sequences* $\{\phi_\nu\} \subseteq \mathscr{X}$ *and* $\{\gamma_\nu\} \subseteq [0, \infty)$ *satisfying* (2.1). *Furthermore, for all* $x \in \mathscr{X}$, $x = \sum_\nu \langle x, U\phi_\nu \rangle_{\mathscr{X}} \phi_\nu$, *where the series converges in* $\mathscr{X}$.

PROOF. Applying the theory in Section 3.3 of Weinberger [25], let $\{\psi_\nu\}$ and $\{\lambda_\nu\}$ be the eigenvectors and eigenvalues of the Rayleigh quotient $\langle x, Ux \rangle_{\mathscr{X}} / \langle x, (U + W)x \rangle_{\mathscr{X}}$, and put $\gamma_\nu = \lambda_\nu^{-1} - 1$ and $\phi_\nu = \gamma_\nu^{1/2} \psi_\nu$. □

For $\rho \in R$, let $\mathscr{X}_\rho^0$ be the set of $x \in \mathscr{X}$ for which the series

$$\|x\|_\rho^2 = \sum_\nu (1 + \gamma_\nu)^\rho \langle x, U\phi_\nu \rangle_{\mathscr{X}}^2$$

converges and let $\mathscr{X}_\rho$ be the completion of $\mathscr{X}_\rho^0$ in the indicated norm. Then $\mathscr{X}_\rho$ is a Hilbert space under the inner product

$$\langle x, \xi \rangle_\rho = \sum_\nu (1 + \gamma_\nu)^\rho \langle x, U\phi_\nu \rangle_{\mathscr{X}} \langle \xi, U\phi_\nu \rangle_{\mathscr{X}}.$$

For Examples 0.3 and 1.1 it will be shown in Section 3 that the $\mathscr{X}_\rho$ spaces correspond to Sobolev spaces, possibly with boundary conditions. Increasing ρ increases the smoothness and number of boundary conditions that must hold. In general, $\langle x_1, Ux_2 \rangle_{\mathscr{X}} = \langle x_1, x_2 \rangle_0$, and T extends to a Hilbert space isomorphism from $\mathscr{X}_0$ to $\mathscr{Y}$. Thus, $\mathscr{X}_0 = L_2$ in Example 0.3, and $\mathscr{X}_0 \cong W_2^p$ in 1.1, where $\cong$ means equal as sets and with equivalent norms. Since $\langle x_1, x_2 \rangle_1 = \langle x_1, (U + W)x_2 \rangle_{\mathscr{X}}$ and $(U + W) \in \mathscr{B}(\mathscr{X})$, we have $\mathscr{X} \subseteq \mathscr{X}_1$, where by convention the inclusion of Banach spaces means continuous imbedding. In Example 0.3, $\mathscr{X} \cong \mathscr{X}_1$, whereas $\mathscr{X} \cong \mathscr{X}_1$ in 1.1. One has $\mathscr{X} \cong \mathscr{X}_1$ more generally whenever $\mathscr{R}(U + W) = \mathscr{X}$. Clearly, $\mathscr{X}_\rho \subseteq \mathscr{X}_\tau$ whenever $\rho \geq \tau$.

THEOREM 2.3. *Let* $x \in \bigcup_{\rho \in R} \mathscr{X}_\rho$, *and fix* $\rho \in R$.

(a) $B_\lambda x = 0$ *for some* $\lambda > 0$ *if and only if* $x \in \mathcal{N}(W)$, *in which case* $B_\lambda x = 0$ *for* $\forall\ \lambda > 0$.

(b) $x \in \mathscr{X}_{\rho+2}$ *if and only if* $\|B_\lambda x\|_\rho = O(\lambda)$ *as* $\lambda \to 0$, *in which case* $\|B_\lambda x\|_\rho^2 = \lambda^2(1 + o(1)) \sum \gamma_\nu^2 (1 + \gamma_\nu)^\rho \langle x, \phi_\nu \rangle_0^2$.

(c) *Let* $\alpha \in [0, 1]$, *then* $\|B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\rho+2\alpha}, \mathscr{X}_\rho)} \leq 1$. *If* $\dim \mathscr{X} = \infty$, *then as* $\lambda \downarrow 0$, $\|B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\rho+2\alpha}, \mathscr{X}_\rho)} = \beta_\alpha \lambda^\alpha (1 + o(1))$, *where* $\beta_\alpha = \sup\{u^{1-\alpha}/(1 + u)\colon u > 0\}$.

PROOF. All claims follow straightforwardly from the formula

$$\|B_\lambda x\|_\rho^2 = \sum_\nu [\lambda\gamma_\nu/(1 + \lambda\gamma_\nu)]^2 (1 + \gamma_\nu)^\rho \langle x, \phi_\nu \rangle_0^2. \tag{2.5}$$

(a) and the "only if" part of (b) are evident. If $\|B_\lambda x\|_\rho = O(\lambda)$, then $\infty > \lambda^{-2} \|B_\lambda x\|_\rho^2 \geq \sum_\nu \gamma_\nu^2 (1 + \gamma_\nu)^\rho \langle x, \phi_\nu \rangle_0^2$, where the last inequality follows from

Fatou's lemma and shows $x \in \mathscr{X}_{\rho+2}$. For (c), the first part follows from $\lambda\gamma_\nu/(1+\lambda\gamma_\nu) \le 1$, and the second from the fact that $\lim_\nu \gamma_\nu = \infty$ if $\dim \mathscr{X} = \infty$. □

Note that $\|B_\lambda \xi\|_\rho$ tends to 0 faster with λ as ξ is in $\mathscr{X}_\tau$ for a larger τ, until $\tau = \rho + 2$, in which case no further gain in convergence rate is made unless $B_\lambda \xi = 0$, $\forall\ \lambda$. This exemplifies the saturation phenomenon well known in approximation theory; see Definition 2.1.1 of [2]. Part (b) of the previous theorem can also be deduced from Corollary 2.5.6 of [2]. Further results on B_λ are given in Theorem 3.3.

We now consider the random part of the estimation error or the "variance." Put $C(\lambda, \rho) = \sigma_n^{-2} E\|\xi_{n\lambda} - E\xi_{n\lambda}\|_\rho^2$. The following result is easily proved by comparison of sums and integrals (e.g., page 401 of [7]).

THEOREM 2.4. *Suppose that for some $r > 0$,*

$$\gamma_\nu \approx \nu^r, \qquad \nu \to \infty. \tag{2.6}$$

Then for $\rho < 2 - 1/r$, $C(\lambda, \rho) < \infty$ for $\lambda \in (0, \infty)$ and as $\lambda \downarrow 0$,

$$\begin{aligned} C(\lambda, \rho) &\approx \lambda^{-(\rho+1/r)}, && \text{if } -1/r < \rho < 2 - 1/r, \\ &\approx \log(1/\lambda), && \text{if } \rho = -1/r, \\ &\approx 1, && \text{if } \rho < -1/r. \end{aligned}$$

REMARK. The assumption (2.6) forces $\dim \mathscr{X} = \infty$. Although it is restrictive, it applies to Examples 0.3 and 1.1, as will be seen. Some such form seems necessary in order to obtain useable results on $C(\lambda, \rho)$.

PROPOSITION 2.5. *For Example 1.1, we have $\gamma_\nu \approx \nu^{2(m-p)}$.*

PROOF. We seek the eigenvalues of the Rayleigh quotient $\mathscr{B}/\mathscr{A}$, where $\mathscr{B}(x, x) = \int\{[x^{(p)}]^2 + bx^2\}f$ and $\mathscr{A}(x, x) = \int[x^{(m)}]^2 + \mathscr{B}(x, x)$. Utilizing the mapping principle of [25], we may replace $\mathscr{B}$ by $\mathscr{B}'(x, x) = \|x\|_{W_2^p}^2$ (as $\mathscr{B} \approx \mathscr{B}'$) and $\mathscr{A}$ by $\mathscr{A}'(x, x) = \|x\|_{W_2^m}^2$ (where we may use any equivalent norm on W_2^p or W_2^m), and the resulting eigenvalues γ_ν' will satisfy $\gamma_\nu' \approx \gamma_\nu$. From the argument of [3], Section 3, there exist sequences $\{\psi_\nu\}$ of functions and $\{\mu_\nu\}$ of nonnegative reals such that $\{\psi_\nu\}$ is an orthonormal basis for L_2 and $|||x|||_q^2 = \sum_\nu (1 + \mu_\nu^{q/m})\langle x, \psi_\nu\rangle_{L_2}^2$ gives an equivalent norm on W_2^q, $q = 0, 1, \ldots, m$. Furthermore, $\mu_\nu \approx \nu^{2m}$, so $\gamma_\nu' = (1 + \mu_\nu)/(1 + \mu_\nu^{p/m}) \approx \nu^{2(m-p)}$. □

For Example 0.3 one can obtain $\gamma_\nu \sim (\pi\nu)^{2m}$ and from this sharper asymptotics for $C(\lambda, \rho)$ as $\lambda \downarrow 0$. See, e.g., Theorem 2.4 of [21].

Theorems 2.3 and 2.4 combine to give convergence rates for $E\|\xi_{n\lambda} - \xi\|_\rho^2 = \|B_\lambda \xi\|_p^2 + \sigma_n^2 C(\lambda, \rho)$. For instance, if $\rho > -1/2(m-p) = -1/r$, $\xi \in \mathscr{X}_{\rho+2\alpha}$ for $0 \le \alpha \le 1$, and $\xi \in \mathscr{X}_0$ (so $T\xi$ is defined), then $E\|\xi_{n\lambda} - \xi\|^2 = O(\lambda^\alpha) + O(n^{-1}\lambda^{-(\rho+1/r)})$ as $n \to \infty$ and $\lambda \downarrow 0$. This is only "useful" if $\rho = 0$ and $\alpha = \frac{1}{2}$ since the only identifications we have are $\mathscr{X}_0 = W_2^p$ and $\mathscr{X}_1 = W_2^m$.

3. Identification of the $\mathscr{X}_\rho$ spaces. We present in this section a tool kit for identifying the spaces defined in Section 2.

Lemma 3.1. (a) *Let* $\rho \in R$ *and* $\delta \in R$. *Then* $\forall\ x \in \mathscr{X}_\rho$, $\|x\|_{\rho+\delta} = \sup\{\langle x, \zeta\rangle_\rho : \zeta \in \mathscr{X}_\rho$ *and* $\|\zeta\|_{\rho-\delta} \le 1\}$.

(b) $\mathscr{X}_{\rho+2}$ *is the range of* $G_\lambda^{-1}U = (I + \lambda A)^{-1}$ *as an operator on* $\mathscr{X}_\rho$ *for any* $\lambda > 0$ *where* $A = U^{-1}W$. *Furthermore,* $\|x\|_{\rho+2} = \|(I + A)x\|_\rho$.

Proof. (a) By a Cauchy–Schwarz argument using the eigenfunction expansions, $\langle x, \zeta\rangle_\rho \le \|x\|_{\rho+\delta}\|\zeta\|_{\rho-\delta}$, $\forall\ x, \zeta \in \mathscr{X}_\rho$. The inequality is valid even if one of the norms is infinite. If $\delta \le 0$, then the inequality is an equality for $\zeta_0 = \Sigma_\nu(1 + \gamma_\nu)^\delta\langle x, \phi_\nu\rangle_0\phi_\nu$, which is in $\mathscr{X}_\rho$, and the supremum is attained at a suitable multiple of ζ_0. If $\delta \ge 0$, then ζ_0 may not be in $\mathscr{X}_\rho$, but the partial sums of its defining series are, and they achieve equality in the limit.

(b) If $\zeta \in \mathscr{X}_\rho$, then $x = G_\lambda^{-1}U\zeta = \Sigma_\nu(1 + \lambda\gamma_\nu)^{-1}\langle\zeta, \phi_\nu\rangle_0\phi_\nu$ is in $\mathscr{X}_{\rho+2}$, and, conversely, $\zeta = \Sigma_\nu(1 + \lambda\gamma_\nu)\langle x, \phi_\nu\rangle_0\phi_\nu$ is in $\mathscr{X}_\rho$ whenever $x \in \mathscr{X}_{p+2}$. Also

$$\|(I + A)x\|_\rho^2 = \sum_\nu (1 + \gamma_\nu)^\rho\langle x, (I + A)\phi_\nu\rangle_0^2 = \sum_\nu (1 + \gamma_\nu)^{\rho+2}\langle x, \phi_\nu\rangle_0^2 = \|x\|_{\rho+2}^2.$$

□

Theorem 3.2. *In Example* 1.1, *assume* $f \in W_2^p$. *Then*

$$\mathscr{X}_2 \cong \mathscr{Z} \stackrel{\text{def}}{=} \{x \in W_2^{2m-p} : x^{(j)}(t) = 0 \text{ for } t = 0, 1 \text{ and } m \le j < 2m - p\}$$

equipped with W_2^{2m-p} *norm.*

Proof. We first show $\mathscr{Z} \subseteq \mathscr{X}_2$. If $x \in \mathscr{Z}$ and $\zeta \in \mathscr{X}_1 = W_2^m$, then

$$\begin{aligned}\langle x, \zeta\rangle_1 &= \int x^{(m)}\zeta^{(m)} + b\langle x, \zeta\rangle_0 \\ &= (-1)^{m-p}\int x^{(2m-p)}\zeta^{(p)} + b\langle x, \zeta\rangle_0 \\ &\le \|x^{(2m-p)}\|_{L_2}\|\zeta\|_{W_2^p} + b\|x\|_0\|\zeta\|_0 \le K\|x\|_{\mathscr{Z}}\|\zeta\|_0.\end{aligned}$$

This shows $\mathscr{Z} \subseteq \mathscr{X}_2$ by Lemma 3.1(a).

Now suppose $x \in \mathscr{X}_1$ and $S = \sup\{\langle x, \zeta\rangle_1 : \zeta \in \mathscr{X}_1, \|\zeta\|_0 \le 1\} < \infty$. We will show $x \in \mathscr{Z}$, and hence $\mathscr{X}_2$ is a subset of $\mathscr{Z}$, which implies $\mathscr{Z} \simeq \mathscr{X}_2$ by the open mapping theorem. Now $S < \infty$ if and only if $S_1 \stackrel{\text{def}}{=} \sup\{\int x^{(m)}\zeta^{(m)} : \zeta \in \mathscr{X}_1, \|\zeta\|_0 \le 1\} < \infty$, as $\langle x, \zeta\rangle_0 \le \|x\|_0$ for $\|\zeta\|_0 \le 1$.

Since $|||\zeta|||^2 = \|\zeta^{(p)}\|_{L_2}^2 + \Sigma_{j=0}^{p-1}[\zeta^{(j)}(0)]^2$ defines an equivalent norm on $\mathscr{X}_0$, $S_1 < \infty$ is equivalent to $S_2 < \infty$, where

$$\begin{aligned}S_2 &= \sup\left\{\int x^{(m)}\zeta^{(m)} : \zeta \in \mathscr{X}_1, |||\zeta||| \le 1\right\} \\ &= \sup\left\{\int x^{(m)}\zeta^{(m)} : \zeta \in \mathscr{X}_1, \zeta^{(j)}(0) = 0 \text{ for } 0 \le j < p \text{ and } \|\zeta^{(p)}\|_{L_2} \le 1\right\} \\ &= \sup\left\{\int x^{(m)}\zeta^{(m-p)} : \zeta \in W_2^{m-p} \text{ and } \|\zeta^{(m-p)}\|_{L_2} \le 1\right\}.\end{aligned}$$

The density of $\mathscr{X}_1$ in W_2^p was used at the last step. We claim $S_2 < \infty$ implies

$$\theta = x^{(p)} \in W_{2,NBC}^{2(m-p)}$$

$$\stackrel{\text{def}}{=} \{h \in W_2^{2(m-p)}: h^{(j)}(0) = h^{(j)}(1) = 0,\ m-p \le j < 2(m-p)\},$$

which is the desired result.

Letting $q = (m-p)$, our claim is that for all $\theta \in W_2^q$,

$$S' = \sup\left\{\int \theta^{(q)}\zeta^{(q)}: \zeta \in W_2^q,\ \|\zeta\|_{L_2} \le 1\right\} < \infty$$

implies $\theta \in W_{2,NBC}^{2q}$. Now $S' < \infty$ entails $S_1' = \sup\{\langle\theta, \zeta\rangle_{W_2^q}: \zeta \in W_2^q, \|\zeta\|_{L_2} \le 1\} < \infty$. Consider Example 0.3 with $\mathscr{X} = W_2^q = \mathscr{X}_1$, and then $\mathscr{X}_0 = L_2$, and then $S_1' < \infty$ if $\theta \in \mathscr{X}_2$ in the context of that example. Now $G_1^{-1}U$ as an operator on $L_2 = \mathscr{X}_0$ maps ζ to the minimizer over $\eta \in \mathscr{X}$ of $\|\eta^{(q)}\|_{L_2}^2 + \|\zeta - \eta\|_{L_2}^2$. The minimizer is by Section 2 of [3] the element of $W_{2,NBC}^{2q}$ satisfying $[(-D^2)^q + 1]\eta = \zeta$, with the indicated boundary conditions. This shows $G_1^{-1}U$ maps $\mathscr{X}_0$ onto $W_{2,NBC}^q$, which proves the claim by Lemma 3.1(b) with $\rho = 0$. □

We will identify $\mathscr{X}_\rho$ for $0 < \rho < 2$ in Example 1.1 using another technique, the so-called K-method of interpolation, which is regularization in another guise. Let $\mathscr{Z}_1$ and $\mathscr{Z}_2$ be Banach spaces with $\mathscr{Z}_2 \subseteq \mathscr{Z}_1$, and take $z \in \mathscr{Z}_1$ and $u \in (0, \infty)$. Put

$$K[u, z; (\mathscr{Z}_1, \mathscr{Z}_2)] = K(u, z) = \inf\left\{\left(\|z_1\|_{\mathscr{Z}_1} + u^2\|z_2\|_{\mathscr{Z}_2}\right)^{1/2}: z_i \in \mathscr{Z}_i,\ i = 1, 2 \text{ and } z = z_1 + z_2\right\}. \tag{3.1}$$

For $0 < \theta < 1$, $1 \le q < \infty$, define

$$\|z\|_{\theta,q} = \left\{\int_0^\infty [u^{-\theta}K(u, z)]^q u^{-1}\,du\right\}^{1/q},$$

and for $0 \le \theta \le 1$,

$$\|z\|_{\theta,\infty} = \operatorname{ess\,sup}\{u^{-\theta}K(u, z)\}.$$

Let

$$(\mathscr{Z}_1, \mathscr{Z}_2)_{\theta,q} = \{z \in \mathscr{Z}_1: \|z\|_{\theta,q} < \infty\}.$$

Then $(\mathscr{Z}_1, \mathscr{Z}_2)_{\theta,q}$ is a Banach space under the norm $\|\cdot\|_{\theta,q}$, and $\mathscr{Z}_2 \subseteq (\mathscr{Z}_1, \mathscr{Z}_2)_{\theta,q} \subseteq \mathscr{Z}_1$. Also set $(\mathscr{Z}_2, \mathscr{Z}_1)_{\theta,q} = (\mathscr{Z}_1, \mathscr{Z}_2)_{1-\theta,q}$. This is not the usual definition of the K functional, but gives equivalent norms and spaces. We have for $\beta \ge 0$ that $\mathscr{X}_{\rho+\theta\beta} \cong (\mathscr{X}_\rho, \mathscr{X}_{\rho+\beta})_{\theta,2}$, which follows from the definitions via direct calculation. See [1], [2], [15] and [24].

Now for $1 \le q \le \infty$ define $\mathscr{X}_{\rho,q} = (\mathscr{X}_{\rho-1}, \mathscr{X}_{\rho+1})_{1/2,q}$ and let $\|\cdot\|_{\rho,q}$ denote the associated norm. Then $\mathscr{X}_{\rho,2} \cong \mathscr{X}_\rho$, and if $\rho_1 < \rho < \rho_2$, $1 \le q \le \infty$ and $1 \le q' \le \infty$, then $\mathscr{X}_{\rho,q} \cong (\mathscr{X}_{\rho_1,q'}, \mathscr{X}_{\rho_2,q'})_{\theta,q}$ if $\theta = (\rho - \rho_1)/(\rho_2 - \rho_1)$. This is the theorem of reiteration (Section 1.10 of [24] or Section 3.2.4 of [2]). Also, $\mathscr{X}_{\rho,q} \subseteq \mathscr{X}_{\rho',q'}$ if either $\rho > \rho'$, or $\rho = \rho'$ and $q \le q'$ (Theorem 1.3.3 of [24]).

THEOREM 3.3. (a) *Suppose* $1 \le q \le \infty$ *and* $\rho < \beta < \rho + 2$. *Then*

$$\|B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\beta,q}, \mathscr{X}_\rho)} \le K(1 \wedge \lambda)^{(\beta-\rho)/2}.$$

(b) *For* $x \in \mathscr{X}_\rho$, $\|B_\lambda x\|_\rho = O(\lambda^\alpha)$ *for some* $\alpha \in (0,1)$ *if and only if* $x \in \mathscr{X}_{\rho+2\alpha,\infty}$.

PROOF. (a) We have $(\mathscr{X}_\rho, \mathscr{X}_{\rho+2})_{\theta,q} \cong \mathscr{X}_{\beta,q}$ if $\theta = (\beta - \rho)/2$ and $(\mathscr{X}_\rho, \mathscr{X}_\rho)_{\theta,q} = \mathscr{X}_\rho$. The result follows from Theorem 2.3(c) and the fact that K-method interpolation is "exact" [Section 3.2.5 of [2] or Theorem 1.3.3(g) of [24]].

(b) One direction follows from (a). We claim that there is a constant M such that for all $x \in \mathscr{X}_\rho$, $MK^2(\lambda^2, x) + O(\lambda^2) \le \|B_\lambda x\|_\rho^2$. Then if $\|B_\lambda x\|_\rho^2 = O(\lambda^{2\alpha})$, $0 < \alpha < 1$, this implies $x \in \mathscr{X}_{\rho+2\alpha,\infty}$ from the definition. Let $m = \dim \mathscr{N}(W)$, and let

$$M_1(x,\lambda) = \sum_{\nu=1}^{m} \lambda^2(1+\gamma_\nu)^{\rho+2}\left[1 + \lambda^2(1+\gamma_\nu)^2\right]^{-1}\langle x, \phi_\nu\rangle_0^2 = O(\lambda^2), \quad \text{as } \lambda \downarrow 0.$$

Then if $x \notin \mathscr{N}(W)$,

$$\|B_\lambda x\|_\rho^2/\left[K^2(\lambda^2, x) - M_1(x,\lambda)\right] \ge \left[\gamma_{m+1}^2/(1+\gamma_{m+1})^2\right] M_2(\lambda),$$

where

$$M_2(\lambda) = \inf_{\gamma \ge 0}\left\{\left[1 + \lambda^2(1+\gamma)^2\right]/(1+\lambda\gamma)^2\right\} \quad \text{and} \quad M_2(\lambda) \to 1, \quad \text{as } \lambda \to 0. \quad \square$$

The Besov spaces B_{2q}^γ on $[0,1]$ are now defined. For $0 < s < \infty$ and $1 \le q \le \infty$, let k denote the greatest integer less than s, then for $q < \infty$,

$$\|h\|_{B_{2q}^s} = \|h\|_{L_2} + \left\{\int_0^1 u^{-(s-k)q}\left\{\int_0^{1-u}\left[u^{-1}\left(D^k h(t+u) - D^k h(t)\right)\right]^2 dt\right\}^{q/2} du\right\}^{1/q}$$

and

$$\|h\|_{B_{2\infty}^s} = \|h\|_{L_2} + \sup_{0 \le u \le 1}\left\{\int_0^{1-u}\left[D^k h(t+u) - D^k h(t)\right]^2 dt\right\}^{1/2}.$$

This definition is based on equation (6) of Section 4.4.1 of [24]. We have $B_{22}^m \cong W_2^m$ for $m = 1, 2, \dots$. The Hölder space C^s (Section 4.5 of [24]) satisfies $C^s \subseteq B_{2,q}^s$ for any q, and $B_{2,q}^s \subseteq C^t$ provided $s > t + \frac{1}{2}$. See Theorem 4.6.1(e) and (f) of [24]. This implies the functional $h \to D^k h(t)$ (evaluation of kth derivative at t) is continuous on $B_{2,q}^s$ if $s > k + \frac{1}{2}$.

THEOREM 3.4. *In the setting of Example* 1.1, *assume* $f \in W_2^p$ *and put* $r = 2(m - p)$.

(a) *If* $0 < \rho < 1 + 1/r$, *then* $\mathscr{X}_{\rho,q} \cong B_{2q}^s$, *where* $s = p + \rho r/2$.

(b) *If* $1 + 1/r < \rho < 2$ *and* $(\rho r - 1)/2$ *is not an integer, then* $\mathscr{X}_{\rho,q} \cong \mathscr{Z}$, *where* $\mathscr{Z} = \{h \in B_{2q}^s\colon h^{(j)}(0) = h^{(j)}(1) = 0$ *for* $m \le j < \rho m - \frac{1}{2}\}$, *where* $s = p + \rho r/2$ *and* $\mathscr{Z}$ *is equipped with* B_{2q}^s *norm.*

(c) *If* $1 + 1/2m < \rho < 2$ *and* $(\rho r - 1)/2 = k$, *an integer, then* $\mathscr{X}_{\rho,2} = \{h \in B_{22}^s\colon h^{(j)}(0) = h^{(j)}(1) = 0,\ m \le j < k$, *and* $\int_0^1 [t(1-t)]^{-1}(h^{(k)}(t))^2\,dt < \infty\}$.

(d) *If* $f \equiv 1$, *then* $\mathscr{X}_{-\rho} \subseteq B_{2,2}^s$, *where* $s = p - \rho(m-p)$, *provided* $0 \le \rho \le [p/(m-p) \wedge 2]$.

PROOF. By definition and the reiteration theorem, $\mathscr{X}_{\rho,q} \cong (\mathscr{X}_0, \mathscr{X}_2)_{\beta,q}$, $\beta = \rho/2$, provided $0 < \rho < 2$. By Theorem 3.2, $\mathscr{X}_2 \cong$ a Besov space $B_{22,\{B_j\}}^{2m-p}$ with boundary conditions as defined in Section 4.3.3 of [24]. Note that $B_{22}^s \cong H_2^s$ by Remark 2.3.3.4 and Definition 4.2.1 of [24]. Also, $\mathscr{X}_0 \cong W_2^p \cong (L_2, H_{2,\{B_j\}}^{2m-p})_{\alpha,2}$ with $\alpha = p/(2m-p)$ [[24], Theorem 4.3.3(a)], so by the reiteration theorem $\mathscr{X}_{\rho,q} \cong ((L_2, \mathscr{X}_2)_{\alpha,2}, \mathscr{X}_2)_{\beta q} \cong (L_2, \mathscr{X}_2)_{\theta,q}$, where $\theta = [p + \rho(m-p)]/(2m-p)$. Parts (a) and (b) follow from [24], Equation 4.3.3(6), and (c) follows from 4.3.3(10).

For part (d), let $\mathscr{Z}_1 = \{\zeta \in W_2^{2p}\colon \zeta^{(j)}(0) = \zeta^{(j)}(1) = 0,\ m \le j < (2m-p) \wedge 2p\}$ equipped with the W_2^{2p} norm. Let $\mathscr{Z}_0$ be the closure of $\mathscr{X}_0$ in the norm

$$x \mapsto \|x\|_{L_2}^2 + \sum_{j=1}^{k} \left[x^{(p-j)}(0)^2 + x^{(p-j)}(1)^2\right],$$

where $k = p \wedge (m-p)$. Note that $\mathscr{Z}_0 \subseteq L_2$, so by a straightforward argument from the definition

$$(\mathscr{Z}_0, \mathscr{Z}_1)_{\theta,2} \stackrel{\text{def}}{=} \mathscr{Z}_\theta \subseteq (L_2, \mathscr{Z}_1)_{\theta,2} \stackrel{\text{def}}{=} \bar{\mathscr{Z}}_\theta.$$

From Theorem 4.3.3 of [24], $\bar{\mathscr{Z}}_\theta \cong \mathscr{X}_\rho$, $\rho = (\theta - 1/2)(2p)/(m-p)$ provided $0 \le \rho \le 2$, $\frac{1}{2} \le \theta \le 1$. (We have used $f \equiv 1$ here.) Let $\mathscr{Z}_1^*$ be the dual of $\mathscr{Z}_1$, and as $\mathscr{Z}_1 \subseteq \mathscr{Z}_\theta$ we define in the usual manner $\mathscr{Z}_\theta^* = \{\zeta \in \mathscr{Z}_1^*\colon \sup\{\zeta z\colon z \in \mathscr{Z}_1, \|z\|_{\mathscr{Z}_\theta} \le 1\} < \infty\}$ for $0 \le \theta < 1$. Then by the duality theorem (1.11.2 of [24]), $\mathscr{Z}_\theta^* = (\mathscr{Z}_0^*, \mathscr{Z}_1^*)_{\theta,2}$, for $0 < \theta < 1$. Now we represent $\mathscr{Z}_1^*$ with $\mathscr{Z}_0$ by using the $\mathscr{X}_0$ duality pairing. Each $x \in \mathscr{X}_0$ defines a bounded linear functional $\zeta \in \mathscr{Z}_1^*$ through

$$\zeta z = \langle x, z\rangle_0 = \int\left[(-1)^p z^{(2p)} + z\right]x + \sum_{j=1}^{m} (-1)^{j-1}\left[z^{(p+j-1)}(1)x^{(p-j)}(1) - z^{(p+j-1)}(0)x^{(p-j)}(0)\right].$$

By Remark 5.4.5.1 of [24],

$$z \mapsto \left((-1)^p z^{(2p)} + z, \left\{z^{(p+j-1)}(1), z^{(p+j-1)}(0)\colon 1 \le j \le m\right\}\right)$$

defines a bicontinuous linear bijection of $\mathscr{Z}_1$ onto $L_2 \times R^{2m}$, so $\|\zeta\|_{\mathscr{Z}_1^*} \approx \|x\|_{\mathscr{Z}_0}$. Since $\mathscr{Z}_1$ is dense in $\mathscr{X}_0$, all $\zeta \in \mathscr{Z}_1^*$ correspond in this manner to an element in $\mathscr{Z}_0$, the closure of $\mathscr{X}_1$ under this norm. The representation of $\mathscr{Z}_0^*$ obtained in the usual manner as before is $\mathscr{Z}_1$. Thus, by the duality theorem $\mathscr{Z}_\theta^*$ is represented by $\mathscr{Z}_{1-\theta}$. Also, by Lemma 3.1(a) the representation of $\mathscr{X}_\rho^*$ under the $\mathscr{X}_0$ duality pairing is $\mathscr{X}_{-\rho}$. Now we have $\mathscr{X}_\rho^* \cong \bar{\mathscr{Z}}_\theta^* \subseteq \mathscr{Z}_\theta^*$, so $\mathscr{X}_{-\rho} \subseteq \mathscr{Z}_{1-\theta} \subseteq \bar{\mathscr{Z}}_{1-\theta} \cong B_{22}^{(1-\theta)2p}$, for $\frac{1}{2} \le \theta \le 1$, $\rho = (2\theta - 1)p/(m-p)$, $0 \le \rho \le 1$. □

Numerous results on $E\|\xi_{n\lambda} - \xi\|_\rho^2$ for Example 1.1 emanate from Theorems 2.3 and 2.4 and 3.2–3.4. What is most often believed in practice (at least in the practice of Monte Carlo experiments) is that the unknown function ξ is quite smooth but does not satisfy any of the boundary conditions in Theorem 3.2 or 3.4(b)(c).

PROPOSITION 3.5. *Assume $\xi \in B_{2,\infty}^{m+1/2}$ and $f \in W_2^p$ in Example* 1.1. *The following hold as $n \to \infty$ and $\lambda \to 0$:*

(a) *If $p - 1/2 < u < m + 1/2$, then*

$$E\|\xi_{n\lambda} - \xi\|_{W_2^u}^2 = O\left(\lambda^{(m+1/2-\delta-u)/(m-p)} + n^{-1}\lambda^{-(2u-2p-1)/(2m-2p)}\right), \tag{3.2}$$

for any $\delta > 0$.

(b) *If $f \equiv 1$ and $0 \vee (2p + 1/2 - m) \le u < p - 1/2$, then*

$$E\|\xi_{n\lambda} - \xi\|_{W_2^u}^2 = O\left(\lambda^{(m+1/2-\delta-u)/(m-p)} + n^{-1}\right), \tag{3.3}$$

for any $\delta > 0$. In particular, if for some $\delta > 0$,

$$\lambda = O\left(n^{-(m-p)/(m+1/2-\delta-u)}\right), \tag{3.4}$$

then (3.3) *is $O(n^{-1})$.*

(c) *If $f \equiv 1$ and $0 \le u < (2p + 1/2 - m)$, then*

$$E\|\xi_{n\lambda} - \xi\|_{W_2^u}^2 = O(\lambda^2 + n^{-1}), \tag{3.5}$$

which is $O(n^{-1})$ if $\lambda = O(n^{-1/2})$.

PROOF. (a) and (b) follow from Theorems 2.3(c), 2.4 and 3.4 as $B_{2,\infty}^{m+1/2} \subseteq \mathscr{X}_\rho$ for any $\rho < 1 + 1/r$. Part (c) follows likewise, but one should note that $W_2^u \supseteq \mathscr{X}_{-\tau}$, $\tau > 1/r$, so the upper bound on ρ in Theorem 3.4(d) has no effect. □

4. Approximation of the discrete problem. In this section, we show how the discrete bias and variance can be approximated by their continuous analogs under the following postulates.

ASSUMPTION 4.1. (a) For each n, $\eta \mapsto \langle \varepsilon_n, \eta \rangle_{\mathscr{Y}_n}$ satisfies the same properties as in 2.1(a) with $\mathscr{Y}$ replaced by $\mathscr{Y}_n$.

(b) $W \in \mathscr{B}(\mathscr{X})$ satisfies 2.1(c).

(c) $T_n \in \mathscr{B}(\mathscr{X}, \mathscr{Y}_n)$ for all n, and for all n sufficiently large $\mathscr{N}(T_n) \cap \mathscr{N}(W) = \{0\}$.

(d) There is a compact $U \in \mathscr{B}(\mathscr{X})$ with $\mathscr{R}(U)$ dense in $\mathscr{X}$ satisfying (e) and (f).

(e) The eigenvalues $\{\gamma_\nu\}$ from Proposition 2.2 satisfy $\gamma_\nu \approx \nu^r$ as $\nu \to \infty$ for some $r > 1$.

(f) There exists $s \in (0, 1 - 1/r)$, $\{\rho_1, \rho_2, \dots, \rho_j\} \subseteq [0, s]$ with $\rho_1 < \rho_2 < \cdots < \rho_j$, and $\{k_n\} \subseteq [0, \infty)$ with $k_n \to 0$ such that for all $x_1, x_2 \in \mathscr{X}$,

$$\left|\left\langle x_1, (U - U_n)x_2\right\rangle_{\mathscr{X}}\right| \le k_n \sum_{i=1}^{J} \|x_1\|_{\rho_i}\|x_2\|_{s-\rho_i},$$

where $\|\cdot\|_\rho$ denotes the $\mathscr{X}_\rho$ norm generated by U.

The operator U in (d) is typically obtained from a continuous analog of the discrete problem, which for Example 0.4 is Example 1.1. Part (f) specifies the manner in which U approximates U_n. For Example 0.4, we have

$$\left\langle x_1, (U - U_n)x_2\right\rangle_{\mathscr{X}} = \int \left[x_1x_2 + bx_1^{(p)}x_2^{(p)}\right] d(F - F_n).$$

Applying integration by parts and Hölder inequalities gives

$$\begin{aligned}\left|\left\langle x_1, (U - U_n)x_2\right\rangle_{\mathscr{X}}\right| &\le (\sup|F - F_n|)\big(\|x_1\|_{L_2}\|x_2\|_{W_2^1} + \|x_1\|_{W_2^1}\|x_2\|_{L_2} \\ &\qquad + b\|x_1\|_{W_2^p}\|x_2\|_{W_2^{p+1}} + b\|x_1\|_{W_2^{p+1}}\|x_2\|_{W_2^p}\big) \\ &\le k_n\big(\|x_1\|_0\|x_2\|_s + \|x_1\|_s\|x_2\|_0\big),\end{aligned}$$

with $s = 1/(m - p)$ and $k_n = K\sup|F - F_n|$. Here, $j = 2$, but multivariate problems require $j > 2$. See the proof of Lemma 4.2(i) in [3].

Lemma 4.2. *Let $R_{n\lambda}$ be given by* (0.10). *Then for all $x \in \mathscr{X}_s$ and all $\rho \in R$,*

$$\|R_{n\lambda}x\|_\rho^2 \le k_n^2 \sum_{i=1}^{j} C(\lambda, \rho + s - \rho_i)\|x\|_{\rho_i}^2.$$

Proof. For all $x \in \mathscr{X}$,

$$\begin{aligned}\|R_{n\lambda}x\|_\rho^2 &= \sum_\nu (1 + \gamma_\nu)^\rho \left\langle(\lambda W + U)^{-1}(U - U_n)x, U\phi_\nu\right\rangle_{\mathscr{X}}^2 \\ &= \sum_\nu (1 + \gamma_\nu)^\rho (1 + \lambda\gamma_\nu)^{-2}\left\langle(U - U_n)x, \phi_\nu\right\rangle_{\mathscr{X}}^2.\end{aligned}$$

Applying 4.1(d) and $\|\phi_\nu\|_\rho^2 = (1 + \gamma_\nu)^\rho$ gives

$$\|R_{n\lambda}x\|_\rho^2 \le k_n^2 \sum (1 + \gamma_\nu)^\rho (1 + \lambda\gamma_\nu)^{-2} \sum_{i=1}^{j} \|x\|_{\rho_i}^2 (1 + \gamma_\nu)^{s-\rho_i}.$$

The result follows from this and the density of $\mathscr{X}$ in $\mathscr{X}_s$. □

Theorem 4.3. *Let $\rho < 2 - s - 1/r$ and suppose β and q satisfy $\rho < \beta < \rho + 2$ and $1 \le q \le \infty$, or $\beta = \rho + 2$ and $q = 2$. Assume $\rho_1 = 0$ and $\rho_j = s$ in* 4.1(f). *Then as $n \to \infty$,*

$$\|B_{n\lambda} - B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\beta,q}, \mathscr{X}_\rho)} = o\left(\|B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\beta,q}, \mathscr{X}_\rho)}\right), \quad \textit{uniformly in } \lambda \in [\lambda_n, \infty],$$

provided $\lambda_n \to 0$ *and one of the following holds*:

(i) $\rho > -1/r$, $s \le \beta < 2$ *and* $k_n^2\lambda_n^{-(s+1/r)} \to 0$.
(ii) $\beta = 2$, $1 \le q \le 2$ *and* $k_n^2\lambda_n^{-(s+1/r)} \to 0$.
(iii) $\beta = 2$, $2 < q \le \infty$ *and* $\exists\, \delta > 0$ *such that* $k_n^2\lambda_n^{-(s+1/r+\delta)} \to 0$.
(iv) $\beta > 2$ *and* $k_n^2\lambda_n^{-(s+1/r+\beta-2)} \to 0$.
(v) $\rho = -1/r$, $\beta \ge s$ *and* $k_n^2\lambda_n^{-(s+1/r)}\log(1/\lambda_n) \to 0$.
(vi) $\rho < -1/r$, $\beta \ge s$ *and* $k_n^2\lambda_n^{\rho-s} \to 0$.

PROOF. From (1.12), Lemma 4.2 and the triangle inequality we have $\forall\, \rho < 2 - s - 1/r$, $\forall\, \xi \in \mathscr{X}_\rho$,

$$\|B_{n\lambda}\xi\|_\rho \le \|B_\lambda \xi\|_\rho + k_n \sum_{i=1}^{j} C^{1/2}(\lambda, \rho + s - \rho_i)\|B_{n\lambda}\xi\|_{\rho_i}. \tag{4.1}$$

Put

$$\mu(\lambda, \tau) = (1 \wedge \lambda)^{\tau/2}, \quad \text{if } \tau < s + 1/r,$$
$$= 1, \quad \text{if } \tau \ge s + 1/r.$$

Note that all of the hypotheses on λ_n imply $k_n^2\lambda_n^{-(s+1/r)} \to 0$. Thus, by Theorem 2.4 and 4.1(f),

$$\forall\, \rho < 2 - 1/r - s, \forall\, \delta > 0, \exists\, N, \forall\, n \ge N, \forall\, \lambda \in [\lambda_n, \infty],$$
$$k_n C^{1/2}(\lambda, \rho + s - \rho_i) \le \delta\mu(\lambda, \rho_i - \rho).$$

Using this with (4.1), one can show

$$\exists\, K, \forall\, \delta > 0, \exists\, N, \forall\, n \ge N, \forall\, \lambda \in [\lambda_n, \infty], \forall\, k \in \{1, \dots, j\}, \forall\, \xi \in \mathscr{X}_s,$$

$$\|B_{n\lambda}\xi\|_{\rho_k} \le K\left\{\|B_\lambda\xi\|_{\rho_k} + \delta \sum_{\substack{i=1\\ i \ne k}}^{j} \mu(\lambda, \rho_i - \rho_k)\|B_\lambda \xi\|_{\rho_i}\right\}. \tag{4.2}$$

This follows from the next inequality, which can be proved by induction on h:

$$\forall\, h \in \{1, \dots, j\}, \exists\, K, \forall\, \delta > 0, \exists\, N,$$
$$\forall\, n \ge N, \forall\, \lambda \in [\lambda_n, \infty], \forall\, k \in \{1, \dots, j\}, \forall\, \xi \in \mathscr{X}_s,$$
$$\|B_{n\lambda}\xi\|_{\rho_k} \le K\left\{\|B_\lambda\xi\|_{\rho_k} + \delta \sum_{\substack{i=1\\ i \ne k}}^{h} \mu(\lambda, \rho_i - \rho_k)\|B_\lambda\xi\|_{\rho_i}\right.$$
$$\left. + \delta \sum_{\substack{i=h+1\\ i \ne k}}^{j} \mu(\lambda, \rho_i - \rho_k)\|B_\lambda\xi\|_{\rho_i}\right\}.$$

From Lemma 4.2 and (4.2) we have

$$\|B_{n\lambda} - B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\beta,q}, \mathscr{X}_\rho)} \le K k_n \sum_{i=1}^{j} \sum_{k=1}^{j} C^{1/2}(\lambda, \rho + s - \rho_i) \times \mu(\lambda, \rho_k - \rho_i)\|B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\beta,q}, \mathscr{X}_{\rho_k})}. \tag{4.3}$$

The theorem is a straightforward consequence of this, Theorem 2.4 and the estimates on the norm of B_λ in Theorems 2.3 and 3.3. For instance, take condition (iv) and assume $\rho_i \neq \beta - 2$ for all i, or else that $1 \le q \le 2$. Then

$$\|B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\beta,q},\mathscr{X}_{\rho_k})} \le K(1 \wedge \lambda)^{(\beta-\rho_k-(0\vee(\beta-\rho_k-2)))/2}.$$

When this is put into (4.3) and one notes that $C(\lambda, \rho + s - \rho_i) \approx (1 \wedge \lambda)^{-(\rho+s-\rho_i+1/r)/2}$ as $\rho + s - \rho_i > -1/r$ for all i, there results

$$\begin{aligned}\|B_{n\lambda} - B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\beta,q},\mathscr{X}_\rho)} &\le Kk_n \sum_{i=1}^{j}\sum_{k=1}^{j} \\ &\qquad \times (1 \wedge \lambda)^{-(\rho+s-\rho_i+1/r)/2+(\rho_k-\rho_i)/2+(\beta-\rho_k-(0\vee(\beta-\rho_k-2)))/2} \\ &\le Kk_n(1 \wedge \lambda)^{-(s+1/r+\beta-\rho_i-2)/2}(1 \wedge \lambda)^{(\beta-\rho)/2} \\ &= o\big((1 \wedge \lambda)^{(\beta-\rho)/2}\big) = o\big(\|B_\lambda\|_{\mathscr{B}(\mathscr{X}_{\beta,q},\mathscr{X}_\rho)}\big).\end{aligned}$$

The proof under (iv) when $\rho_k = \beta - 2$ for some $k > 1$, and under (i), (ii) or (iii) is similar. The proof under (iv) and (v) makes use of the different forms of the estimates on $C(\lambda, \rho + s - \rho_j)$ in Theorem 2.4 and the assumption that $\rho_j = s$. Note in (vi) that it suffices to prove the result for $\rho = -(\delta + 1/r)$ and then it holds for any $\rho' \le \rho$. □

There are numerous obvious extensions to other values of β. We assumed $\rho_1 = 0$ and $\rho_j = s$ merely for convenience. Next we turn to the variance.

LEMMA 4.4. *Let* $\rho < 2 - s - 1/r$ *and* $\{\lambda_n\} \subseteq (0, \infty)$ *with* $\lambda_n \to 0$. *Then* $E\|(\lambda W + U)^{-1}T_n^*\varepsilon_n\|_\rho^2 = \sigma_n^2 C(\lambda, \rho)(1 + o(1))$ *uniformly in* $\lambda \in [\lambda_n, \infty]$ *provided one of the following holds*:

(i) $-1/r < \rho < 2 - s - 1/r$ *and* $k_n\lambda_n^{-s} \to 0$.
(ii) $\rho < 1 - 1/r$ *and* $k_n^2\lambda_n^{-(s+1/r)} \to 0$.

PROOF. One can show that

$$\begin{aligned}&\left|E\left\|(\lambda W + U)^{-1}T_n^*\varepsilon_n\right\|_\rho^2 - \sigma_n^2 C(\lambda, \rho)\right| \\ &\qquad = \sigma_n^2\left|\sum_\nu (1 + \gamma_\nu)^\rho(1 + \lambda\gamma_\nu)^{-2}\langle\phi_\nu, (U - U_n)\phi_\nu\rangle_{\mathscr{X}}\right|.\end{aligned} \tag{4.4}$$

Now consider the various ranges for ρ. Under (i), apply the triangle inequality and 4.1(f) to show that the l.h.s. of (4.4) is $\le K\sigma_n^2 k_n C(\lambda, \rho + s)$. The conclusion now follows from Theorem 2.4.

Under (ii), a more complicated argument can be used. We have

$$\begin{aligned}\left|(1 + \lambda\gamma_\nu)^{-1}\langle\phi_\nu, (U - U_n)\phi_\nu\rangle_{\mathscr{X}}\right| &= |\langle\phi_\nu, R_{n\lambda}\phi_\nu\rangle_0| \\ &\le \|\phi_\nu\|_0\|R_{n\lambda}\phi_\nu\|_0 \\ &\le k_n \sum_{i=1}^{j} C^{1/2}(\lambda, s - \rho_i)(1 + \gamma_\nu)^{\rho_i/2},\end{aligned} \tag{4.5}$$

where the last expression results from Lemma 4.2. After some calculation one can show that the r.h.s. of (4.5) is

$$\leq \sigma_n^2 k_n \sum_i C^{1/2}(\lambda, s - \rho_i) \sum_\nu (1 + \gamma_\nu)^{\rho + \rho_i/2}(1 + \lambda\gamma_\nu)^{-1}.$$

An argument similar to the one in Theorem 2.4 will show $\Sigma_\nu(1 + \gamma_\nu)^\tau(1 + \lambda\gamma_\nu)^{-1}$ obeys the same estimates as $C(\lambda, \tau)$ for $\tau < 1 - 1/r$. Plug these into the previous display to obtain the result. □

Theorem 4.5. *If $\rho < 2 - s - 1/r$ and $\lambda_n \to 0$, then*

$$E\left\|(\lambda W + U_n)^{-1} T_n^* \varepsilon_n\right\|_\rho^2 = \sigma_n^2 C(\lambda, \rho)(1 + o(1)),$$

uniformly in $\lambda \in [\lambda_n, \infty]$ provided one of the following holds:

(i) $\rho < 1 - 1/r$ *and* $k_n^2 \lambda_n^{-(s+1/r)} \to 0$.
(ii) $1 - 1/r \leq \rho$ *and both* $k_n^2 \lambda_n^{-(s+1/r)} \to 0$ *and* $k_n \lambda_n^{-s} \to 0$.

Proof. Some simple algebra shows

$$(\lambda W + U)^{-1} - (\lambda W + U_n)^{-1} = R_{n\lambda}(\lambda W + U_n)^{-1}.$$

Hence, by Lemma 4.2, if $\rho < 2 - s - 1/r$, then

$$\begin{aligned} &E\left\|\left[(\lambda W + U)^{-1} - (\lambda W + U_n)^{-1}\right] T_n^* \varepsilon_n\right\|_\rho^2 \\ &\quad \leq k_n^2 \sum_{k=1}^{j} C(\lambda, \rho + s - \rho_k) E\left\|(\lambda W + U_n)^{-1} T_n^* \varepsilon_n\right\|_{\rho_k}^2. \end{aligned} \tag{4.6}$$

Similar to the proof of (4.2), one can "invert" this linear inequality to obtain $\exists\, K, \exists\, N, \forall\, n \geq N, \forall\, \lambda \in [\lambda_n, \infty], \forall\, k \in \{1, \ldots, j\}$,

$$E\left\|(\lambda W + U_n)^{-1} T_n^* \varepsilon_n\right\|_{\rho_k}^2 \leq K \sum_{i=1}^{j} (1 \wedge \lambda)^{(\rho_i - \rho_k)} E\left\|(\lambda W + U)^{-1} T_n^* \varepsilon_n\right\|_{\rho_i}^2,$$

provided $k_n^2 \lambda_n^{-(s+1/r)} \to 0$ as $n \to \infty$.

The theorem follows from this, Lemma 4.4 and (4.6). □

There are numerous possible results one can now obtain regarding Example 0.4. We content ourselves with the following analog of Proposition 3.5.

Proposition 4.6. *Assume in Example* 0.4 *that $\xi \in B_{2,\infty}^{m+1/2}$ and $m > p + 1$. Then the following hold as $n \to \infty$ and $\lambda \to 0$:*

(a) *If $p - 1/2 < u < m + 1/2$ and $k_n \lambda^{-3/(4m-4p)} \to 0$, then* (3.1) *holds for any $\delta > 0$.*

(b) *If* $0 \vee (2p + 1/2 - m) \leq u < p - 1/2$ *and*

$$k_n \lambda^{(u-p-1)/(2m-2p)} \to 0, \tag{4.7}$$

then (3.3) *holds for any* $\delta > 0$. *If, additionally,* (3.4) *holds, then*

$$E\|\xi_{n\lambda} - \xi\|^2_{W_2^u} = O(n^{-1}). \tag{4.8}$$

(c) *If* $0 \leq u < 2p + 1/2 - m$ *and* (4.7) *holds, then* (3.5) *holds, and if* $\lambda = O(n^{-1/2})$, *then* (4.8) *holds.*

Always $n^{-1} = O(k_n)$, so the set of λ wherein (4.7) and (3.4) hold is nonempty, and similarly for (3.5) and $\lambda = O(n^{-1/2})$. In view of [22] it is perhaps surprising that one can achieve $O(n^{-1})$ rate of convergence for a mean squared norm of error-type quantity in a nonparametric regression problem. The key here is the derivative data and the choice of norms ($u < p - 1/2$ so $\rho < -1/r$). There is an interesting difference between the discrete and continuous problems. One can put $\lambda = 0$ in Proposition 3.6(b) and (c) and obtain $O(n^{-1})$ convergence rate. However, ξ_{n0} still interpolates the data for the discrete problem and so is not consistent.

REFERENCES

[1] ADAMS, R. (1975). *Sobolev Spaces*. Academic, New York.

[2] BUTZER, P. and BERENS, H. (1967). *Semi-groups of Operators and Approximation*. Springer, Berlin.

[3] COX, D. (1984). Multivariate smoothing spline functions. *SIAM J. Numer. Anal.* **21** 789–813.

[4] COX, D. (1984). Gaussian approximation of smoothing splines. *Probab. Theory Related Fields*. To appear.

[5] COX, D. and NYCHKA, D. (1984). Convergence rates for regularized solutions of integral equations from noisy data. Unpublished.

[6] COX, D. and O'SULLIVAN, F. (1985). Analysis of penalized likelihood type estimators with application to generalized smoothing on Sobolev spaces. Unpublished.

[7] CRAVEN, P. and WAHBA, G. (1979). Smoothing noisy data with spline functions. *Numer. Math.* **31** 377–403.

[8] ENGL, H. (1981). Necessary and sufficient conditions for convergence of regularization methods for solving linear operator equations of the first kind. *Numer. Funct. Anal. Optim.* **3** 201–222.

[9] GROETSCH, C. (1977). *Generalized Inverses of Linear Operators*. Dekker, New York.

[10] HOERL, A. and KENNARD, R. (1970). Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics* **12** 55–67.

[11] KIMELDORF, G. and WAHBA, G. (1970). Spline functions and stochastic processes. *Sankhyā Ser. A* **32** 173–180.

[12] KUELBS, J. (1971). Expansions of vectors in a Banach space related to Gaussian measures. *Proc. Amer. Math. Soc.* **27** 364–370.

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

[14] LUKAS, M. (1984). Convergence rates for regularized solutions. Unpublished.

[15] RAGOZIN, D. (1983). Error bounds for derivative estimates based on spline smoothing of exact or noisy data. *J. Approx. Theory* **37** 335–355.

[16] RAO, C. R. (1973). *Linear Statistical Inference and Its Applications*, 2nd ed. Wiley, New York.

[17] RICE, J. and ROSENBLATT, M. (1983). Smoothing splines: Regression, derivatives and deconvolution. *Ann. Statist.* **11** 141–156.

[18] RUDIN, W. (1973). *Functional Analysis*. McGraw-Hill, New York.

[19] SCHWARZ, K. (1979). Geodetic improperly posed problems and their regularization. *Bolletino di Geodesia e Scienze Affini* **38** 389–416.

[20] SPECKMAN, P. (1979). Minimax estimates of linear functionals on a Hilbert space. Unpublished.
[21] SPECKMAN, P. (1981). On the asymptotic mean squared error for smoothing noisy data by splines. Unpublished.
[22] STONE, C. (1982). Optimal global rates of convergence for nonparametric regression. *Ann. Statist.* **10** 1040–1053.
[23] TIKHONOV, A. (1963). Solution of incorrectly posed problems and the method of regularization. *Soviet Math. Dokl.* **5** 1035–1038.
[24] TRIEBEL, H. (1978). *Interpolation Theory, Function Spaces, Differential Operators*. North-Holland, Amsterdam.
[25] WEINBERGER, H. (1974). *Variational Methods for Eigenvalue Approximation*. Arrowsmith, Bristol, England.

DEPARTMENT OF STATISTICS
UNIVERSITY OF ILLINOIS
725 SOUTH WRIGHT STREET
CHAMPAIGN, ILLINOIS 61820