JOURNAL OF APPROXIMATION THEORY 37, 335–355 (1983)

Error Bounds for Derivative Estimates Based on Spline Smoothing of Exact or Noisy Data

DAVID L. RAGOZIN

Department of Mathematics, University of Washington, GN-50, Seattle, Washington 98195, U.S.A.

Communicated by Carl de Boor

Received October 27, 1981; revised May 11, 1982

Estimates are found for the L_2 error in approximating the jth derivative of a given smooth function f by the corresponding derivative of the $2m$th order smoothing spline based on an n-point sample from the function. The results cover both the case of an exact sample from f and the case when the sample is subject to some random noise. In the noisy case, the estimates are for the expected value of the approximation error. These bounds show that, even in the presence of noise, the derivatives of the smoothing splines of order less than m can be expected to converge to those of f as the number of (uniform) sample points increases, and the smoothing parameter approaches zero at a rate appropriately related to m, n, and the order of differentiability of f.

1. INTRODUCTION

Error bounds for L_2 approximation by spline interpolants of exact data generated by a sooth function f have been known for some time. See, e.g., Schultz [17] and the references cited there. These bounds and related theoretical work such as [12] show that spline interpolation is essentially an optimal method for function (and derivative) estimation from finite exact data samples, which, moreover, is quite practical at the computational level.

Spline interpolation, however, along with most interpolation processes, is subject to significant distortion due to noise in the data, particularly when derivatives of f are to be estimated by differentiation of the interpolating spline. Spline smoothing, as described, for instance, in [14, 16], is quite similar to a Tikhonov regularization (cf. [19]), and so it can be expected to be much less sensitive to noise in the data, particularly for derivative estimation. In fact, several results in [11, 18] show that spline smoothing has optimal properties for function (and derivative) estimation based on data samples subject to *errors*. However general error bounds for spline smoothing, analogous to those for spline interpolation, do not seem to be known, although some related estimates are given in Cullum [6] on numerical differentiation.

0021-9045/83 $3.00

In this paper error bounds are obtained for the order of approximation to a function f possessing at least $k \leqslant m$ L_2 derivatives by the $2m$th order smoothing spline with smoothing parameter λ based on either exact or inexact data (see Eq. (2.1) for the precise definition of this smoothing spline). For an exact n-point sample our L^2 estimates show that the jth derivative of the smoothing spline ($j \leqslant k$) converges to $f^{(j)}$ at a rate bounded, roughly, by $\lambda^{(k-j)/2m}$ as $\lambda \to 0$ and $n \to \infty$ with $1/n\lambda^{1/2m}$ bounded. For inexact data our results confirm the expected stability of the spline smoothing process and they give bounds which indicate that derivative estimation is possible even in the presence of a noisy sample.

Specifically, when the n-point sample of function values used to generate the smoothing spline is subject to some random noise, we estimate the expected value of the error in approximating derivatrives of f by the corresponding derivatives of the smoothing spline based on this noisy data. When combined with some estimates due to Craven and Wahba [5] these results show that for equally spaced samples if $n \to \infty$ and $\lambda \sim n^{-2m/(2k+1)}$, then the jth derivative of the smoothing spline converges to $f^{(j)}$ at a rate bounded by $n^{-(k-j)/(2k+1)}$. This extends results in [5] for the case $j=0$, $k=m$.

Since the computation of the smoothing spline from any data sample is relatively easy once the smoothing parameter λ has been chosen (see [8, 9] for Algol and Fortran code), these estimates lead to a practical method for computing derivative estimates. One apparently effective way to select the smoothing parameter is by the method of generalized cross-validation, as described in [5], which is known to procedure essentially optimal values for λ for the function estimation problem. Portable Fortran code for this method exists in several places, including the IMSL library (Edition 8) and [7] for the cubic ($m=2$) smoothing spline case and in [20]. Evidence for the efficacy of this method of spline smoothing for function (and first derivative) estimation can be found in [5, 22]. A number of examples of the successful use of these techniques to estimate first through third derivatives of several functions, based on both artificial and real data sets, are given in [13].

In outline, the paper proceeds as follows. First the definitions of the basic spaces and (spline) smoothing operators are given in Section 2. In addition this section contains of the basic error estimates for the Tikhanov regularizer which will serve as a model for the spline smoothing estimates. In Section 3, we provide some background estimates which relate the Euclidean norm of the exact sample from f to various Sobolev norms of f and the size of the mesh at which f is sampled. In Section 4 these estimates are combined with some facts derived from the minimizing definition of the smoothing spline to get the desired error bounds for the exact data case. Finally in Section 5, we treat the case of error bounds for data subject to random noise.

2. Notation, Definitions, and Basic Properties of the Smoothing Operators

In this section we shall give the basic definitions of the function spaces, W_2^m and the smoothing operators we work with. In addition several basic facts are proved about these operators and spaces and their relation via interpolation theory.

We work in an L^2 setting for most of this paper and the basic Sobolev spaces we use are

$$W_2^m[a, b] = \{f : f, f^{(1)}, \ldots, f^{(m-1)} \text{ abs. cont. } f^{(m)} \text{ in } L_2[a, b]\}.$$

For any $k \leqslant m$, there are seminorms on this space defined by

$$|f|_k^2 = \int_a^b (f^{(k)}(x))^2 dt$$

derived from the semi-inner product

$$\langle f, g \rangle_k = \int_a^b f^{(k)}(x)\, g^{(k)}(x)\, dx.$$

(We shall drop the integration variable from here on.) For most of our work only these seminorms will be needed rather than any norm, but when a norm is required we shall use

$$\|f\|_m^2 = |f|_0^2 + (b-a)^{2m}\, |f|_m^2,$$

where the factor $(b-a)^{2m}$ is used to make the norm relatively invariant under rescaling of the interval. This means that if f is in $W_2^m[a, b]$ and h in $W_2^m[0, 1]$ is defined by $h(y) = f(a + y(b-a))$, then with the given definitions of the norms $\|h\|_m^2 = (b-a)^{-1} \|f\|_m^2$.

Let $\Delta = \Delta_n = \{a \leqslant x_1 < \cdots < x_n \leqslant b\}$ denote an n-point partition of $[a, b]$. For any function f defined on all of $[a, b]$ we denote by $\mathbf{f}_\Delta$ the column vector in R^n given by

$$\mathbf{f}_\Delta = [f(x_i)].$$

In general for vectors $\mathbf{y} = [y_i]$, $\mathbf{z} = [z_i]$ in R^n we shall use the norm and inner product given by

$$\|\mathbf{y}\|^2 = \|\mathbf{y}\|_{R^n}^2 = \sum_{i=1}^n y_i^2, \qquad \langle \mathbf{y}, \mathbf{z} \rangle = \sum_{i=1}^n y_i z_i.$$

Our basic object of attention will be the *spline smoothing operators* $S_{n,\lambda,m}\colon R^n \to W_2^m[a, b]$, defined for a smoothing order $m \geqslant 1$, a partition Δ_n

with $n \geqslant m$, a smoothing parameter $\lambda > 0$, and a vector of data values $\mathbf{y}$ in R^n by

$$S_{n,\lambda,m}(\mathbf{y}) = g_{n,\lambda,m} \text{ if and only if } g_{n,\lambda,m} \text{ solves the problem:}$$

$$\text{Find } f \text{ in } W_2^m[a,b] \text{ which minimizes } \frac{b-a}{n} \|\mathbf{y} - \mathbf{g}_\Delta\|^2 + \lambda\, |g|_m^2. \qquad (2.1)$$

The minimizer $g_{n,\lambda,m}$ exists and is unique since $n \geqslant m$ implies the map I_λ: $W_2^m[a,b] \to R^n \oplus L_2$, with $I_\lambda f = (\mathbf{f}_\Delta, \lambda^{1/2} f^{(m)})$ is an isomorphism onto a *closed* subspace H_λ of the Hilbert space direct sum and $g_{n,\lambda,m}$ is just the inverse image under I_λ of the orthogonal projection of $(\mathbf{y}, 0)$ onto H_λ. It is known [16] that $S_{n,\lambda,m}(\mathbf{y})$ is a $2m$th order *natural spline* on $[a,b]$ with knots at the x_i, i.e., $S_{n,\lambda,m}(\mathbf{y})$ is a C^{2m-2} function which is a polynomial of degree $2m-1$ on each $[x_i, x_{i+1}]$ and a polynomial of degree $m-1$ on $(-\infty, x_1]$ and $[x_n, \infty)$. We extend the definition to $\lambda = 0$ by letting $S_{n,0,m}(\mathbf{y})$ be the $2m$th order natural spline interpolant to the data y_i at the knots x_i, since the $\lim_{\lambda \to 0} S_{n,\lambda,m} = S_{n,0,m}$ [16].

Most of our results in this paper concern norm estimates for $|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_k^2$, $k \leqslant m$. Almost all of these estimates parallel corresponding estimates for the Tikhonov regularizer $S_{\lambda,m} f$ of f defined for any f in $L_2[a,b]$, $\lambda > 0$, by

$$S_{\lambda,m} f = g_{\lambda,m} \text{ if and only if } g_{\lambda,m} \text{ solves the problem:}$$

$$\text{Find } g \text{ in } W_2^m[a,b] \text{ which minimizes } |f - g|_0^2 + \lambda\, |g|_m^2. \qquad (2.2)$$

Again the existence of the required minimizer follows from some projection facts about Hilbert space and the closure of the range of the injection taking f in W_2^m to $(f, \lambda^{1/2} f^{(m)})$ in $L_2 \oplus L_2$. Since we shall need the known estimates for $S_{\lambda,m}$ and some of the techniques from their proof, we now turn to these results.

As a start we note a simple fact which allows us to prove all our estimates only on $[0,1]$. Namely, we have

PROPOSITION 2.3. *For f defined on $[a,b]$, let $h(y) = f(x)$, $x = a + y(b-a)$. Then $S_{n,\lambda,m}(\mathbf{f}_\Delta)(x) = S_{n,\lambda/(b-a)^{2m},m}(\mathbf{h}_\Delta)$ and $S_{\lambda,m} f(x) = S_{\lambda/(b-a)^{2m},m} h(y)$.*

Proof. This follows by simply effecting the change of variables $x = a + y(b-a)$ in the defining equations (2.2) and (2.1) for $S_{n,\lambda,m}(\mathbf{f}_\Delta)$ and $S_{\lambda,m} f$. ∎

Next we give a simple extension of some basic facts about the K-method of interpolation applied to the pair of spaces $(W_2^k[a,b], W_2^m[a,b])$. Adapting the notation of Bergh and Löfström [3, Sect. 3], we shall write $(A, B)_{\theta,q}$ for the K-method intermediate space between A and B with parameters Θ, q; in fact we shall only use $q = 2$. Specifically we shall need the Besov spaces B^s_{22} which are given by

$$B^s_{22}[a, b] = (W^k_2[a, b], W^m_2[a, b])_{\Theta,2}, \qquad s = \Theta k + (1 - \Theta)\, m.$$

Let us also introduce the notation

$$\Pi_m = \{p \text{ in } W^m_2 : p^{(m)} = 0\}$$

for the polynomials of degree less than m. Then we can summarize some useful facts in

PROPOSITION 2.4. *For integers k, m with $0 \leqslant k < m$ and any real Θ, $0 < \Theta < 1$, let $s = \Theta k + (1 - \Theta)\, m$.*

(i) *If s is integral, then $B^s_{22}[a, b] = W^s_2[a, b]$.*

(ii) *The quotient norm on $W^m_2[a, b]/\Pi_m$ is equivalent to $|\,|_m$.*

(iii) *For p in Π_m, $S_{\lambda,m}(p) = p = S_{n,\lambda,m}(\mathbf{p}_\Delta)$.*

Proof. When $[a, b] = (-\infty, \infty)$ (i) is just [3, (6.4.5)]. The finite interval case is in [10, p. 166], and follows from the full line case once one notes that there exist extension operators $I_m: L_2[a, b] \to L_2[R]$ with $I_m f = f$ on $[a, b]$ and $I_m(W^k_2[a, b]) \subset W^k_2[R]$. See [1, 4.28].

For (ii) note that Π_m is the nullspace of the seminorm $|\,|_m$ and this seminorm is continuous with respect to the norm $\|\,\|_m$. Hence the quotient norm and $|\,|_m$ are equivalent Hilbert space norms on W^m_2/Π_m.

(iii) This part follows immediately from definitions (2.1) and (2.2) since $g = p$ in Π_m makes the quantities to be minimized equal to zero. ∎

Now we can prove a basic theorem which simultaneously estimates the accuracy of approximation of the Tikhonov regularizer in L_2, and bounds $S_{\lambda,m} f$ in the seminorm $|\,|_m$.

THEOREM 2.5. *For $0 \leqslant k \leqslant m$ there exist constants $\alpha = \alpha(m, k)$ (independent of $[a, b]$) such that for f in $W^k_2[a, b]$*

$$|f - S_{\lambda,m} f|^2_0 + \lambda\, |S_{\lambda,m} f|^2_m \leqslant \alpha \lambda^{k/m}\, |f|^2_k. \tag{2.6}$$

Proof. This follows from some simple consequences of the minimization property of $S_{\lambda,m}$ together with the previous propositions. A simple derivative computation shows it will suffice to prove this theorem when $[a, b] = [0, 1]$

as a result of the change of variables Proposition 2.3. So we shall proceed to prove (2.6) for the interval $[0, 1]$ and suppress all references to the interval.

First note that for f in W_2^m,

$$|f - S_{\lambda,m}f|_0^2 + \lambda\, |S_{\lambda,m}f|_m^2 \leqslant \lambda\, |f|_m^2 \ (\leqslant \lambda\, \|f\|_m^2)$$

since $g = f$ is admissible in the minimization (2.2) used to define $S_{\lambda,m}f$. This gives (2.6) for $k = m$ with $\alpha(m, m) = 1$. Moreover, by letting $g = 0$ in that minimization it follows that

$$|f - S_{\lambda,m}f|_0^2 + \lambda\, |S_{\lambda,m}f|_m^2 \leqslant |f|_0^2 \tag{2.7}$$

which is the desired result for $k = 0$ with $\alpha(m, 0) = 1$.

The two preceding inequalities show that for $k = 0$ and $k = m$ the linear transformation T, with

$$T(f) = (f, 0) - (S_{\lambda,m}f, \lambda^{1/2}(S_{\lambda,m}(f))^{(m)}) \tag{2.8}$$

maps W_2^k to $H = L_2 \oplus L_2$ and the norm of T satisfies $\|T\|^2 \leqslant \lambda^{k/m}$. If we apply the standard interpolation theorems for the K-method to T as a map between the pairs (W_2^0, W_2^m) and (H, H), then we get (see [3, (3.1.2)])

$$T\colon B_{22}^{(1-\Theta)m} = (W_2^0, W_2^m)_{\Theta,2} \to H, \qquad \|T\|^2 \leqslant \lambda^{1-\Theta} \tag{2.9}$$

when H on the right is given the equivalent norm it receives from the K-interpolation method. When $\Theta = (m-k)/m$, Proposition 2.4(i) says $B_{22}^{(1-\Theta)m} = W_2^k$. Since Proposition 2.4(iii) shows $T(p) = 0$ for p in Π_m, T may be considered as a map from W_2^k/Π_k. Hence (2.6) follows from 2.4(ii) and the preceding inequality, once α is set to account for the K-interpolation and other norm equivalences. ∎

For bounds on the accuracy of the Tikhonov regularizer as an estimate of derivatives we have

THEOREM 2.10. *For $j \leqslant k \leqslant m$ there exist constants $\beta = \beta(m, k, j)$ such that for f in $W_2^k[a, b]$*

$$|f - S_{\lambda,m}f|_j^2 \leqslant \beta\lambda^{(k-j)/m}\, |f|_k^2.$$

Proof. Again we restrict attention to $[0, 1]$. Now we start by noting that the minimizing property of $S_{\lambda,m}f$ is equivalent to the fact that $T(f)$, as defined in (2.8) above, is orthogonal to $(h, \lambda^{1/2}h^{(m)})$ in $L_2 \oplus L_2$ for all h in W_2^m. In particular when f is in W_2^m we may set $h = f - S_{\lambda,m}f$ to get

$$|f - S_{\lambda,m}f|_0^2 - \lambda\langle f - S_{\lambda,m}f, S_{\lambda,m}f\rangle_m = 0.$$

Hence

$$|f|_m^2 = |f - S_{\lambda,m} f|_m^2 + |S_{\lambda,m} f|_m^2 + (2/\lambda)\, |f - S_{\lambda,m} f|_0^2$$

which implies

$$|f - S_{\lambda,m} f|_m^2 \leqslant |f|_m^2. \tag{2.11}$$

When we add to this the corresponding inequality with $m = 0$, which follows from Eq. (2.7), the result implies that the map $I - S_{\lambda,m}$: $W_2^m \to W_2^m$ has norm one. The norm of $I - S_{\lambda,m}$ considered as a map between B_{22}^s and W_2^0, however, can be bounded by $\lambda^{s/2m}$ from (2.9). So if we apply the interpolation theorems and [3, (6.4.5)] to the map $I - S_{\lambda,m}$ between the pairs (B_{22}^s, W_2^m) and (W_2^0, W_2^m), with $\Theta = (m-k)/m$, and $s = (k-j)\, m/(m-k)$, we deduce that

$$I - S_{\lambda,m}\colon B_{22}^k \to B_{22}^j, \qquad \|I - S_{\lambda,m}\|^2 \leqslant \lambda^{(k-j)/m}.$$

Again the fact that for s integral $B_{22}^s = W_2^s$ can be used to yield

$$\|f - S_{\lambda,m} f\|_j^2 \leqslant \beta \lambda^{(k-j)/m}\, \|f\|_k^2$$

when β is set to account for the various norm equivalences. Since $I - S_{\lambda,m} = 0$ on Π_m, by Proposition 2.4(iii), $I - S_{\lambda,m}$ maps W_2^k/Π_k into W_2^j. Hence $\|\;\|_k$ on the right can be replaced by $|\;|_k$, in view of 2.4(ii). This yields the desired result Theorem 2.10, once we note that $|\;|_j$ is dominated by $\|\;\|_j$. ∎

3. Relations Between Discrete and Standard L_2-Norms

In order to extend the results of the last section from the Tikhonov regularizer to the spline smoothing operator, we shall need a number of results which replace some of the intermediate space theorems we have used. These relate the Euclidean norm of the sample $\mathbf{f}_\Delta$ and the various W_2^m seminorms of f. In some ways the results we need resemble the more classical interpolation theorems in the standard theory of Sobolev spaces (e.g., [2, Sect. 3]). The statements and proofs of these results are collected here as they may be of independent interest.

All of the estimates to follow depend on the global mesh ratio of the partition Δ of $[a, b]$. This is measured by $\bar{\Delta}/\underline{\Delta}$ whose entries are defined by

$$\bar{\Delta} = \max\{x_{i+1} - x_i, x_1 - a, b - x_n\}, \qquad \underline{\Delta} = \min\{x_{i+1} - x_i\}. \tag{3.1}$$

Our first basic result quantifies, in a form useful for our subsequent work, the obvious fact that $f^{(m)}$ and $f(x_i)$, $i = 1,\ldots, n \geqslant m$ determine $f^{(k)}$, $k < m$.

THEOREM 3.2. *Let $\Delta = \{x_1, \ldots, x_n\}$ be a partition of $[a, b]$. For any integers $k < m \leqslant n$ there exist constants $C = C(m, k, \bar{\Delta}/\underline{\Delta})$, $D = D(m, k) \geqslant 1$ such that for any f in W_2^m*

$$\bar{\Delta}^{2k} |f|_k^2 \leqslant C\bar{\Delta} \|\mathbf{f}_\Delta\|^2 + D\bar{\Delta}^{2m} |f|_m^2, \tag{3.3}$$

where C is a polynomial in $\bar{\Delta}/\underline{\Delta}$ of degree $2(m-1)$. In particular for $m = k+1$, $C = C_k(\bar{\Delta}/\underline{\Delta})^{2k}$, and

$$\bar{\Delta}^{2k} |f|_k^2 \leqslant C_k(\bar{\Delta}/\underline{\Delta})^{2k} \bar{\Delta} \|\mathbf{f}_\Delta\|^2 + D\bar{\Delta}^{2(k+1)} |f|_{k+1}^2. \tag{3.4}$$

Remark. In later work we shall denote the $k = 0$ constants by $C(m, \bar{\Delta}/\underline{\Delta})$ and $D(m)$.

Proof. If the second inequality (3.4) is true for all k, then the first follows by induction on $m > k$. The induction step is achieved by estimating the second term in (3.3) via (3.4) with $k = m$.

To prove (3.4), first consider any interval $[y, z]$ containing at least $k+1$ of the x_i's, say $x_1, \ldots, x_{k+1}$. Then for some η in $[y, z]$ we have the equality $f[x_1, \ldots, x_{k+1}] = f^{(k)}(\eta)/k!$, where $f[x_1, \ldots, x_{k+1}]$ is the kth divided difference. Hence,

$$f^{(k)}(x) = k!\, f[x_1, \ldots, x_{k+1}] + \int_\eta^x f^{(k+1)}.$$

Since it is easy to prove by induction that

$$|f[x_1, \ldots, x_{k+1}]| \leqslant \left(\sum_{j=0}^{k} \binom{k}{j} |f(x_{j+1})| \right) \Big/ k!\, \underline{\Delta}^k \tag{3.5}$$

it follows from the previous two inequalities, via Cauchy–Schwarz, that for all x in $[a, b]$

$$|f^{(k)}(x)|^2 \leqslant 2 \binom{2k}{k} \underline{\Delta}^{-2k} \sum_{j=0}^{k} f^2(x_{j+1}) + 2\,|x - \eta| \left| \int_\eta^x |f^{(k+1)}|^2 \right|,$$

where we have used the fact that $\sum_{j=0}^{k} \binom{k}{j}^2 = \binom{2k}{k}$. Hence integrating from y to z yields

$$\int_y^z |f^{(k)}(x)|^2 \leqslant 2 \binom{2k}{k} \underline{\Delta}^{-2k}(z-y) \sum_{y \leqslant x_i \leqslant z} f^2(x_i) + (z-y)^2 \int_y^z |f^{(k+1)}|^2. \tag{3.6}$$

Now choose $y_0 = a < y_1 < \cdots < y_l = b$ such that (i) $(k+1)\bar{\Delta} \leqslant y_{i+1} - y_i < 2(k+1)\bar{\Delta}$ and (ii) $x_j \neq y_i$ for all j and all $i \neq 0, l$. Then each $[y_i, y_{i+1}]$ contains at least $k+1$ x_j's by (i) and the definition of $\bar{\Delta}$. Hence (3.6) can be applied to each interval $[y_i, y_{i+1}]$, for $i = 1, l-1$. Since each x_j occurs in exactly one of the resulting inequalities, when we sum over i and estimate $y_{i+1} - y_i$ by $2(k+1)\bar{\Delta}$, we get

$$|f|_k^2 \leqslant 4(k+1)\binom{2k}{k}\bar{\Delta}\underline{\Delta}^{-2k}\sum_{j=1}^{n} f^2(x_j) + 4(k+1)^2\bar{\Delta}^2|f|_{k+1}^2.$$

Multiplication by $\bar{\Delta}^{2k}$ yields (3.5). ∎

Notice that all the dependence of the constants has been made very explicit. In particular none of the constants depend on the interval length $b-a$ or on n, the number of x_i's. Also note that the case when $k=0$ gives estimates for the l_2 norm in terms of discrete L_2 and W_2^m norms.

In the opposite direction we can estimate the discrete l_2 norm by W_2^m type (semi-) norms. Specifically we have the following inequalities:

THEOREM 3.7. *Let Δ be a partition of $[a, b]$ with $n \geqslant k \geqslant 1$. There exist constants $E = E(k, \bar{\Delta}/\underline{\Delta})$ and $F = F(k)$ such that*

$$\bar{\Delta}\,\|\mathbf{f}_\Delta\|^2 \leqslant E\,|f|_0^2 + F\bar{\Delta}^{2k}\,|f|_k^2. \tag{3.8}$$

Proof. For $k=1$ a standard Sobolev-type argument works. Set $y_1 = a$, $y_i = (x_{i-1} + x_i)/2$. $2 \leqslant i \leqslant n$, and $y_{n+1} = b$. Then x_i is in $[y_i, y_{i+1}]$ and $f(x_i) = f(x) + \int_x^{x_i} f^{(1)}$. Hence

$$f^2(x_i) \leqslant 2f^2(x) + 2\,|y_{i+1} - y_i| \int_{y_i}^{y_{i+1}} (f^{(1)})^2.$$

Integration over $[y_i, y_{i+1}]$ yields

$$(y_{i+1} - y_i) f^2(x_i) \leqslant 2\int_{y_i}^{y_{i+1}} f^2 + 2(y_{i+1} - y_i)^2 \int_{y_i}^{y_{i+1}} (f^{(1)})^2.$$

Now we divide by $(y_{i+1} - y_i)$, sum over i, and use the estimates $\frac{1}{2}\underline{\Delta} \leqslant \min\{y_{i+1} - y_i\}$ and $\max\{y_{i+1} - y_i\} \leqslant \frac{3}{2}\bar{\Delta}$ to get

$$\|\mathbf{f}_\Delta\|^2 \leqslant 4\underline{\Delta}^{-1}\,|f|_0^2 + 3\bar{\Delta}\,|f|_1^2.$$

This becomes (3.8) for $k=1$ when we multiply by $\bar{\Delta}$.

To prove (3.8) for $k > 1$ we just use Lemma 3.9 with $t = \bar{\Delta}$ to estimate the $|f|_1^2$ term in the $k=1$ case by $\bar{\Delta}^{2k}\,|f|_k^2$. The lemma is just a version of the standard interpolation inequalities. ∎

The interpolation inequality we need is

LEMMA 3.9. *There exist constants* $\gamma=\gamma(m,k)$ *such that for any* f *in* $W_2^m[a,b]$ *and any* $t\leqslant(b-a)$

$$t^{2k}\,|f|_k^2\leqslant\gamma(|f|_0^2+t^{2m}\,|f|_m^2).$$

Proof. This is essentially [2, Theorem 3.3] except for making γ independent of $b-a$. To achieve this just use this theorem when $[a,b]=[0,1]$ to get γ. To see that the same constant works on $[a,b]$ apply the $[0,1]$ case to $h(y)=f(x)$, $x=a+y(b-a)$, with t replaced by $t/(b-a)$ which is less than one as required in Agmon's theorem. ∎

4. CONVERGENCE RATES FOR EXACT DATA

This section contains estimates for the spline smoothing operator analogous to those for the Tikhonov regularizer in Theorem 2.10. These show that for f in W_2^k, $1\leqslant k\leqslant m$, the smoothing splines $S_{n,\lambda,m}(\mathbf{f}_\Delta)$ based on exact data converge to f, as $n\to\infty$, $\lambda\to 0$. Moreover, they also show that derivatives of *order less than* k of the smoothing splines converge to the corresponding derivatives of f. These convergence results require that the partitions be quasi-uniform, i.e., the mesh ratios $\bar{\Delta}_n/\underline{\Delta}_n$ remain bounded. Our theorems give convergence rates which are similar to those from Theorem 2.10 once λ is replaced by a quantity $L\sim\lambda+\bar{\Delta}^{2m}$.

Most of the estimates we prove depend on minimizing property (2.1) used to define the smoothing spline $S_{n,\lambda,m}(\mathbf{f}_\Delta)$. Some useful consequences of this minimizing property are in

LEMMA 4.1. *Given* $\mathbf{y}$ *in* R^n *the residual vector* $\mathbf{y}-\mathbf{S}_{n,\lambda,m}(\mathbf{y})_\Delta$ *satisfies*

$$\frac{b-a}{n}<\mathbf{y}-\mathbf{S}_{n,\lambda,m}(\mathbf{y})_\Delta,\mathbf{h}_\Delta\rangle=\lambda\langle S_{n,\lambda,m}(\mathbf{y}),h\rangle_m \tag{4.2}$$

for all h *in* $W_2^m[a,b]$. *In particular*

$$\frac{b-a}{n}\langle\mathbf{y}-\mathbf{S}_{n,\lambda,m}(\mathbf{y})_\Delta,\mathbf{S}_{n,\lambda,m}(\mathbf{y})_\Delta\rangle=\lambda\,|S_{n,\lambda,m}(\mathbf{y})|_m^2 \tag{4.3}$$

and when f *is in* $W_2^m[a,b]$

$$\frac{b-a}{n}\|\mathbf{f}_\Delta-\mathbf{S}_{n,\lambda,m}(\mathbf{f}_\Delta)_\Delta\|^2=\lambda\langle S_{n,\lambda,m}(\mathbf{f}_\Delta),f-S_{n,\lambda,m}(\mathbf{f}_\Delta)\rangle_m. \tag{4.4}$$

Proof. If a Hilbert space norm (and inner product) is defined on $R^n \oplus L_2$ by

$$\|(\mathbf{y}, h)\|^2 = \frac{b-a}{n} \|\mathbf{y}\|^2 + \lambda |h|_0^2,$$

then (4.2) says $(\mathbf{y}, 0) - (\mathbf{S}_{n,\lambda,m}(\mathbf{y})_\Delta, (S_{n,\lambda,m}(\mathbf{y}))^{(m)})$ is orthogonal (in $R^n \oplus L_2$) to $(\mathbf{h}_\Delta, h^{(m)})$, for all h in W_2^m. Standard Hilbert space projection arguments say this is equivalent to $g_{n,\lambda,m} = S_{n,\lambda,m}(\mathbf{y})$ in W_2^m minimizes the norm of $(\mathbf{y}, 0) - (\mathbf{g}_\Delta, g^{(m)})$ over all g in W_2^m, which was our definition of $S_{n,\lambda,m}(\mathbf{y})$. Now (4.3) is immediate and (4.4) follows since f in W_2^m means $h = f - S_{n,\lambda,m}(\mathbf{f}_\Delta)$ can be substituted into (4.2) when $\mathbf{y} = \mathbf{f}_\Delta$. ∎

This result can be used to give two further equalities, one of which, Eq. (4.7), is the analog of the first integral relation for interpolating splines (see [17, Theorem 3.2]).

PROPOSITION 4.5. *For* $\mathbf{y}$ *in* R^n *and* f *in* $W_2^m[a, b]$,

$$\|\mathbf{y}\|^2 = \|\mathbf{y} - \mathbf{S}_{n,\lambda,m}(\mathbf{y})_\Delta\|^2 + \|\mathbf{S}_{n,\lambda,m}(\mathbf{y})_\Delta\|^2 + 2\frac{n\lambda}{b-a} |S_{n,\lambda,m}(\mathbf{y})|_m^2 \tag{4.6}$$

and

$$|f|_m^2 = |f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_m^2 + |S_{n,\lambda,m}(\mathbf{f}_\Delta)|_m^2 + \frac{2}{\lambda}\frac{b-a}{n} \|\mathbf{f}_\Delta - \mathbf{S}_{n,\lambda,m}(\mathbf{f}_\Delta)_\Delta\|^2. \tag{4.7}$$

Proof. Write $\mathbf{y} = (\mathbf{y} - \mathbf{S}_{n,\lambda,m}(\mathbf{y})) + \mathbf{S}_{n,\lambda,m}(\mathbf{y})$ and $f = f - S_{n,\lambda,m}(\mathbf{f}_\Delta)) + S_{n,\lambda,m}(\mathbf{f}_\Delta)$ and expand the left side. Use (4.3) and (4.4) to replace the middle inner product terms by norm terms. ∎

The first part of this proposition allows us to make the estimate

$$|S_{n,\lambda,m}(\mathbf{y})|_m^2 \leqslant \frac{b-a}{2n\lambda} \|\mathbf{y}\|^2. \tag{4.8}$$

For small λ, however, this is a poor bound and can be replaced using

PROPOSITION 4.9. *For any partition* Δ *of* $[a, b]$ *there exists a constant* $G(m, \Delta)$ *independent of* λ, *such that*

$$|S_{n,\lambda,m}(\mathbf{y})|_m^2 \leqslant G(m, \Delta) \|\mathbf{y}\|^2.$$

In fact

$$G(m, \Delta) \leqslant \delta(m)\, \bar{\Delta}\underline{\Delta}^{-2m}$$

for some constant $\delta(m)$.

Proof. The existence of a constant dependent on λ follows from the continuity of $S_{n,\lambda,m}: R^n \to W_2^m$. If, however, we substitute the *interpolating* spline, $g = S_{n,0,m}(\mathbf{y})$, in minimization equation (2.1) defining the smoothing spline $S_{n,\lambda,m}(\mathbf{y})$, then it follows that

$$\lambda\, |S_{n,\lambda,m}(\mathbf{y})|_m^2 \leqslant \frac{b-a}{n} \|\mathbf{y} - \mathbf{S}_{n,\lambda,m}(\mathbf{y})\|^2 + \lambda\, |S_{n,\lambda,m}(\mathbf{y})|_m^2$$
$$\leqslant \lambda\, |S_{n,0,m}(\mathbf{y})|_m^2.$$

So $G(m, \Delta)$ may be taken as the norm of the interpolation operator $S_{n,0,m}$ which is independent of λ.

The bound on $G(m, \Delta)$ for *uniform* Δ follows from the last inequality in Schoenberg [15]. For more general partitions it follows from de Boor [4]. In particular the inequality in [4, p. 115] implies that if $f = S_{n,0,m}(\mathbf{y})$, then for some constant M depending on m

$$|S_{n,0,m}(\mathbf{y})|_m^2 \leqslant M\bar{\Delta} \sum_{j=1}^{n-m} (f[x_j, \ldots, x_{j+m}])^2.$$

But estimate (3.5) for the divided differences shows that

$$\sum_{j=1}^{n-m} (f[x_j, \ldots, x_{j+m}])^2 \leqslant (m+1) \binom{2m}{m} (m!)^{-2}\, \underline{\Delta}^{-2m} \sum_{j=1}^{n} f^2(x_j).$$

These two inequalities combine to give the desired result. ∎

From these simple propositions and the estimates in Theorem 3.2 we can derive the first main theorem on error bounds for spline smoothing.

THEOREM 4.10. *Given any partition* Δ *of* $[a, b]$ *let*

$$L = C(m, \bar{\Delta}/\underline{\Delta}) \frac{n\bar{\Delta}}{b-a} \frac{\lambda}{2} + D(m)\, \bar{\Delta}^{2m}, \tag{4.11}$$

where $C(m, \bar{\Delta}/\underline{\Delta})$ *and* $D(m)$ *are as in Theorem* 3.2. *Then given integers* $0 \leqslant k \leqslant m$ *there exist constants* $H = H(m, k)$ *such that for any* f *in* $W_2^m[a, b]$,

$$|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_k^2 \leqslant H(1 + (L/(b-a)^{2m}))^{k/m}\, L^{(m-k)/m}\, |f|_m^2. \tag{4.12}$$

Remark. For $\lambda = 0$ this gives error bounds for the *natural* spline interpolant which have the same dependence on $\bar{\Delta}$ as those in Schultz [17, Theorem 3.4].

Proof. As for the Tikhonov regularizer it will suffice to prove the theorem on the interval $[0, 1]$ since under the standard change of variables, Proposition 2.3, both λ and $\bar{\Delta}^{2m}$ are scaled by $(b-a)^{2m}$. We proceed to the proof under the assumption that $[a, b] = [0, 1]$.

Now for $k = 0$ estimate (3.3) shows that

$$|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_0^2 \leqslant C(m, \bar{\Delta}/\underline{\Delta})\, \bar{\Delta}\, \| \mathbf{f}_\Delta - \mathbf{S}_{\mathbf{n},\lambda,\mathbf{m}}(\mathbf{f}_\Delta)_\Delta \|^2 + D(m)\, \bar{\Delta}^{2m}\, |f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_m^2 .$$

But (4.7) allows us to estimate each summand in terms of $|f|_m^2$ and leads to

$$|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_0^2 \leqslant L\, |f|_m^2, \qquad L = C(m, \bar{\Delta}/\underline{\Delta})\, n\bar{\Delta}\lambda/2 + D(m)\, \bar{\Delta}^{2m} \tag{4.13}$$

which is (4.12) for $k = 0$ with $H(m, 0) = 1$.

For $k = m$, (4.7) also shows that $|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_m^2 \leqslant |f|_m^2$ which is better than the desired result with $H(m, m) = 1$. When we add the $k = 0$ case we get

$$\|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)\|_m^2 \leqslant (1 + L)\, |f|_m^2 . \tag{4.14}$$

Now if we define the map T by $T(f) = f - S_{n,\lambda,m}(\mathbf{f}_\Delta)$, then Proposition 2.4(iii) shows that T is a map *from* W_2^m/Π_m, while 2.4(ii) implies that (4.13) and (4.14) give norm estimates for T as a map *into* W_2^m and W_2^0, respectively. The interpolation results from [3] can be applied to conclude

$$T\colon W_2^m/\Pi_m \to B_{22}^{(1-\Theta)m}, \qquad \|T\|^2 \leqslant L^\Theta (1 + L)^{1-\Theta}. \tag{4.15}$$

Hence, when $\Theta = (m-k)/m$ the equality $B_{22}^k = W_2^k$, 2.4(i), shows

$$|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_k^2 \leqslant \|Tf\|_k^2 \leqslant H(m, k)(1 + L)^{k/m}\, L^{m-k/m}\, |f|_m^2,$$

where $H(m, k)$ accounts for the norm equivalences in the interpolation theorems. ∎

This theorem shows that given exact samples from f in $W_2^m[a, b]$ the spline smoothing operators produce good approximations to f and to its derivatives of order less than m. For instance, for quasi-uniform partitions we have the convergence result of

COROLLARY 4.16. *If a sequence of n-point partitions Δ_n of $[a, b]$ satisfies $\bar{\Delta}_n/\underline{\Delta}_n \leqslant r$, all n, then for any f in $W_2^m[a, b]$*

$$|f - S_{n,\lambda,m}(\mathbf{f}_{\Delta_n})|_k^2 \leqslant O((\lambda + ((b-a)/n)^{2m})^{(m-k)/m})\, |f|_m^2$$

as $\lambda \to 0$, $n \to \infty$. *The coefficient in the O relation depends only on m, k, and r.*

Proof. The quasi-uniformity, $\bar{\Delta}_n/\underline{\Delta}_n \leqslant r$, implies that $\bar{\Delta}_n = O((b-a)/n)$ and $C(m, \bar{\Delta}/\underline{\Delta}) = O(1)$. Hence this corollary follows directly from Theorem 4.10, since $L \to 0$ as $\lambda \to 0$ and $n \to \infty$. ∎

Even when f has fewer than m derivatives, there are convergence rate results similar to the ones just proved. Unfortunately the simple interpolation proofs of Section 2 for $S_{\lambda,m}$ do not carry over directly. The problem is that the analog of the bound $|S_{\lambda,m}f|_0^2 \leqslant |f|_0^2$ is the discrete norm inequality $\|\mathbf{S}_{n,\lambda,m}(\mathbf{f}_\Delta)_\Delta\|^2 \leqslant \|\mathbf{f}_\Delta\|^2$, which can only lead to estimates of the continuous L_2-type norms by introducing some *derivative* information. Once this is done, however, most of the natural estimates follow.

We begin with the W_2^0 estimates when f has $k(<m)$ derivatives.

THEOREM 4.17. *Given integers* $0 < k < m$ *there exist constants* $J = J(m, k, \bar{\Delta}/\underline{\Delta})$ *and* $K = K(m, k, \bar{\Delta}/\underline{\Delta})$ *such that for any n-point partition* Δ *of* $[a, b]$ *and any* f *in* $W_2^k[a, b]$, *if* $G = G(m, \Delta)$ *as in Proposition* 4.9 *and*

$$L = C(m, \bar{\Delta}/\underline{\Delta}) \frac{n\bar{\Delta}}{b-a} \frac{\lambda}{2} + D(m)\,\bar{\Delta}^{2m},$$

then

$$|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_0^2 \leqslant \left(J + K\bar{\Delta}^{2m-1} \min\left(G, \frac{b-a}{2n\lambda}\right)\right) L^{k/m}\, |f|_k^2.$$

Proof. We exploit our previous error bounds for functions in $W_2^m[a, b]$ as follows. Let g in $W_2^m[a, b]$ be arbitrary for the moment. Then adding and subtracting appropriate terms in g and $S_{n,\lambda,m}(\mathbf{g}_\Delta)$ to the left-hand side in the desired inequality and applying the triangle and Cauchy–Schwarz inequalities yields

$$\begin{aligned} |f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_0^2 \leqslant 3(&|f-g|_0^2 + |g - S_{n,\lambda,m}(\mathbf{g}_\Delta)|_0^2 \\ &+ |S_{n,\lambda,m}((\mathbf{f}-\mathbf{g})_\Delta)|_0^2). \end{aligned} \tag{4.18}$$

We shall work on estimates for the last two terms.

The middle term can be estimated using our basic mth order error bounds from (4.13) which say

$$|g - S_{n,\lambda,m}(\mathbf{g}_\Delta)|_0^2 \leqslant L\,|g|_m^2, \qquad L = C(m, \bar{\Delta}/\underline{\Delta}) \frac{n\bar{\Delta}}{b-a} \frac{\lambda}{2} + D(m)\,\bar{\Delta}^{2m}. \tag{4.19}$$

To handle the last term in (4.18) let $\mathbf{y} = (\mathbf{f} - \mathbf{g})_\Delta$. Then the continuous-discrete norm estimates of (3.3) show

$$|S_{n,\lambda,m}((\mathbf{f} - \mathbf{g})_\Delta)|_0^2 \leqslant C(m, \bar{\Delta}/\underline{\Delta})\,\bar{\Delta}\,\|\mathbf{S}_{n,\lambda,m}(\mathbf{y})_\Delta\|^2 + D(m)\,\bar{\Delta}^{2m}\,|S_{n,\lambda,m}(\mathbf{y})|_m^2.$$

Also Eq. (4.8) and Proposition 4.9 allow us to estimate each summand on the right in terms of $\|\mathbf{y}\|^2$. These estimates yield

$$|S_{n,\lambda,m}(\mathbf{y})|_0^2 \leqslant B\bar{\Delta}\,\|\mathbf{y}\|^2,$$

$$B = C(m, \bar{\Delta}/\underline{\Delta})\,\bar{\Delta} + D(m)\,\bar{\Delta}^{2m-1} \min\left(G, \frac{b-a}{2n\lambda}\right). \tag{4.20}$$

Now the Euclidean norm of $\mathbf{y} = (\mathbf{f} - \mathbf{g})_\Delta$ can be estimated in terms of L_2-norms via Theorem 3.7 to get

$$\bar{\Delta}\,\|\mathbf{y}\|^2 \leqslant E(k, \bar{\Delta}/\underline{\Delta})\,|f-g|_0^2 + F(k)\,\bar{\Delta}^{2k}\,|f-g|_k^2.$$

When we combine the last two inequalities and use the result along with (4.19) in (4.18) we get

$$|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_0^2 \leqslant 3((1 + E(k, \bar{\Delta}/\underline{\Delta})\,B)\,|f-g|_0^2 + L\,|g|_m^2 + F(k)\,B\,|f-g|_k^2).$$

Now we can choose $g = S_{L,m}f$, the Tikhonov regularizer of f, and apply the estimate of Theorem 2.5 for $|f - S_{L,m}f|_0^2$ and $L\,|S_{L,m}f|_m^2$ and of (2.10) for $|f - S_{L,m}f|_k^2$ to get

$$|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_0^2 \leqslant 3((1 + E(k, \bar{\Delta}/\underline{\Delta})\,B)\,\alpha L^{k/m} + \alpha L^{k/m} + F(k)\,\beta B\bar{\Delta}^{2k})\,|f|_k^2,$$

where $\alpha = \alpha(m,k)$, $\beta = \beta(m,k)$ as in Theorems 2.5 and 2.10. Since $\bar{\Delta}^{2k} \leqslant L^{k/m}$ as $D(m) \geqslant 1$, the preceding inequality yields the desired result with $J = 3(2\alpha + (\beta F(k) + \alpha E(k, \bar{\Delta}/\underline{\Delta}))\,C(m, \bar{\Delta}/\underline{\Delta})$, and $K = 3(\beta F(k) + \alpha E(k, \bar{\Delta}/\underline{\Delta}))$, once eqs. (4.19) and (4.20) for L and B are considered. ∎

Now we can apply interpolation arguments to get error bounds for derivative estimates for spline smoothing when f has fewer than m derivatives.

THEOREM 4.21. *Given integers $1 \leqslant j < k < m$ there exist constants $M = M(m,k,j)$ such that for any partition Δ of $[a,b]$, and any f in $W_2^k[a,b]$, if $J = J(m, 1, \bar{\Delta}/\underline{\Delta})$, $K = K(m, 1, \bar{\Delta}/\underline{\Delta})$ as in Theorem* 4.17, *$G = G(m, \Delta)$ as in Proposition* 4.9, *and L is as in* (4.11), *then*

$$|f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_j^2 \leqslant M \left\{J + K\bar{\Delta}^{2m-1}\min\left(G, \frac{b-a}{2n\lambda}\right)\right\}^{(m-k)/(k-1)}$$
$$\times\,(1 + L/(b-a)^{2m})^{j/m}\,L^{(k-j)/m}\,|f|_k^2.$$

Proof. As usual we restrict ourselves to $[a, b] = [0, 1]$. If $T(f) = f - S_{n,\lambda,m}(\mathbf{f}_\Delta)$ as in Theorem 4.10, then replacing $|f|_1^2$ by the larger quantity $\|f\|_1^2$ in the $k = 1$ case of Theorem 4.17 implies

$$T: W_2^1 \to W_2^0, \qquad \|T\|^2 \leqslant (J + K\bar{\Delta}^{2m-1} \min(G, 1/2n\lambda)) L^{1/m}.$$

A similar replacement of $|f|_m^2$ by $\|f\|_m^2$ in norm estimate (4.15) for T says that

$$T: W_2^m \to B_{22}^s, \qquad \|T\|^2 \leqslant (1 + L)^{s/m} L^{(m-s)/m}.$$

Now all the standard K-method interpolation and norm replacement arguments we have made, repeatedly, yield the desired results when $\Theta = (m - k)/(m - 1)$ and $s = j(m - 1)/(k - 1)$. ∎

This gives new convergence rates for spline smoothing and spline interpolation.

THEOREM 4.22. *Suppose $\{\Delta_n\}$ is a sequence of n-point partitions of $[a, b]$ which are quasi-uniform, i.e., $\bar{\Delta}_n/\underline{\Delta}_n \leqslant r$ for some fixed $r \geqslant 1$. Then for f in $W_2^k[a, b]$, and $0 \leqslant j < k \leqslant m$, the smoothing spline $S_{n,\lambda,m}(\mathbf{f}_\Delta)$ satisfies*

$$|f - S_{n,\lambda_n,m}(\mathbf{f}_\Delta)|_j^2 = O((\lambda_n + \bar{\Delta}_n^{2m})^{(k-j)/m}) \, |f|_k^2, \tag{4.23}$$

where the O coefficient depends only on m, k, j, and r. In particular the interpolating spline $S_{n,0,m}(\mathbf{f}_\Delta)$ satisfies

$$|f - S_{n,0,m}(\mathbf{f}_\Delta)|_j^2 = O(\bar{\Delta}_n^{2(k-j)}) \, |f|_k^2 \tag{4.24}$$

as $n \to \infty$.

Proof. If we recall the arguments in Corollary 4.16 we see that the quasi-uniformity guarantees that each of the constants J, K, C in our previous estimates which involve $\bar{\Delta}/\underline{\Delta}$ are bounded independent of n. Moreover, the quantity L (in (4.11)) is such that $L = O(\lambda_n + \bar{\Delta}_n^{2m})$. The estimates for $G(m, \Delta_n)$ in Proposition 4.9 and the quasi-uniformity guarantee that the factor multiplying K in Theorem 4.21 is bounded. Parts (i) and (ii) follow when these remarks are applied to Theorems 4.10, 4.17, and 4.21. ∎

5. ERROR BOUNDS FOR INEXACT DATA

Real data samples from a function f normally contain some random noise, even when all systematic errors have been corrected for. In this situation the spline interpolants ($\lambda = 0$) are greatly affected (consider the size of $G(m, \Delta)$ in Proposition 4.9 when $\underline{\Delta}$ is small) and may well cease to provide the good estimates they produce for exact data. This is even more dramatically true for derivative estimates. It is in this context of noisy data that the spline smoothing operators, with an appropriate choice of λ, demonstrate their power. In this section we shall show the types of convergence rates which can be achieved, on the average, in the case of data subject to random noise.

Let us fix the following model for the errors in the data:

$$\mathbf{y} = \mathbf{f}_\Delta + \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} = [\varepsilon_i],$$

where ε_i are uncorrelated, mean zero random variables with a common variance σ^2, i.e.,

$$E(\varepsilon_i) = 0, \qquad E(\varepsilon_i \varepsilon_j) = \delta_{ij}\sigma^2,$$

with E standing for expectation. With this model it follows from the linearity of $S_{n,\lambda,m}$ that the expected value of the approximation errors for noisy data satisfy the relationship

$$E(|f - S_{n,\lambda,m}(\mathbf{f}_\Delta + \boldsymbol{\varepsilon})|_k^2) = |f - S_{n,\lambda,m}(\mathbf{f}_\Delta)|_k^2 + E(|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_k^2). \tag{5.1}$$

So the problem of determining the expected errors and convergence rates for noisy data is reduced to the previously derived estimates, together with the problem of estimating the second term above.

Now for random data $\boldsymbol{\varepsilon}$ the smoothing nature of $S_{n,\lambda,m}$ with $\lambda > 0$ acts to force $S_{n,\lambda,m}(\varepsilon)$ to be closer to zero on the average than would be expected from the average size of $\|\varepsilon\|^2$. A precise version of this fact can be derived in the case of a uniform partition from some estimates due to Craven and Wahba in [5].

Let us denote by $A(\lambda)$ the symmetric linear transformation defined for $\mathbf{y}$ in R^n by $A(\lambda)\,\mathbf{y} = \mathbf{S}_{n,\lambda,m}(\mathbf{y})_\Delta$. (The symmetry of $A(\lambda)$ follows from (4.2) with $h = S_{n,\lambda,m}(\mathbf{z})$.) We need estimates for $A(\lambda)$ from [5, Lemma 4.3] which we restate in a suitably altered form as

PROPOSITION 5.2. *There exist constants $M_j(m)$, $j = 1, 2$, such that when Δ is the uniform n-point partition of* $[0, 1]$ *and $\lambda \leqslant 1$, then*

$$\operatorname{Tr}(A^j(\lambda)) \leqslant M_j(m)/\lambda^{1/2m}.$$

Proof. This follows from the proof of [5, Lemma 4.3], in particular from p. 401 once the restriction $\lambda \leqslant 1$ is considered. As Dennis Cox kindly pointed out the (heuristic) arguments in that lemma, specifically the approximate equalities on p. 402, are exact in the uniform partition case ($w(t) = 1$), once one takes $D_{ll} = \sum_l' 1/(2\pi\nu)^{2m}$, where $\sum_l'$ is the sum over $\nu \neq 0$, $\nu \equiv l \bmod n$. (Utreras [21, 22] has given another approach toward removing the lack of rigor in [5, Lemma 4.3].) ∎

Now we estimate the expected size of $S_{n,\lambda,m}(\boldsymbol{\varepsilon})$.

THEOREM 5.3. *For any uniform n-point partition Δ of $[a, b]$ and any $\lambda > 0$ let*

$$N = C(m, \bar{\Delta}/\underline{\Delta}) \frac{n\bar{\Delta}}{b-a} M_2(m) + D(m) M_1(m) \bar{\Delta}^{2m}/\lambda.$$

Then provided $\lambda \leqslant (b-a)^{2m}$

$$E(|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_m^2) \leqslant \sigma^2 (b-a)^2 M_1(m)/n\lambda^{(2m+1)/2m}, \tag{5.4}$$

$$E(|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_0^2) \leqslant \sigma^2 (b-a)^2 N/n\lambda^{1/2m}, \tag{5.5}$$

and more generally for $0 < k < m$

$$E(|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_k^2) \leqslant \sigma^2 (b-a)^2 (N + M_1(m)) \gamma/n\lambda^{(2k+1)/2m}, \tag{5.6}$$

where $\gamma = \gamma(m, k)$ as in Lemma 3.9.

Remark. Bounded estimates for very small values of λ could be derived using Proposition 4.9 but we shall not need these.

Proof. We prove this using the interpolation Lemma 3.9 which we adapted from [2] starting from the extreme cases $k = m$ and $k = 0$. Moreover, as usual we assume $[a, b] = [0, 1]$. For $k = m$ we use the definition of $A(\lambda)$ and (4.3) to get

$$|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_m^2 = \frac{1}{n\lambda} \langle (I - A(\lambda))\, \boldsymbol{\varepsilon}, A(\lambda)\, \boldsymbol{\varepsilon} \rangle.$$

So taking expectations and using the mean and correlation properties of the ε_i yields

$$E(|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_m^2) = \frac{\sigma^2}{n\lambda} \operatorname{Tr}((I - A^t(\lambda)) A(\lambda)) \leqslant \frac{\sigma^2}{n\lambda} \operatorname{Tr}(A(\lambda)). \tag{5.7}$$

Now use the $j = 1$ case of the preceding proposition to deduce (5.4) from (5.7).

As usual for the $k = 0$ case we turn to the continuous-discrete norm estimates in Theorem 3.2 which show that

$$|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_0^2 \leqslant C(m)\, \bar{\Delta}\, \|A(\lambda)\, \boldsymbol{\varepsilon}\|^2 + D(m)\, \bar{\Delta}^{2m}\, |S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_m^2.$$

When we take expectations and use the previous case, (5.4), we find

$$E(|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_0^2) \leqslant \sigma^2\{C(m)\,\bar{\Delta}\,\mathrm{Tr}(A^2(\lambda)) + D(m)\,\bar{\Delta}^{2m}M_1(m)/n\lambda^{(2m+1)/2m}\}.$$

Once we apply the $j=2$ case of Proposition 5.2 we get (5.5).

Now we use the interpolation inequality from Lemma 3.9 with $t=\lambda^{1/2m}$ to deduce that

$$E(|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_k^2) \leqslant \gamma(\lambda^{-k/m}E(|S_{n,\lambda,m}(\boldsymbol{\varepsilon})|_0^2) + \lambda^{(m-k)/m}E(|S_{n,k,m}(\boldsymbol{\varepsilon})|_m^2)).$$

Once we use (5.4) and (5.5) to estimate each summand we get (5.6). ∎

Finally we combine expression (5.1) for the expected error in the face of noisy data with the estimates from the previous theorem to get the major theorem on convergence rates for spline smoothing of inexact data.

THEOREM 5.8. *Suppose samples $y_i = f(x_i) + \varepsilon_i$ are taken from a function f in $W_2^k[a,b]$, $k>0$, on uniform n-point partitions Δ_n of $[a,b]$ subject to uncorrelated, mean zero errors ε_i with common variance σ^2. If these samples are used to construct the smoothing spline approximation $S_{n,\lambda,m}(\mathbf{y})$ of order 2m, then for any c and for all n and λ satisfying $1/n\lambda_{1/2m} \leqslant c$ the expected value of the integrated mean square error in the jth derivative, $j<k$, satisfies*

$$E(|f - S_{n,\lambda,m}(\mathbf{y})|_j^2)$$
$$\leqslant P(\lambda + ((b-a)/n)^{2m})^{(k-j)/m}\,|f|_k^2 + Q\sigma^2(b-a)^2/n\lambda^{(2j+1)/2m} \qquad (5.9)$$

for constants P, Q depending only on m, k, and c.

In particular if $n \to \infty$ and the λ_n are chosen to satisfy $\lambda_n \sim n^{-2m/(2k+1)}$ (so $\bar{\Delta}_n^{2m}/\lambda_n$ is bounded), then

$$E(|f - S_{n,\lambda_n,m}(\mathbf{y})|_j^2) = O(n^{-2(k-j)/(2k+1)}). \qquad (5.10)$$

Proof. Combine the previous theorem and Theorem 4.22 with the equality in (5.1) to get (5.9). For (5.10) just note that $\lambda_n \sim n^{-2m/(2k+1)}$ implies the dominant term in each summand of (5.9) is $O(n^{-2(k-j)/(2k+1)})$. ∎

Remark. Similar bounds for the case $j=0$, $k=m$, can be found in various papers of Wahba, e.g., [5, 23]. Also, it seems most likely, in light of [22], that the estimates in Proposition 5.2 from [5] continue to hold for partitions with bounded mesh ratios, $\bar{\Delta}_n/\underline{\Delta}_n \leqslant r$. If this is so, then the previous theorem will carry over to this more general setting.

Acknowledgment

The Author would like to express his thanks to Walt Powers for introducing him to the problems which lead to this work and for providing a source of technical and intellectual support.

References

1. R. A. Adams, "Sobolev Spaces," Academic Press, New York, 1975.
2. S. Agmon, "Lectures on Elliptic Boundary Value Problems," Van Nostrand, Princeton, N.J., 1965.
3. J. Bergh and J. Löfström, "Interpolation Spaces," *in* Grundlehren der mathematischen Wissenschaften, Springer-Verlag, Berlin, 1976.
4. C. de Boor, How small can one make the derivatives of an interpolating function? *J. Approx. Theory* **13** (1975), 105–116.
5. P. Craven and G. Wahba, Smoothing noisy data with spline functions, *Numer. Math.* **31** (1979), 377–403.
6. J. Cullum, Numerical differentiation and regularization, *SIAM J. Numer. Anal.* **8** (1971), 254–265.
7. J. Fleisher, "Spline Smoothing Routines," Reference manual for the 1110, Academic Computing Center, University of Wisconsin, Madison, 1979.
8. T. Lyche and L. Schumaker, Procedures for computing smoothing and interpolating natural splines, *Comm. ACM* **17** (1974), 453–467.
9. T. Lyche, L. Schumaker, and K. Sepehrnoori, FORTRAN Subroutines for Computing Smoothing and Interpolating Natural Splines, May 1980, preprint.
10. E. Magenes, Spazi di interpolazione ed equazioni a derivate parziali, *in* "Atti del VII Congresso dell'Unione Matematica Italiana, Genova, 1963," pp. 134–197, Edizione Cremonese, Roma, 1964.
11. A. A. Melkman and C. A. Micchelli, Optimal estimation of linear operators in Hilbert spaces from inaccurate data, *SIAM J. Numer. Anal.* **16** (1979), 87–105.
12. C. A. Micchelli and T. J. Rivlin, "A Survey of Optimal Recovery," Plenum, New York, 1977.
13. D. L. Ragozin, W. E. Powers, and T. L. Tong, "Numerical Differentiation Using Regularization and Generalized Crossvalidation," 1981, preprint.
14. C. H. Reinsch, Smoothing by spline functions, *Numer. Math.* **10** (1967), 177–183.
15. I. J. Schoenberg, Spline interpolation and the higher derivatives, *Proc. Nat. Acad. Sci. U.S.A.* **51** (1964), 24–28.
16. I. J. Schoenberg, Spline functions and the problem of graduation, *Proc. Nat. Acad. Sci. U.S.A.* **52** (1964), 947–950.
17. M. H. Schultz, Error bounds for polynomial spline interpolation, *Math. Comp.* **24** (111) (1970), 507–515.
18. P. Speckman, "Minimax Estimates of Linear Functionals in a Hilbert Space," 1981, preprint.
19. A. N. Tikhonov and V. Y. Arsenin, "Solutions of Ill-Posed Problems," Winston, Washington, D.C., 1977. From Russian (F. John, Ed.).
20. F. Utreras D., "A Package for Smoothing Noisy Data with Splines," Tech. Rept. MA-80-B-209, Department of Mathematics, University of Chile, Santiago, 1980.
21. F. Utreras D., Sur le choix du paramètre d'adjustment dans le lissage par fonctions spline, *Numer. Math.* **34** (1980), 15–28.

22. F. UTRERAS D., Optimal smoothing of noisy data using spline functions, *SIAM J. Sci. Stastist. Comput.* **2** (1981), 349–362.
23. G. WAHBA, Smoothing noisy data with spline functions, *Numer. Math.* **24** (1975), 383–393.