JOURNAL OF APPROXIMATION THEORY **86**, 293–309 (1996)
ARTICLE NO. 0071

Almost Optimal Differentiation Using Noisy Data

Klaus Ritter

Mathematisches Institut, Universität Erlangen-Nürnberg, Bismarckstraße 1 1/2, 91054 Erlangen, Germany
E-mail: ritter@mi.uni-erlangen.de

Communicated by Charles A. Micchelli

Received January 1, 1995; accepted in revised form October 18, 1995

We study differentiation of functions f based on noisy data $f(t_i)+\varepsilon_i$. We recover $f^{(k)}$ either at a single point or on the interval $[0, 1]$ in L_2-norm. Under stochastic assumptions on f and ε_i, we determine the order of the errors of the best linear methods which use n noisy function values. Polynomial interpolation for the pointwise problem and smoothing splines for the problem in L_2-norm are shown to be almost optimal. The analysis involves worst case estimates in reproducing kernel Hilbert spaces and a Landau inequality.

1. INTRODUCTION

Suppose we wish to compute a derivative of a function f, if only inaccurate data $f(t_i)+\varepsilon_i$ are available. How do we select the sampling points t_i, and how do we approximate the derivative? Which error do we expect?

We study two variants of the differentiation problem for real-valued functions f on $[0, 1]$. We wish either to recover the derivative $f^{(k)}(t)$ at a single point $t\in[0, 1]$, or to recover $f^{(k)}$ on the whole interval $[0, 1]$. In the latter case we consider the distance from $f^{(k)}$ in L_2-norm. In order to treat also the recovery of the function itself we assume $k\geqslant 0$.

The function f may be observed at a finite number of points $t_i\in[0, 1]$; however, the values $f(t_i)$ are corrupted by some noise ε_i. The data $f(t_i)+\varepsilon_i$ are used to recover the kth derivative of f approximately. We consider linear methods (estimators)

$$S_n(f, \varepsilon)=\sum_{i=1}^{n} (f(t_i)+\varepsilon_i)\cdot g_i, \tag{1}$$

where $g_i\in\mathbb{R}$ for recovering $f^{(k)}(t)$, and $g_i\in L_2([0, 1])$ for recovering $f^{(k)}$.

Assumptions on the function f and on the noise ε are needed to derive error bounds for methods S_n. In this paper $\varepsilon_1, ..., \varepsilon_n$ are uncorrelated random variables with zero mean and common variance $\sigma^2>0$, i.e., $E(\varepsilon_i)=0$

0021-9045/96 $18.00

and $E(\varepsilon_i \varepsilon_j) = \delta_{ij}\sigma^2$. The variance σ^2 is assumed to be known. As with the noise, we also consider the average case with respect to the functions. More precisely, f is a measurable random function (stochastic process) with zero mean, $E(f(t)) = 0$, and covariance kernel $R(s, t) = E(f(s) f(t))$ where $s, t \in [0, 1]$, see [7, 16]. The regularity of f in mean square sense is specified by the regularity of the covariance kernel R. In order to let $f^{(k)}$ be well defined and continuous in mean square sense for $k \leqslant r$, we assume that the partial derivative $R^{(r, r)}$ is at least continuous on $[0, 1]^2$. Finally, we assume that function values and noise are uncorrelated, i.e., $E(f(t)\varepsilon_i) = 0$ for all $t \in [0, 1]$ and $1 \leqslant i \leqslant n$. Observe that we use E to denote the expectation with respect to the random function and the noise (if applicable). We allow repeated observations at the same point with uncorrelated noise by taking $t_i = \cdots = t_{i+l}$ for some i and l. The stochastic approach to numerical differentiation follows Anderssen and Bloomfield [1].

The error of a linear method S_n is defined in mean square sense by

$$e(S_n, k, t) = (E(f^{(k)}(t) - S_n(f, \varepsilon))^2)^{1/2}$$

for recovering $f^{(k)}(t)$ or by

$$e(S_n, k) = (E \, \|f^{(k)} - S_n(f, \varepsilon)\|_2^2)^{1/2} = \left(E \int_0^1 (f^{(k)}(t) - S_n(f, \varepsilon)(t))^2 \, dt \right)^{1/2}$$

for recovering $f^{(k)}$ on $[0, 1]$.

The following questions arise. What are the minimal errors in the class of all methods which use n function values? Which selection of sampling points t_i and elements g_i leads to these minimal errors? The answers depend on the random function f through the covariance kernel R. In this paper we study kernels R such that the corresponding reproducing Hilbert spaces coincide or differ only slightly from Sobolev spaces $W_2^{r+1}([0, 1])$. This holds true, for instance, if R satisfies the Sacks–Ylvisaker regularity conditions, see [21–23]. In particular, f may be the r-fold integrated Brownian motion.

It turns out that the minimal errors are of order

$$\min(\sigma/\sqrt{n}, 1)^{1 - k/(r+1/2)} \qquad \text{for recovering } f^{(k)}(t) \tag{2}$$

if $n \geqslant r + 1$ and

$$\min(\sigma/\sqrt{n}, 1)^{1 - (k+1/2)/(r+1)} \qquad \text{for recovering } f^{(k)} \tag{3}$$

if $\sigma \geqslant n^{-(r+1/2)}$. In the latter problem, a noise with $\sigma \leqslant n^{-(r+1/2)}$ does not effect the order; we get $n^{-(r-k+1/2)}$ as in the case of exact data. For recovering $f^{(k)}(t)$, the optimal order is obtained by polynomial interpolation of sample means at points which are concentrated around t; without repetitions we get the same order. For recovering $f^{(k)}$ on $[0, 1]$, equidistant points and natural smoothing splines yield the optimal order.

Plaskota [17] has already obtained (3) in the case $r=k=0$, which corresponds to recovering a Brownian motion f from noisy data $f(t_i)+\varepsilon_i$. While Plaskota uses conditional probability measures, our analysis involves worst case estimates in reproducing kernel Hilbert spaces. The proof of the upper bound in (3) relies on a Landau inequality for the norms $\|h^{(k)}\|_\infty$, $\|h\|_2$, and $\|h^{(r+1)}\|_2$ on the interval $[0, 1]$, which is due to Gabushin [6] and Kwong and Zettl [9]. Relations between Landau problems and numerical differentiation are used by several authors, see [2, 4, 9, 12] for results and further references.

Recovery from noisy data by means of smoothing splines is studied in the monographs of Eubank [5], Wahba [29], and Plaskota [19]. In particular for differentiation, upper bounds for the resulting errors are obtained under different assumptions concerning the functions f and the noise ε. A partial list of references includes Ragozin [20], Cox [3], Utreras [26, 27], and Vershinin and Pavlov [28], who study the worst case with respect to f. The noise is assumed to be deterministic in [28] and stochastic in the other papers. Univariate functions f are considered in [20, 28], as in the present paper, while [3, 26, 27] study the multivariate case.

2. OPTIMAL METHODS FOR GIVEN SAMPLING POINTS

This section contains known results concerning the optimal choice of the scalars or functions g_i, given fixed points t_i, see (1). The optimal g_i yield minimal errors in the class of all linear estimators which use the sampling points t_i. We discuss the role of reproducing kernel Hilbert spaces and the optimality of smoothing spline methods.

For recovering $f^{(k)}$, Fubini's Theorem gives

$$e(S_n, k)^2 = \int_0^1 E\left(f^{(k)}(t) - \sum_{i=1}^{n} (f(t_i)+\varepsilon_i)\cdot g_i(t)\right)^2 dt. \tag{4}$$

Therefore minimizing this error with respect to the functions g_i is equivalent to minimizing the integrand with respect to the scalars $g_i(t)$ for almost every t. Hence an optimal method to recover $f^{(k)}$ on $[0, 1]$ coincides almost everywhere with optimal methods to recover $f^{(k)}(t)$.

Let $H(R)$ denote the Hilbert space with reproducing kernel R, see [16, 29]. The scalar product and the norm on $H(R)$ are denoted by

$\langle \cdot, \cdot \rangle_R$ and $\|\cdot\|_R$, respectively. By h we denote an arbitrary function $h \in H(R)$. Since $R^{(r,r)}$ is continuous on $[0, 1]^2$ by assumption, we have $H(R) \subset C^r([0, 1])$. Observe that $h^{(k)}(t) = \langle h, \eta \rangle_R$ with $\eta = R^{(0,k)}(\cdot, t)$. Furthermore, $E(f^{(k)}(s)\, f^{(l)}(t)) = R^{(k,l)}(s, t)$. Here $s, t \in [0, 1]$ and $k, l \leqslant r$.

Consider the Hilbert space $H(R) \times \mathbb{R}^n$, equipped with the scalar product

$$\langle (h_1, p_1), (h_2, p_2) \rangle = \langle h_1, h_2 \rangle_R + \sigma^2 \sum_{i=1}^{n} p_{1,i} \cdot p_{2,i}.$$

Let e_i denote the ith unit vector in $\mathbb{R}^n$, and let BX denote the unit ball in a normed space X.

For recovering $f^{(k)}(t)$, the error of S_n satisfies

$$\begin{aligned} e(S_n, k, t)^2 &= E\left(f^{(k)}(t) - \sum_{i=1}^{n} f(t_i) \cdot g_i\right)^2 + \sigma^2 \sum_{i=1}^{n} g_i^2 \\ &= \left\| \eta - \sum_{i=1}^{n} g_i \cdot R(\cdot, t_i) \right\|_R^2 + \sigma^2 \sum_{i=1}^{n} g_i^2 \\ &= \left\| (\eta, 0) - \sum_{i=1}^{n} g_i \cdot (R(\cdot, t_i), e_i) \right\|^2. \end{aligned}$$

We conclude that

$$e(S_n, k, t) = \sup_{(h, p) \in B(H(R) \times \mathbb{R}^n)} \left| h^{(k)}(t) - \sum_{i=1}^{n} (h(t_i) + \sigma^2 p_i) \cdot g_i \right|. \tag{5}$$

Hence we are dealing with a linear problem, the recovery of $h^{(k)}(t)$, on the unit ball in a Hilbert space. Formally, exact data are available. In this situation, the optimality of an abstract spline algorithm is known, see [25, Section 4.5.7]. For fixed $t_i \in [0, 1]$ and corrupted function values $y_i = f(t_i) + \varepsilon_i$ let (h^*, p^*) denote the unique solution of the minimization problem

$$\|(h, p)\| \to \min \qquad \text{on} \quad \{(h, p) \in H(R) \times \mathbb{R}^n : h(t_i) + \sigma^2 p_i = y_i\}.$$

Equivalently, h^* is the unique solution of the minimization problem

$$\|h\|_R^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} (h(t_i) - y_i)^2 \to \min. \tag{6}$$

The solutions of problems of this type are called smoothing splines. The function h^* depends linearly on $y_1, ..., y_n$ and $S_n^*(f, \varepsilon) = (h^*)^{(k)}(t)$ has minimal error among all linear methods which use sampling points t_i. Summarizing we obtain the following.

PROPOSITION 1 (see [8, 18]). *The smoothing spline algorithms*

$$S_n^*(f, \varepsilon) = (h^*)^{(k)} \qquad \text{and} \qquad S_n^*(f, \varepsilon) = (h^*)^{(k)}(t)$$

are optimal for recovering $f^{(k)}$ *and* $f^{(k)}(t)$, *respectively.*

Optimality properties of smoothing spline algorithms for linear problems with noisy data are known in several settings. See [5, 19, 25, 29] for results and further references.

EXAMPLE 1. Let $K_1(s, t) = \min(s, t)$ be the Brownian motion kernel, let $K_2(s, t) = \exp(-|s-t|)$ be the Ornstein–Uhlenbeck kernel, and let $K_3(s, t) = 1 - |s-t|$. We have $H(K_1) = \{h \in W_2^1([0, 1]) : h(0) = 0\}$ and $H(K_2) = H(K_3) = W_2^1([0, 1])$. The norms in these reproducing kernel Hilbert spaces are given by

$$\|h\|_{K_1} = \|h'\|_2,$$

$$\|h\|_{K_2}^2 = (\|h'\|_2^2 + \|h\|_2^2 + h(0)^2 + h(1)^2)/2,$$

$$\|h\|_{K_3}^2 = (\|h'\|_2^2 + (h(0) + h(1))^2)/2,$$

see [14, 29].

Let f_i be a random function with covariance kernel K_i. The partial derivative $K_i^{(1,1)}(s, t)$ does not exist for $s = t$, and therefore f_i is not differentiable in mean square sense. To obtain a random function f with regularity $r > 0$ we integrate f_i r-fold with deterministic or stochastic boundary conditions, i.e.,

$$f(s) = \sum_{k=0}^{r} \theta_k s^k/k! + (r-1)!^{-1} \int_0^1 f_i(u) \cdot (s-u)_+^{r-1}\, du,$$

with suitable constants or random variables θ_k. We denote the covariance kernel of f by R.

Consider the Brownian motion f_1. If $\theta_0 = \cdots = \theta_r = 0$ then

$$R(s, t) = r!^{-2} \int_0^1 (s-u)_+^r \cdot (t-u)_+^r\, du,$$

and f is called r-fold integrated Brownian motion. Furthermore,

$$H(R) = \{h \in W_2^{r+1}([0, 1]) : h(0) = \cdots = h^{(r)}(0) = 0\}$$

and $\|h\|_R = \|h^{(r+1)}\|_2$. In this case recovery and integration of functions with noisy data are studied in [17]. If $(\theta_0, ..., \theta_r)$ is independent of f_1 and normally distributed with zero mean and covariance matrix $\sigma_0^2\, Id$ then $H(R) = W_2^{r+1}([0, 1])$ and

$$\|h\|_R^2 = \frac{1}{\sigma_0^2} \sum_{k=0}^{r} h^{(k)}(0)^2 + \|h^{(r+1)}\|_2^2. \tag{7}$$

See [29, Chapter 1].

Observe that the kernels K_2 and K_3 correspond to (wide sense) stationary random functions; i.e., K_i is a function of $s-t$ only, see [7]. Stationarity is an interesting assumption as there are no distinguished points in the interval [0, 1]. In [10, 14] it is shown how to preserve stationarity by r-fold integration using random boundary conditions.

Remark 1. The classical smoothing splines are defined by the minimization problem

$$\lambda \cdot \|h^{(r+1)}\|_2^2 + \frac{1}{n} \sum_{i=1}^{n} (h(t_i) - y_i)^2 \to \min \tag{8}$$

on the space $W_2^{r+1}([0, 1])$. This corresponds to (6) with $\lambda = \sigma^2/n$ and with $\|h\|_R$ replaced by the seminorm $\|h^{(r+1)}\|_2$. If $n \geqslant r+1$ then (8) has a natural polynomial spline of degree $2r+1$ as its unique solution. The solution of (6) tends to the solution of (8), if $\|\cdot\|_R$ is given by (7) and $\sigma_0 \to \infty$.

In our approach the smoothing parameters σ^{-2} in (6) and λ in (8) are determined by the variance of the noise. Other choices of λ, in particular in the case of an unknown variance σ^2, are studied in the literature. See [5, 29].

3. DESIGN PROBLEM AND ERROR BOUNDS

We turn to the question of optimal selection of the sampling points t_i and elements g_i simultaneously, see (1). Only the number n of evaluations is fixed. Due to Proposition 1 we basically have to vary over the points t_i; the corresponding minimization problem is often called the design problem. Moreover, we are interested in the respective minimal errors

$$\inf_{S_n} e(S_n, k, t) \qquad \text{and} \qquad \inf_{S_n} e(S_n, k)$$

in the class of all estimators which use n sampling points.

Let $t, t_1, ..., t_n \in [0, 1]$ be arbitrary, and consider the corresponding optimal method S_n^* for recovering $f^{(k)}(t)$. Then

$$e(S_n^*, k, t) = \sup\{|h^{(k)}(t)| : (h, p) \in B(H(R) \times \mathbb{R}^n), h(t_i) = -\sigma^2 p_i\}$$

$$= \sup\left\{|h^{(k)}(t)| : \|h\|_R^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} h(t_i)^2 \leqslant 1\right\}, \tag{9}$$

which follows from (5) and [13, Theorem 5]. Using (4) and Proposition 1 we obtain

$$e(S_n^*, k)^2 = \int_0^1 \sup\left\{|h^{(k)}(t)|^2 : \|h\|_R^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} h(t_i)^2 \leqslant 1\right\} dt \tag{10}$$

for the optimal method to recover $f^{(k)}$.

Clearly the minimal errors depend on the regularity of the functions f, and hence on the regularity of the covariance kernel R. The latter determines the regularity of the functions in $H(R)$. In the following we assume that $H(R) = W_2^{r+1}([0, 1])$ as sets, related assumptions are discussed in Remark 4. Due to the closed graph theorem, the norm $\|\cdot\|_R$ is equivalent to any norm on $W_2^{r+1}([0, 1])$ which turns this space into a reproducing kernel Hilbert space.

Furthermore, the minimal errors depend on k (and t), as well as on the number n of sampling points and on the variance σ^2 of the noise. If $\sigma/\sqrt{n}$ is large then the following turns out. For both variants of the differentiation problem the error of any method is bounded from below by a positive constant. Conversely, the error of the zero algorithm $S_n = 0$ is finite and does not depend on n or σ. We put

$$q = \min(\sigma/\sqrt{n}, 1).$$

Hence small values of q form the interesting case.

We can determine the minimal errors and solve the design problem only up to multiplicative constants. For the corresponding almost optimal designs we analyze not only the optimal spline algorithms, see Proposition 1, but also methods which are easier to implement and which depend on R only through the condition $H(R) = W_2^{r+1}([0, 1])$, see Remarks 2 and 3. In particular, these methods do not use the norm $\|\cdot\|_R$, which may be hard to determine. Henceforth, positive constants γ and γ_i with possibly different values may only depend on R.

The following theorem gives the order of the minimal errors for recovering $f^{(k)}(t)$ at the point t.

THEOREM 1. *Suppose that* $H(R) = W_2^{r+1}([0,1])$ *where* $r \in \mathbb{N}_0$. *Then there are constants* $\gamma_1, \gamma_2 > 0$ *such that*

$$\gamma_1 \cdot q^{1-k/(r+1/2)} \leqslant \inf_{S_n} e(S_n, k, t) \leqslant \gamma_2 \cdot q^{1-k/(r+1/2)}$$

for all $\sigma > 0$ *and* $n \geqslant r+1$, *and all* $0 \leqslant k \leqslant r$ *and* $t \in [0,1]$.

Proof. First we prove the lower bound. Using $|h(t_i)| \leqslant \|h\|_\infty$ and (9) we obtain

$$\inf_{S_n} e(S_n, k, t) \geqslant \sup\{|h^{(k)}(t)| : \|h\|_R^2 + q^{-2}\|h\|_\infty^2 \leqslant 1\}.$$

Note that the norms $\|h\|_R$ and $\max(\|h\|_\infty, \|h^{(r+1)}\|_2)$ are equivalent on $H(R)$. Hence $q \leqslant 1$ implies

$$\inf_{S_n} e(S_n, k, t) \geqslant \gamma \cdot \sup\{|h^{(k)}(t)| : \|h^{(r+1)}\|_2 \leqslant 1, \|h\|_\infty \leqslant q\}.$$

Take $\psi \in C^{r+1}(\mathbb{R})$ with

$$\int_{\mathbb{R}} \psi^{(r+1)}(s)^2 \, ds \leqslant 1, \qquad \sup_{s \in \mathbb{R}} |\psi(s)| \leqslant 1, \qquad \psi^{(k)}(0) > 0,$$

and put

$$h(s) = q \cdot \psi((s-t)/q^{1/(r+1/2)}).$$

Then

$$\|h^{(r+1)}\|_2 \leqslant 1, \qquad \|h\|_\infty \leqslant q, \qquad h^{(k)}(t) = q^{(2(r-k)+1)/(2r+1)} \cdot \psi^{(k)}(0),$$

and the lower bound follows.

We obtain the upper bound by polynomial interpolation if $\sigma^2 \leqslant n$. Assume that $n = l(r+1)$ with $l \in \mathbb{N}$. Take $x_0, ..., x_r \in [0,1]$ such that $x_i - x_{i-1} = v$ and $t = \lambda x_0 + (1-\lambda) x_r$ for some $\lambda \in [0,1]$ and some positive v. Consider sample means

$$y_i = \frac{1}{l} \sum_{j=1}^{l} (f(x_i) + \varepsilon_{il+j}) = f(x_i) + \frac{1}{l} \sum_{j=1}^{l} \varepsilon_{il+j} \tag{11}$$

of l-fold noisy evaluation of f at the points x_i. Let p denote the polynomial of degree at most r with $p(x_i) = y_i$ for $i = 0, ..., r$ and define $S_n(f, \varepsilon) = p^{(k)}(t)$.

We have

$$p(s) = \sum_{i=0}^{r} y_i \cdot p_i((s - x_0)/v),$$

where p_i are the polynomials of degree r with $p_i(j)=\delta_{ij}$ for $i, j=0, ..., r$. Hence

$$p^{(k)}(t)=\frac{1}{v^k}\sum_{i=0}^{r} y_i\cdot p_i^{(k)}((1-\lambda)r)$$

and

$$e(S_n, k, t)^2=E(f^{(k)}(t)-S_n(f, 0))^2+E(S_n(0, \varepsilon))^2$$

$$=\sup_{h\in BH(R)} |h^{(k)}(t)-S_n(h, 0)|^2+\frac{\sigma^2}{lv^{2k}}\cdot\gamma$$

with

$$\gamma=\sum_{i=0}^{r}(p_i^{(k)}((1-\lambda)r))^2.$$

Clearly $h^{(k)}(t)=S_n(h, 0)$ for all polynomials h of degree at most r. Therefore

$$\sup_{h\in BH(R)} |h^{(k)}(t)-S_n(h, 0)|\leqslant\gamma\cdot v^{r-k+1/2}$$

with a suitable constant γ. We conclude that

$$e(S_n, k, t)^2\leqslant\gamma\cdot\left(v^{2(r-k)+1}+\frac{\sigma^2}{lv^{2k}}\right).$$

Take

$$v=\min((\sigma/\sqrt{l})^{1/(r+1/2)}, 1/r) \tag{12}$$

to obtain

$$e(S_n, k, t)^2\leqslant\gamma\cdot(\sigma/\sqrt{l})^{(2(r-k)+1)/(r+1/2)}.$$

Furthermore, the error of the method $S_n=0$ is given by $\sup\{|h^{(k)}(t)|: \|h\|_R\leqslant 1\}=R^{(k,k)}(t, t)^{1/2}$, see (5). ∎

Remark 2. According to the proof of Theorem 1, the following method is almost optimal to recover $f^{(k)}(t)$ if $n\geqslant\max(r+1, \sigma^2)$. Put $l=\lfloor n/(r+1)\rfloor$ and define v by (12). Approximate $f^{(k)}(t)$ by $S_n(f, \varepsilon)=p^{(k)}(t)$, where p is the polynomial of degree at most r which interpolates the sample means (11) at equidistant points $x_0, ..., x_r$ with $x_i-x_{i-1}=v$ and $x_0\leqslant t\leqslant x_r$.

The following theorem gives the order of the minimal errors for recovering $f^{(k)}$ on the interval $[0, 1]$.

THEOREM 2. *Suppose that* $H(R) = W_2^{r+1}([0,1])$ *where* $r \in \mathbb{N}_0$. *Then there are constants* $\gamma_1, \gamma_2 > 0$ *with the following properties for all* $\sigma > 0$ *and* $n \in \mathbb{N}$, *and all* $0 \leqslant k \leqslant r$.

If $\sigma \leqslant n^{-(r+1/2)}$ *then*

$$\gamma_1 \cdot n^{-(r-k+1/2)} \leqslant \inf_{S_n} e(S_n, k) \leqslant \gamma_2 \cdot n^{-(r-k+1/2)}. \tag{13}$$

If $\sigma \geqslant n^{-(r+1/2)}$ *then*

$$\gamma_1 \cdot q^{1-(k+1/2)/(r+1)} \leqslant \inf_{S_n} e(S_n, k) \leqslant \gamma_2 \cdot q^{1-(k+1/2)/(r+1)}. \tag{14}$$

Proof. First we establish the lower bounds. Let A denote the operator of k-fold integration, i.e.,

$$Ah(s) = (k-1)!^{-1} \int_0^1 h(u) \cdot (s-u)_+^{k-1}\, du.$$

Due to (10) we have

$$e(S_n^*, k)^2 \geqslant \int_0^1 \sup\{|h^{(k)}(t)|^2 : \|h\|_R \leqslant 1,\, h(t_i) = 0\}\, dt$$

$$\geqslant \gamma \int_0^1 \sup\{|h(t)|^2 : h \in X\}\, dt,$$

where

$$X = \{h \in W_2^{r-k+1}([0,1]) : h(0) = \cdots = h^{(r-k)}(0) = 0,$$
$$\|h^{(r-k+1)}\|_2 \leqslant 1,\, Ah(t_i) = 0\}.$$

By this estimate, $e(S_n^*, k)$ is, modulo a constant, bounded from below by an average error for recovering an $(r-k)$-fold integrated Brownian motion, see Example 1. The data are the exact values of the k-fold integrals, evaluated at t_i. The error of the latter problem is bounded from below by $\gamma \cdot n^{-(r-k+1/2)}$, see [15], and hence we get the lower bound in (13).

To obtain the lower bound in (14) we proceed similar to the proof of the lower bound in Theorem 1. Let $\psi \in C^{r+1}(\mathbb{R})$ with

$$\int_{\mathbb{R}} \psi^{(r+1)}(s)^2\, ds \leqslant 1, \qquad \sup_{s \in \mathbb{R}} |\psi(s)| \leqslant 1, \qquad \psi^{(k)}(0) > 0,$$

and

$$\psi(s)=0 \qquad \text{if} \quad |s| \geqslant 1/2. \tag{15}$$

For any $t \in [0, 1]$ we define

$$h_t(s)=q^{(2r+1)/(2r+2)} \cdot \psi((s-t)/q^{1/(r+1)}).$$

Clearly

$$\|h_t\|_\infty \leqslant q^{(2r+1)/(2r+2)}, \qquad h_t^{(k)}(t)=q^{(2(r-k)+1)/(2r+2)} \cdot \psi^{(k)}(0),$$

and

$$h_t(s)=0 \qquad \text{if} \quad |s-t| \geqslant 1/2 \cdot q^{1/(r+1)}.$$

In particular, h_t vanishes with all derivatives at $s=0$ or $s=1$, and therefore

$$\|h_t\|_R \leqslant \gamma \cdot \|h_t^{(r+1)}\|_2 \leqslant \gamma.$$

Consider an arbitrary set $T=\{t_1, ..., t_n\} \subset [0, 1]$. Without loss of generality we may assume that the points t_i are pairwise different. Put $v = \min(q^{1/(r+1)}, 1/4)$ and define $I_j =](j-1)\,2v, j2v[$ for $j=1, ..., \lfloor 1/(2v) \rfloor$. Observe that

$$\#\{j\colon \#(I_j \cap T) \leqslant 4nv\} \geqslant \lfloor 1/(4v) \rfloor,$$

where $\#$ is used to denote the number of elements in a set. Taking into account only points $t \in](j-1)\,2v+v/2, j2v-v/2[$, we conclude that the Lebesgue measure of the set

$$U=\{t \in [0, 1]\colon \#(]t-v/2, t+v/2[\cap T) \leqslant 4nv\}$$

is at least $v \cdot \lfloor 1/(4v) \rfloor \geqslant 1/8$.

For any $t \in U$ the respective function h_t satisfies

$$\frac{1}{\sigma^2} \sum_{i=1}^{n} h_t(t_i)^2 \leqslant \frac{4nv \cdot \|h_t\|_\infty^2}{\sigma^2} \leqslant 4.$$

Let S_n^* denote the optimal estimator which uses the points from T. From (10) we get the lower bound in (14) by

$$\begin{aligned} e(S_n^*, k)^2 &\geqslant \int_U \sup \left\{ |h^{(k)}(t)|^2\colon \|h\|_R^2 + \frac{1}{\sigma^2} \sum_{i=1}^{n} h(t_i)^2 \leqslant 1 \right\} dt \\ &\geqslant \gamma \cdot q^{(2(r-k)+1)/(r+1)}. \end{aligned}$$

If $\sigma > \sqrt{n}$ then the zero algorithm $S_n = 0$ yields the upper bound in (14). We show that otherwise equidistant sampling points $t_i = (i-1)/(n-1)$ and natural smoothing splines yield the matching upper bounds.

From (4) and (5) we obtain

$$e(S_n, k)^2 = \int_0^1 \sup_{(h,p) \in B(H(R) \times \mathbb{R}^n)} |h^{(k)}(t) - S_n(h, \sigma^2 p)(t)|^2 \, dt$$

$$\leqslant \gamma \int_0^1 \sup\{|h^{(k)}(t) - S_n(h, \sigma^2 p)(t)|^2 : (h, p) \in Y\} \, dt$$

for any linear method S_n, where

$$Y = \left\{ (h, p) \in W_2^{r+1}([0,1]) \times \mathbb{R}^n : \|h^{(r+1)}\|_2^2 + \sigma^2 \sum_{i=1}^n p_i^2 \leqslant 1 \right\}.$$

Assume that $n \geqslant \max(r+1, \sigma^2)$ and let $h^\dagger$ be the solution of (8) with $y_i = f(t_i) + \varepsilon_i$ and $\lambda = \sigma^2/n$. Note that $h^\dagger$ also solves the minimization problem

$$\|h^{(r+1)}\|_2^2 + \sigma^2 \sum_{i=1}^n p_i^2 \to \min \qquad \text{on}$$

$$\{(h, p) \in W_2^{r+1}([0,1]) \times \mathbb{R}^n : h(t_i) + \sigma^2 p_i = y_i\}.$$

Consider the method $S_n(f, \varepsilon) = (h^\dagger)^{(k)}$. A general theorem on optimality of abstract spline algorithms, see [25, Section 4.5.7], implies the following. For fixed $t \in [0,1]$, the method $(h, p) \mapsto (S_n(h, \sigma^2 p))^{(k)}(t)$ is worst case optimal on Y to recover $h^{(k)}(t)$ from the data $h(t_i) + \sigma^2 p_i$. Furthermore,

$$\sup\{|h^{(k)}(t) - S_n(h, \sigma^2 p)(t)| : (h, p) \in Y\}$$
$$= \sup\{|h^{(k)}(t)| : (h, p) \in Y, h(t_i) = -\sigma^2 p_i\}$$
$$\leqslant \sup\{\|h^{(k)}\|_\infty : (h, p) \in Y, h(t_i) = -\sigma^2 p_i\}.$$

We conclude that

$$e(S_n, k) \leqslant \gamma \cdot \sup \left\{ \|h^{(k)}\|_\infty : \|h^{(r+1)}\|_2 \leqslant 1, \sum_{i=1}^n h(t_i)^2 \leqslant \sigma^2 \right\}.$$

According to [20, Theorem 3.2],

$$\|h\|_2^2 \leqslant \gamma \cdot \left(n^{-1} \sum_{i=1}^n h(t_i)^2 + n^{-(2r+2)} \cdot \|h^{(r+1)}\|_2^2 \right)$$

for any $h \in W_2^{r+1}([0, 1])$. Therefore

$$e(S_n, k) \leqslant \gamma \cdot \sup\{\|h^{(k)}\|_\infty : \|h^{(r+1)}\|_2 \leqslant 1, \|h\|_2 \leqslant q + n^{-(r+1)}\}.$$

Take $\psi \in C^{r+1}(\mathbb{R})$ with

$$\sup_{s \in \mathbb{R}} |\psi(s)| \leqslant 1, \qquad \psi(0) = 1, \qquad \psi^{(j)}(0) = 0 \qquad \text{if} \quad 1 \leqslant j \leqslant r,$$

and (15). For $h \in H(R)$ with $|h^{(k)}(t^*)| = \|h^{(k)}\|_\infty$ we define

$$h_0(s) = \psi(s - t^*) \cdot h(s).$$

If $\|h\|_R \leqslant 1$ then $\|h_0^{(r+1)}\|_2 \leqslant \gamma$, $\|h_0\|_2 \leqslant \|h\|_2$, and $\|h_0^{(k)}\|_\infty \geqslant |h_0^{(k)}(t^*)| = \|h^{(k)}\|_\infty$. Moreover, h_0 vanishes with all derivatives at $s = 0$ or 1. Hence

$$e(S_n, k) \leqslant \gamma \cdot \sup\{\|h^{(k)}\|_\infty : h(s) = \cdots = h^{(r)}(s) = 0 \text{ for } s = 0 \text{ or } 1,$$
$$\|h^{(r+1)}\|_2 \leqslant 1, \|h\|_2 \leqslant q + n^{-(r+1)}\}.$$

Due to [6] and [9, p. 21], a Landau inequality

$$\|h^{(k)}\|_\infty \leqslant \gamma \cdot \|h\|_2^{(2(r-k)+1)/(2r+2)} \cdot \|h^{(r+1)}\|_2^{(2k+1)/(2r+2)}$$

holds for all $h \in W_2^{r+1}([0, 1])$ such that $h, ..., h^{(r)}$ have zeros in $[0, 1]$. This implies

$$e(S_n, k) \leqslant \gamma \cdot (q + n^{-(r+1)})^{(2(r-k)+1)/(2r+2)},$$

and the upper bounds in (13) and (14) follow. ∎

Remark 3. In the previous proof we have shown in particular that the following method is almost optimal to recover $f^{(k)}$ if $n \geqslant \max(r+1, \sigma^2)$. Define $S_n(f, \varepsilon) = (h^\dagger)^{(k)}$ where $h^\dagger$ is the natural smoothing spline of degree $2r+1$ which solves (8) with sampling points $t_i = (i-1)/(n-1)$, data $y_i = f(t_i) + \varepsilon_i$, and smoothing parameter $\lambda = \sigma^2/n$.

Remark 4. The condition $H(R) = W_2^{r+1}([0, 1])$ in Theorems 1 and 2 is not easy to check, in general. However, the upper bounds from these theorems hold for the methods described in Remarks 2 and 3 whenever

$$H(R) \subset W_2^{r+1}([0, 1]). \tag{16}$$

Sacks and Ylvisaker [22, 23] have introduced regularity conditions for covariance kernels R to study the design problem for weighted integration, based on exact data. The conditions say that

(a) $L = R^{(r,r)}$ is continuous on $[0,1]^2$ with partial derivatives of order two off the diagonal in $[0,1]^2$,

(b) along the diagonal $L^{(1,0)}$ has a discontinuity of constant height,

(c) $L^{(2,0)}(s,\cdot) \in H(L)$ with uniformly bounded norms $\|L^{(2,0)}(s,\cdot)\|_L$.

Several authors have studied weighted integration or recovery of functions in the exact data case under Sacks–Ylvisaker conditions. See [21] for a list of references. It is easy to check that the kernels $L = K_i$ in Example 1 satisfy the conditions.

In [21] it is shown that the Sacks–Ylvisaker conditions imply the inclusion (16), and hence the upper bounds from Theorems 1 and 2 hold. Moreover, if

$$R^{(r,k)}(\cdot, 0) = 0 \tag{17}$$

for $k = 0, ..., r-1$, additionally, then

$$\{h \in W_2^{r+1}([0,1]) : h^{(k)}(0) = h^{(k)}(1) = 0 \text{ for } k = 0, ..., r\}$$
$$\subset H(R) \subset W_2^{r+1}([0,1]). \tag{18}$$

The proofs of Theorem 1 and 2 are easily modified to see that the lower bounds also hold if $H(R)$ satisfies (18). However, in Theorem 1 the constants γ_1 and γ_2 may depend on t and may vanish for $t = 0$ and 1. We add that the boundary condition (17) states that $f^{(k)}(0)$ and $f^{(r)}(t)$ are uncorrelated for any $0 \leqslant t \leqslant 1$.

Remark 5. In particular for $k = 0$, i.e., for recovering the function f, and for fixed $\sigma > 0$ we have

$$\gamma_1 \cdot n^{-1/2 + 1/4(r+1)} \leqslant \inf_{S_n} e(S_n, 0) \leqslant \gamma_2 \cdot n^{-1/2 + 1/4(r+1)}.$$

For $r = 0$ this result is due to Plaskota [17]. Furthermore, Plaskota shows that noisy observation of rth derivatives yields minimal errors of order $n^{-1/2}$ for the r-fold integrated Brownian motion if $r > 0$. Hence Theorem 2 answers an open problem from [17]: Observation of noisy rth derivatives is more powerful to recover f than observation of noisy function values.

Remark 6. We briefly discuss the case when exact data $f(t_i)$ are available. Suppose that $H(R)$ satisfies (18). For recovering $f^{(k)}(t)$, polynomial interpolation at $n = k+1$ points yields arbitrarily small errors. For recovering $f^{(k)}$, the minimal errors satisfy

$$\gamma_1 \cdot n^{-(r-k+1/2)} \leqslant \inf_{S_n} e(S_n, k) \leqslant \gamma_2 \cdot n^{-(r-k+1/2)}. \tag{19}$$

In the case $k=0$ this result is due to Speckman [24]. Due to results of Papageorgiou and Wasilkowski [15] the minimal errors for recovering an $(r-k)$-fold integrated Brownian motion from exact values of linear functionals are of order $n^{-(r-k+1/2)}$. Hereby the lower bound bound in (19) follows. The upper bound in (19) holds for the error of methods which are based on natural splines which interpolate at equidistant points, cf. Remark 3.

Even in the case of exact data, asymptotic constants are only known for $r=k=0$. Lee [11] has shown that

$$\lim_{n\to\infty} \inf_{S_n} e(S_n, 0)\cdot n^{1/2} = 6^{-1/2}$$

for recovering a Brownian motion from data $f(t_i)$. Plaskota [17] has obtained explicit estimates for $\inf_{S_n} e(S_n, 0)\cdot n^{1/4}$ in the latter problem in the presence of noise.

Remark 7. Stronger conclusions hold if the noise ε is normally distributed and the random function f is Gaussian.

The order of minimal errors remains unchanged if we consider a much broader class of methods S_n, see [19] for general results. These methods may choose the sampling points adaptively (sequentially) and determine the total number of observations by means of an adaptive stopping rule. Any approximation to a derivative is permitted which depends (measurably) on the data $f(t_i)+\varepsilon_i$ and uses n observations on the average.

Furthermore, Theorem 2 extends to the case of recovering $f^{(k)}$ in L_p-norm with $1\leqslant p<\infty$; we omit the details. It would be interesting to know the order of minimal errors in the case $p=\infty$.

ACKNOWLEDGMENTS

I thank T. Müller-Gronbach, E. Novak, M. Nussbaum, and L. Plaskota for valuable remarks and comments.

REFERENCES

1. R. S. Anderssen and P. Bloomfield, Numerical differentiation procedures for non-exact data, *Numer. Math.* **22** (1974), 157–182.
2. V. V. Arestov, On the best approximation of the operators of differentiation and related questions, *in* "Approximation Theory, Proc. Conf. Poznań 1972" (Z. Ciesielski and J. Musielak, Eds.), pp. 1–9, Reidel, Dordrecht, 1975.
3. D. D. Cox, Multivariate smoothing spline functions, *SIAM J. Numer. Anal.* **21** (1984), 789–813.

4. D. L. Donoho, Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery, *Probab. Theory Relat. Fields* **99** (1994), 145–170.
5. R. L. Eubank, “Spline Smoothing and Nonparametric Regression,” Dekker, New York, 1988.
6. V. N. Gabushin, Inequalities for the norms of a function and its derivatives in metric L_p, *Math. Notes* **1** (1967), 194–198.
7. I. I. Gihman and A. V. Skorohod, “The Theory of Stochastic Processes I,” Springer-Verlag, Berlin, 1974.
8. G. S. Kimeldorf and G. Wahba, Spline functions and stochastic processes, *Sankhya Ser. A* **32**, Part 2 (1970), 173–180.
9. Man Kam Kwong and A. Zettl, “Norm Inequalities for Derivatives and Differences,” Lecture Notes in Mathematics, Vol. 1536, Springer-Verlag, Berlin, 1992.
10. R. Lasinger, Integration of covariance kernels and stationarity, *Stochastic Process. Appl.* **45** (1993), 309–318.
11. D. Lee, Approximation of linear operators on a Wiener space, *Rocky Mountain J. Math.* **16** (1986), 641–659.
12. C. A. Micchelli, On an optimal method for the numerical differentiation of smooth functions, *J. Approx. Theory* **18** (1976), 189–204.
13. C. A. Micchelli and T. J. Rivlin, A survey of optimal recovery, *in* “Optimal Estimation in Approximation Theory” (C. A. Micchelli and T. J. Rivlin, Eds.), pp. 1–54, Plenum, New York, 1977.
14. T. Mitchell, M. Morris, and D. Ylvisaker, Existence of smoothed stationary processes on an interval, *Stochastic Process. Appl.* **35** (1990), 109–119.
15. A. Papageorgiou and G. W. Wasilkowski, On the average complexity of multivariate problems, *J. Complexity* **6** (1990), 1–23.
16. E. Parzen, Statistical inference on time series by Hilbert space methods, I, (1959), *in* “Time Series Analysis Papers” (E. Parzen, Ed.), pp. 251–382, Holden–Day, San Francisco, 1967.
17. L. Plaskota, Function approximation and integration on the Wiener space with noisy data, *J. Complexity* **8** (1992), 301–323.
18. L. Plaskota, Optimal approximation of linear operators based on noisy data on functionals, *J. Approx. Theory* **73** (1993), 93–105.
19. L. Plaskota, “Noisy Information and Computational Complexity,” Cambridge Univ. Press, Cambridge, 1996.
20. D. L. Ragozin, Error bounds for derivative estimates based on spline smoothing of exact or noisy data, *J. Approx. Theory* **37** (1983), 335–355.
21. K. Ritter, G. W. Wasilkowski, and H. Woźniakowski, Multivariate integration and approximation for random fields satisfying Sacks–Ylvisaker conditions, *Ann. Appl. Probab.* **5** (1995), 518–540.
22. J. Sacks and D. Ylvisaker, Designs for regression with correlated errors, *Ann. Math. Statist.* **37** (1966), 68–89.
23. J. Sacks and D. Ylvisaker, Statistical designs and integral approximation, *in* “Proc. 12th Bienn. Semin. Can. Math. Congr.” (R. Pyke, Ed.), pp. 115–136, Can. Math. Soc., Montreal.
24. P. Speckman, “L_p Approximation of Autoregressive Gaussian Processes,” Report, University of Oregon, Department of Statistics, Eugene, 1979.
25. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, “Information-Based Complexity,” Academic Press, New York, 1988.
26. F. I. Utreras, Convergence rates for multivariate smoothing spline functions, *J. Approx. Theory* **52** (1988), 1–27.

27. F. I. Utreras, Recent results on multivariate smoothing splines, *in* "Multivariate Approximation and Interpolation" (W. Haussmann and K. Jetter, Eds.), pp. 299–312, Internationale Schriftenreihe zur Numerischen Mathematik, Vol. 94, Birkhäuser Verlag, Basel, 1990.
28. V. V. Vershinin and N. N. Pavlov, Splines in a convex set, and the problem of numerical differentiation, *USSR Comput. Math. Math. Phys.* **27** (1987), 199–202.
29. G. Wahba, "Spline Models for Observational Data," SIAM, Philadelphia, 1990.