Journal of Approximation Theory **93**, 501–515 (1998)
Article No. AT983175

Average Case L_∞-Approximation in the Presence of Gaussian Noise

Leszek Plaskota*

Institute of Applied Mathematics and Mechanics, University of Warsaw, ul. Banacha 2, 02-097, Warsaw, Poland
E-mail: leszekp@mimuw.edu.pl

Communicated by Allan Pinkus

Received January 6, 1997; accepted in revised form May 1, 1997

We consider the average case L_∞-approximation of functions from $C^r([0, 1])$ with respect to the r-fold Wiener measure. An approximation is based on n function evaluations in the presence of Gaussian noise with variance $\sigma^2 > 0$. We show that the nth minimal average error is of order $n^{-(2r+1)/(4r+4)} \ln^{1/2} n$, and that it can be attained either by the piecewise polynomial approximation using repetitive observations, or by the smoothing spline approximation using non-repetitive observations. This completes the already known results for L_q-approximation with $q < \infty$ and $\sigma \geqslant 0$, and for L_∞-approximation with $\sigma = 0$.

1. INTRODUCTION AND MAIN RESULT

Optimal approximation of a function based on exact or noisy observations is one of the leading problems in approximation theory, statistics, and information-based complexity. Many different settings have been proposed to solve this problem. The settings are basically determined by whether we have deterministic or stochastic assumptions on the function and noise.

In this paper, we consider the average case setting, where both the function and the noise are assumed to be random. The problem is to approximate a real function f which belongs to the space

$$F^r = \{ f \in C^r([0, 1]) \colon f^{(j)}(0) = 0, 0 \leqslant j \leqslant r \}$$

and is a realization of the stochastic process corresponding to the r-fold Wiener measure w^r on F^r, $r \geqslant 0$. That is, the mean element of w^r is zero and the covariance kernel

* This research was partially supported by the State Committee for Scientific Research of Poland.

0021-9045/98 $25.00

$$R_r(s,t)=\int_{F^r} f(s)\,f(t)\,w^r(df)$$
$$=\int_0^1 \frac{(s-u)_+^r}{r!}\,\frac{(t-u)_+^r}{r!}\,du, \qquad 0\leqslant s,\ t\leqslant 1.$$

An approximation to f is based only on some *noisy* information y about f. More specifically, the approximation is given as a function $\varphi(y)\colon [0,1]\to\mathbb{R}$, where y is of the form

$$y=N(f)+x,$$

with

$$N(f)=(f(t_1), f(t_2), ..., f(t_n))$$

and x having the zero mean n dimensional normal distribution with covariance matrix $\sigma^2 I_n$,

$$x\sim\pi_\sigma^n=\mathcal{N}(0,\sigma^2 I_n).$$

That is, the noise $x_i=y_i-f(t_i)$ coming from different observations is independent and identically distributed, $x_i\sim_{iid}\mathcal{N}(0,\sigma^2)$. We assume that $\sigma\geqslant 0$ allowing $\sigma=0$ to cover also the exact information case. The quality of approximation is given by its average L_p-error in the L_q-norm with respect to w^r, i.e.,

$$e_q^p(r,\sigma;\varphi,N)=\left(\int_{F^r}\int_{\mathbb{R}^n}(\|f-\varphi(N(f)+x)\|_q)^p\,\pi_\sigma^n(dx)\,w^r(df)\right)^{1/p}.$$

Here $1\leqslant p<\infty$ and $1\leqslant q\leqslant\infty$. Letting $\mathcal{N}_n$ be the set of all N consisting of at most n function evaluations, we are interested in the minimal error that can be attained using $N\in\mathcal{N}_n$,

$$e_q^p(r,\sigma;n)=\inf_{N\in\mathcal{N}_n}\inf_{\varphi} e_q^p(r,\sigma;\varphi,N).$$

The minimal errors $e_q^p(r,\sigma;n)$ have been obtained in the case of exact information where they are roughly equal to the average Kolmogorov n-widths. See, e.g., Speckman [13] and Ritter [11] for $q<\infty$, and Müller–Gronbach and Ritter [5], Ritter [10], and Wasilkowski [15] for $q=\infty$. Results on average n-widths can be found in Maiorov and Wasilkowski [4].

In the case of noisy information with $p=q=2$ the errors $e_2^2(r,\sigma;n)$ have been obtained by Plaskota [6] for $r=0$ and Ritter [12] for $r\geqslant 1$. As noticed in Ritter [12], similar formulas can be derived for all p, q except for $q=\infty$.

We now state those results precisely. We use notation $a_n \asymp b_n$ to denote the same asymptotic behavior of sequences, i.e., this means that there are $0 < m \leqslant M < \infty$ such that $m \leqslant a_n/b_n \leqslant M$ holds for sufficiently large n, $n \geqslant n_0 = n_0(\sigma)$. We stress that in all the asymptotic formulas for $e_q^p(r, \sigma; n)$ with $\sigma > 0$, the constants m and M are independent of σ; however, $n_0(\sigma) \to \infty$ if $\sigma \to 0$ or if $\sigma \to \infty$.

With this notation, for $1 \leqslant q < \infty$ and $\sigma \geqslant 0$ we have

$$e_q^p(r, \sigma; n) \asymp n^{-(r+1/2)} + \left(\frac{\sigma}{\sqrt{n}}\right)^{(2r+1)/(2r+2)}, \tag{1}$$

and for $q = \infty$ and $\sigma = 0$ we have

$$e_\infty^p(r, 0; n) \asymp n^{-(r+1/2)} \ln^{1/2} n.$$

The main purpose of this paper is to solve the missing case of the average case L_∞-approximation with noise. Note that a similar problem, but for the worst case error over f, has been studied by Donoho [2] and Korostelev [3], who obtained the asymptotic constants for equidistant sampling. Results for the multivariate L_∞-approximation in the same setting are given in Plaskota [9]. In the average case setting, we have the following theorem which is the main result of this paper.

Main Theorem. *For $\sigma > 0$ we have*

$$e_\infty^p(r, \sigma, n) \asymp \left(\frac{\sigma}{\sqrt{n}}\right)^{(2r+1)/(2r+2)} \ln^{1/2} n.$$

Thus, taking together the exact and noisy information cases, for $\sigma \geqslant 0$ we obtain

$$e_\infty^p(r, \sigma; n) \asymp n^{-(r+1/2)} \ln^{1/2} n + \left(\frac{\sigma}{\sqrt{n}}\right)^{(2r+1)/(2r+2)} \ln^{1/2} n.$$

In particular, comparing this with (1), we have $e_\infty^p(r, \sigma; n) \asymp e_q^p(r, \sigma; n) \ln^{1/2} n$, i.e., the minimal error of L_∞-approximation is larger by a factor of $\ln^{1/2} n$ than the minimal error of L_q-approximation with $q < \infty$, for both the exact and noisy information.

A proof of the Main Theorem is given in Section 2 and Section 3. In particular, we show that an order optimal L_∞-approximation can be obtained by piecewise polynomial interpolation of degree r based on repetitive observations at equidistant points. In Section 4 we additionally note that the well-known smoothing spline approximation based on non-repetitive and equidistant observations also leads to an order optimal approximation.

The techniques used for proving the upper and lower bounds can be applied to obtain the already known formulas for L_q-approximation with $1 \leqslant q < \infty$, as indicated in a series of remarks.

2. THE UPPER BOUND

We first need some facts from the case of exact information. Let N_n be information about function values at equidistant points,

$$N_n(f) = (f(1/n), f(2/n), ..., f(1)).$$

For exact data ($\sigma = 0$) information N_n is order optimal. This is well known for L_q-approximation with $q < \infty$. A proof for $q = \infty$ and $r = 0$ can be found in Ritter [10]. In what follows, we show optimality of N_n for $q = \infty$ and $r \geqslant 1$. In this case, Maiorov and Wasilkowski [4] give us a lower bound $n^{-(r+1/2)} \ln^{1/2} n$ for the minimal error; hence we only need to construct an approximation using N_n with error of that order.

Assume without loss of generality that n is a multiple of r, i.e., $s = n/r$ is an integer. (Otherwise we take $s = \lfloor n/r \rfloor$.) For $1 \leqslant i \leqslant s$ and $0 \leqslant j \leqslant r$, let

$$t_{i,j} = (i-1)/s + j/n.$$

Consider the approximation φ_n which uses information N_n and relies on piecewise polynomial interpolation of degree r. That is, for t in an interval $[t_{i,0}, t_{i,r}]$, $1 \leqslant i \leqslant s$, we have

$$(\varphi_n(y_1, ..., y_n))(t) = (w_{i,r}(y_{(i-1)r}, y_{(i-1)r+1}, ..., y_{ir}))(t)$$

(with $y_0 = 0$), where $w_{i,r}(z_0, z_1, ..., z_r)$ is the polynomial of degree r taking z_j at the points $t_{i,j}$, $0 \leqslant j \leqslant r$.

LEMMA 1. *In the case of exact information and $r \geqslant 1$ we have*

$$e^p_\infty(r, 0; \varphi_n, N_n) = O(n^{-(r+1/2)} \ln^{1/2} n),$$

i.e., the piecewise polynomial interpolation of degree r based on equidistant points is order optimal.

Proof. Let $f \in F^r$. Let $f(z_1, z_2, ..., z_k)$ denote the divided difference of f of order k, i.e., $f(z_1)$ is the usual function value at z_1, and (for $z_1 \neq z_k$)

$$f(z_1, z_2, ..., z_k) = \frac{f(z_2, ..., z_k) - f(z_1, ..., z_{k-1})}{z_k - z_1}.$$

Using the well-known formula for the error of Lagrange interpolation we obtain that for any $t \in [t_{i,0}, t_{i,r}]$

$$\begin{aligned} f(t) - (\varphi_n(N_n(f)))(t) &= f(t) - (w_{i,r}(f(t_{i,0}), ..., f(t_{i,r})))(t) \\ &= (t - t_{i,0}) \cdots (t - t_{i,r})\, f(t_{i,0}, ..., t_{i,r}, t) \\ &= (t - t_{i,0}) \cdots (t - t_{i,r}) \frac{f(t_{i,1}, ..., t_{i,r}, t) - f(t_{i,0}, ..., t_{i,r-1}, t)}{t_{i,r} - t_{i,0}} \\ &= (t - t_{i,0}) \cdots (t - t_{i,r}) \frac{f^{(r)}(\xi_1(t))/r! - f^{(r)}(\xi_2(t))/r!}{1/s}, \end{aligned}$$

where $\xi_k(t) \in [t_{i,0}, t_{i,r}]$, $k = 1, 2$. Hence

$$|f(t) - (\varphi_n(N_n(f)))(t)| \leqslant \frac{1}{s^r r!} \psi_i(f)$$

with

$$\psi_i(f) = \max\{ |f^{(r)}(\xi_1) - f^{(r)}(\xi_2)| : t_{i,0} \leqslant \xi_1, \xi_2 \leqslant t_{i,r} \}.$$

For $1 \leqslant i \leqslant s$, define the functions

$$g_i(t) = \sqrt{s} \left(f^{(r)} \left(\frac{\lfloor ts \rfloor + t}{s} \right) - f^{(r)} \left(\frac{\lfloor ts \rfloor}{s} \right) \right), \qquad 0 \leqslant t \leqslant 1.$$

Then the g_i's are independent and identically distributed according to the classical Wiener measure w^0 on $C([0, 1])$. Moreover,

$$\psi_i(f) = s^{-1/2} \tilde{\psi}(g_i)$$

with

$$\tilde{\psi}(g) = \sup\{ |g(\xi_1) - g(\xi_2)| : 0 \leqslant \xi_1, \xi_2 \leqslant 1 \}.$$

We obtain that the error of φ_n using N_n can be bounded from above by

$$\begin{aligned} e^p_\infty(r, 0; \varphi_n, N_n) &= \left(\int_{F^r} \| f - \varphi_n(N_n(f)) \|^p_\infty \, w^r(df) \right)^{1/p} \\ &\leqslant (r!)^{-1} s^{-(r+1/2)} \left(\int_{F^0} \cdots \int_{F^0} (\max_{1 \leqslant i \leqslant s} \tilde{\psi}(g_i))^p \, w^0(dg_1) \cdots w^0(dg_s) \right)^{1/p}. \end{aligned}$$

Since $\mathrm{Prob}\{\tilde{\psi}(g_i) > a\} \approx (2\pi)^{-1/2} \int_{|z|>a} e^{-z^2/2}\, dz$ as $a \to \infty$, see Billingsley [1], we have that the multiple integral above is asymptotically equal to the expected value of $\max_{1 \leqslant i \leqslant s} |z_i|^p$ with $z_i \sim_{iid} \mathcal{N}(0, 1)$, which in turn behaves as $\ln^{p/2} s$. Hence

$$e^p_\infty(r, 0; \varphi_n, N_n) \leqslant (r!)^{-1}\, s^{-(r+1/2)} \ln^{1/2} s(1 + o(1))$$
$$= \frac{r^{r+1/2}}{r!}\, n^{-(r+1/2)} \ln^{1/2} n(1 + o(1)) \asymp n^{-(r+1/2)} \ln^{1/2} n,$$

as claimed. ∎

Note that the approximation based on the piecewise polynomial interpolation is defined for $r \geqslant 1$. In the case of $r = 0$ the corresponding approximation would be

$$(\varphi_n(y))(t) = y_{\lfloor tn \rfloor/n}, \qquad 0 \leqslant t \leqslant 1. \tag{2}$$

Although $\varphi_n(y)$ has discontinuities at j/n, $\forall j$, its error is well defined and

$$e^p_\infty(0, 0; \varphi_n, N_n) = \left(\int_{F^0} \|f - \varphi_n(N_n(f))\|^p_\infty\, w^0(df) \right)^{1/p}$$
$$= n^{-1/2} \left(\int_{F^0} \cdots \int_{F^0} (\max_{1 \leqslant i \leqslant n} \tilde{\psi}(g_i))^p\, w^0(dg_1) \cdots w^0(dg_n) \right)^{1/p}$$
$$\approx \sqrt{\frac{2 \ln n}{n}}.$$

From Ritter [10] we know that the minimal error is in this case asymptotically equal to $\sqrt{\ln n/(2n)}$. (Actually the proof was given for $p = 1$ but the same holds for all $p \geqslant 1$.) Hence φ_n is only twice worse than the optimal approximation.

Remark 1. The same piecewise polynomial interpolation is order optimal also for all $q < \infty$. Indeed, in view of Hölder's inequality it suffices to consider $p = q < \infty$. Using the notation from the proof of Lemma 1 we obtain

$$e^q_q(r, 0; \varphi_n, N_n) = \left(\int_{F^r} \int_0^1 |f(t) - (\varphi_n(N_n(f)))(t)|^q\, dt\, w^r(df) \right)^{1/q}$$
$$\leqslant \left(\int_0^1 \int_{F^r} |f(t) - (\varphi_n(N_n(f)))(t)|^q\, w^r(df)\, dt \right)^{1/q}$$
$$\leqslant \frac{1}{r!\, s^r} \left(\sum_{i=1}^{s} \int_{t_{i,0}}^{t_{i,r}} \int_{F^r} (\psi_i(f))^q\, q^r(df)\, dt \right)^{1/q}$$
$$= s^{-(r+1/2)} (r!)^{-1} \left(\int_{F^0} (\tilde{\psi}_i(g))^q\, w^0(dg) \right)^{1/q} \asymp n^{-(r+1/2)}.$$

Obviously, the same holds for $r=0$ and the piecewise constant approximation defined in (2).

We now use the approximation procedure φ_n to prove the upper bound of the Main Theorem. We assume that $\sigma>0$ and $r\geqslant 0$. Let N_m^k be information where observations at equidistant points are repeated k times,

$$N_m^k(f)=(\underbrace{f(1/n), ..., f(1/n)}_{k}, ..., \underbrace{f(1), ..., f(1)}_{k}).$$

For information

$$y=(y_{1,1}, ..., y_{1,k}, ..., y_{m,1}, ..., y_{m,k})\in\mathbb{R}^{n_1},$$

where $n_1=mk$, let $\tilde{y}=(\tilde{y}_1, ..., \tilde{y}_m)\in\mathbb{R}^m$ be defined as

$$\tilde{y}_i=\frac{1}{k}\sum_{j=1}^{k} y_{i,j}, \qquad 1\leqslant i\leqslant m.$$

Define the approximation φ_m^k as

$$\varphi_m^k(y)=\varphi_m(\tilde{y}),$$

where φ_m is as in Lemma 1 if $r\geqslant 1$, and φ_m is the linear spline interpolating data or the piecewise constant approximation defined in (2) if $r=0$. Note that, due to linearity of φ_m, we equivalently have $\varphi_m^k(y)=(1/k)\sum_{j=1}^k \varphi_m(y^j)$ with $y^j=(y_{1,j}, ..., y_{m,j})$.

In what follows, we write $a_n(\sigma)=O(b_n(\sigma))$ iff there is a constant $0<M<\infty$ such that $a_n(\sigma)\leqslant Mb_n(\sigma)$ holds for sufficiently large n, $n\geqslant n_0$. Here M is independent of σ, but n_0 may depend on σ.

THEOREM 1. *Let $\sigma>0$. Let the number m of points and the number k of repetitions be chosen such that $mk=n_1\leqslant n$,*

$$m\asymp\left(\frac{\sigma}{\sqrt{n}}\right)^{-1/(r+1)} \qquad \textit{and} \qquad k\asymp n\left(\frac{\sigma}{\sqrt{n}}\right)^{1/(r+1)}.$$

Then for $1\leqslant p<\infty$ we have

$$e_\infty^p(r,\sigma;\varphi_m^k, N_m^k)=O\left(\left(\frac{\sigma}{\sqrt{n}}\right)^{(2r+1)/(2r+2)}\ln^{1/2} n\right).$$

Proof. Letting $\tilde{x}=\tilde{y}-N_m(f)\sim\mathcal{N}(0,\sigma^2/kI_m)$ we have

$$\begin{aligned}
e^p_\infty(r,\sigma;\varphi^k_m,N^k_m)
&=\left(\int_{F^r}\int_{\mathbb{R}^n}\|f-\varphi^k_m(N^k_m(f)+x)\|^p_\infty\,\pi^{n_1}_\sigma(dx)\,w^r(df)\right)^{1/p}\\
&=\left(\int_{F^r}\int_{\mathbb{R}^m}\|f-\varphi_m(N_m(f)+\tilde{x})\|^p_\infty\,\pi^m_{\sigma/\sqrt{k}}(d\tilde{x})\,w^r(df)\right)^{1/p}\\
&\leqslant\left(\int_{F^r}\int_{\mathbb{R}^m}(\|f-\varphi_m(N_m(f))\|_\infty+\|\varphi_m(\tilde{x})\|_\infty)^p\,\pi^m_{\sigma/\sqrt{k}}(d\tilde{x})\,w^r(df)\right)^{1/p}.
\end{aligned}$$

Observe that for all m and $\tilde{x}\in\mathbb{R}^m$ we have $\|\varphi_m(\tilde{x})\|_\infty\leqslant K\|\tilde{x}\|_\infty$, where

$$K=\sup_{\|z\|_\infty\leqslant 1}\|\varphi_r(z)\|_\infty<\infty.$$

This and the inequality $(|a|+|b|)^p\leqslant 2^{p-1}(|a|^p+|b|^p)$ yield

$$\begin{aligned}
e^p_\infty(r,\sigma;\varphi^k_m,N^k_m)
&\leqslant 2^{1-1/p}\left(\int_{F^r}\int_{\mathbb{R}^m}\|f-\varphi_m(N_m(f))\|^p_\infty+\|\varphi_m(\tilde{x})\|^p_\infty\,\pi^m_{\sigma/\sqrt{k}}(d\tilde{x})\,w^r(df)\right)^{1/p}\\
&\leqslant 2^{1-1/p}\left(\int_{F^r}\|f-\varphi_m(N_m(f))\|^p_\infty\,w^r(df)+\left(\frac{K\sigma}{\sqrt{k}}\right)^p\int_{\mathbb{R}^m}\|z\|^p_\infty\,\pi^m_1(dz)\right)^{1/p}.
\end{aligned}$$

Finally, we use Lemma 1 and the formulas for m and k to obtain

$$\begin{aligned}
e^p_\infty(r,\sigma;\varphi^k_m,N^k_m)&=O((m^{-p(r+1/2)}\ln^{p/2}m+(\sigma/\sqrt{k})^p\ln^{p/2}m)^{1/p})\\
&\asymp\left(\frac{\sigma}{\sqrt{n}}\right)^{(2r+1)/(2r+2)}\ln^{1/2}n,
\end{aligned}$$

as claimed. ∎

In the special case of $r=0$, $p=1$, and φ_m being the linear spline, the constant $K=1$ and the error

$$\begin{aligned}
e^1_\infty(O,\sigma;\varphi^k_m,N^k_m)&\leqslant\int_{F^0}\|f-\varphi_m(N_m(f))\|_\infty\,w^0(df)+\frac{\sigma}{\sqrt{k}}\int_{\mathbb{R}^m}\|z\|_\infty\,\pi^m_1(dz)\\
&\asymp\sqrt{\frac{\ln m}{2m}}+\frac{\sigma}{\sqrt{k}}\sqrt{2\ln m}
\end{aligned}$$

(as $m \to \infty$). The optimal m and k are $m \approx \sqrt{n}/(2\sigma)$ and $k \approx 2\sigma\sqrt{n}$, and then

$$e^1_\infty(0, \sigma; \varphi^k_m, N^k_m) \leqslant \left(\frac{2\sigma \ln n}{\sqrt{n}}\right)^{1/2} (1 + o(1)). \tag{3}$$

Remark 2. We saw in the previous remark that for exact information the approximation φ_n is order optimal not only for $q = \infty$ but also for $1 \leqslant q < \infty$. Similarly, in the "noisy" case the approximation φ^k_m is order optimal for all finite q. Indeed, assume without loss of generality that $p = q$. Let

$$M = \left(\int_{\mathbb{R}^{r+1}} \|z\|^q_\infty \, \pi^{r+1}_1(dz)\right)^{1/q} < \infty.$$

Then

$$\begin{aligned} e^q_q(r, \sigma; \varphi^k_m, N^k_m) &\leqslant 2^{1-1/q} \left(\int_{F^r} \|f - \varphi_m(N_m(f))\|^q_q \, w^r(df) + \left(\frac{KM\sigma}{\sqrt{k}}\right)^q\right)^{1/q} \\ &\asymp \left(\frac{\sigma}{\sqrt{n}}\right)^{(2r+1)/(2r+2)}. \end{aligned}$$

In particular, for $p = q = 2$, $r = 0$, and φ_m the linear spline, we can calculate the exact error; namely,

$$\begin{aligned} &e^2_2(0, \sigma; \varphi^k_m, N^k_m) \\ &\quad = \left(\int_{F^0} \|f - \varphi_m(N_m(f))\|^2_2 \, w^0(df) + \frac{\sigma^2}{k} \int_{\mathbb{R}^m} \int_0^1 |(\varphi_m(\tilde{x}))(t)|^2 \, dt \, \pi^m_1(d\tilde{x})\right)^{1/2} \\ &\quad = \left(\frac{1}{6m} + \frac{\sigma^2}{k}\left(\int_0^{1/m} \int_{\mathbb{R}} (ztm)^2 \, \pi^1_1(dz) \, dt \right.\right. \\ &\qquad \left.\left. + (m-1) \int_0^{1/m} \int_{\mathbb{R}^2} (z_2 tm - z_1(1 - tm))^2 \, \pi^2_1(dz) \, dt\right)\right)^{1/2} \\ &\quad = \left(\frac{1}{6m} + \frac{2\sigma^2}{3k}\left(1 - \frac{1}{2m}\right)\right)^{1/2}. \end{aligned}$$

The optimal choice of m is $m \approx \sqrt{n}/(2\sigma)$. Then the number k of repetitions is roughly $2\sigma\sqrt{n}$ and the error is asymptotically equal to

$$e^2_2(0, \sigma; \varphi^k_m, N^k_m) \approx \left(\frac{2\sigma}{3\sqrt{n}}\right)^{1/2}.$$

This can be compared with the minimal error that can be obtained from non-repetitive n equidistant observations. In the latter case, this error asymptotically equals $(\sigma/(2\sqrt{n}))^{1/2}$, see Plaskota [6]; hence we lose about 15.6% using repetitions.

3. THE LOWER BOUND

We first show the following simple, but useful lemma.

LEMMA 2. *Let μ be a Gaussian measure on $C([0,1])$ with covariance kernel R. Let $\tilde{R}$ be the covariance kernel of the conditional (a posteriori) distribution of μ with respect to independent noisy observations of $f(t_1), ..., f(t_n)$ with variance σ^2. Then*

$$\tilde{R}(t,t) \geqslant \frac{\sigma^2 R(t,t)}{\sigma^2 + nR_{\max}}, \qquad 0 \leqslant t \leqslant 1,$$

where $R_{\max} = \sup_{0 \leqslant u \leqslant 1} R(u,u)$.

Proof. Suppose that a single observation at u is performed with variance

$$\sigma_u^2 = \sigma^2 R(u,u)/R_{\max}$$

instead of σ^2. (Or equivalently that only observations of the functionals $af(t)$, where $a^2 \leqslant R_{\max}/R(u,u)$, with variance σ^2 are allowed.) Let $\tilde{R}_1$ be the corresponding conditional covariance kernel. The problem of maximally reducing $\tilde{R}_1(t,t)$ is equivalent to approximating $f(t)$ in the L_2-norm. It follows from Plaskota [7] (see also Plaskota [8, Sect. 3.8.1]) that the best way of doing this is to make all the n observations at t, and then

$$\tilde{R}_1(t,t) = \frac{\sigma_t^2 R(t,t)}{\sigma_t^2 + nR(t,t)} = \frac{\sigma^2 R(t,t)}{\sigma^2 + nR_{\max}}.$$

The proof is complete with the observation that $\sigma_t^2 \leqslant \sigma^2$ implies $\tilde{R}(t,t) \geqslant \tilde{R}_1(t,t)$. ∎

Using this lemma we now prove the lower bound of Main Theorem. We write $a_n(\sigma) = \Omega(b_n(\sigma))$ iff there is $0 < m < \infty$ independent of σ such that $a_n(\sigma) \geqslant m b_n(\sigma)$ for all sufficiently large n, $n \geqslant n_0 = n_0(\sigma)$. ∎

THEOREM 2. *For $\sigma > 0$ we have*

$$e_\infty^p(r, \sigma; n) = \Omega\left(\left(\frac{\sigma}{\sqrt{n}}\right)^{(2r+1)/(2r+2)} \ln^{1/2} n\right).$$

Proof. Let $m \geqslant 1$. Let $\tilde{w}^r$ be the conditional distribution on F^r given additional exact information that $f^{(j)}(i/m) = 0$ for $1 \leqslant i \leqslant m$ and $0 \leqslant j \leqslant r$. That is, $\tilde{w}^r$ is the well-known r-fold Brownian bridge. Its mean element is zero and its covariance kernel $R(s, t)$ is given as follows. Let $\gamma_r = ((2r+1)(r!)^2)^{-1}$. Then for $(i-1)/m < s \leqslant t < i/m$ we have

$$R(s, t) = R(t, s) = \gamma_r m^{-(2r+1)} (\tilde{s}(1+\tilde{t}))^{2r+1},$$

where $\tilde{s} = sm - (i-1)$ and $\tilde{t} = tm - (i-1)$. On the other hand, if s and t are not in the same interval $((i-1)/m, i/m)$ then the variables $f(s)$ and $f(t)$ are independent and $R(s, t) = 0$.

Now, let $u_i = (2i-1)/(2i)$, $1 \leqslant i \leqslant m$, and

$$\|f\|_* = \max_{1 \leqslant i \leqslant m} |f(u_i)| \leqslant \|f\|_\infty .$$

It is clear that the minimal error for our original problem is bounded from below by the minimal error for a new problem with w^r replaced by $\tilde{w}^r$ and with error measured in the seminorm $\|\cdot\|_*$ instead of $\|\cdot\|_\infty$. That is, we now find

$$\tilde{e}^p_*(r, \sigma; n) = \inf_{\varphi, N} \left(\int_{F^r} \int_{\mathbb{R}^n} \max_{1 \leqslant i \leqslant m} |f(u_i) - \varphi_i(N(f)+x)|^p \, \pi^n_\sigma(dx) \, \tilde{w}^r(df) \right)^{1/p},$$

where $N(f) = (f(t_1), ..., f(t_n))$ and $\varphi_i(y) = (\varphi(y))(u_i)$. To that end, observe that the random variables $z_i = f(u_i)$ are independent (with respect to $\tilde{w}^r$) and $z_i \sim_{iid} \mathcal{N}(0, \lambda)$ with

$$\lambda = R(u_i, u_i) = \sup_{0 \leqslant u \leqslant 1} R(u, u) = \gamma_r (4m)^{-(2r+1)}.$$

Also, for $t \in ((i-1)/m, i/m)$ and $j \neq i$ the variables $f(t)$ and z_j are independent. Thus a single observation at such a t can only reduce the variance of z_i and, in view of Lemma 2, the best way of doing this is to take $t = u_i$. Consequently we can restrict observations to those at the u_i's and

$$\tilde{e}^p_*(r, \sigma; n) = \inf_{\varphi, N} \left(\int_{\mathbb{R}^m} \int_{\mathbb{R}^n} \max_{1 \leqslant i \leqslant m} |z_i - \varphi_i(N(z)+x)|^p \, \pi^n_\sigma(dx) \, \pi^m_\lambda(dz) \right)^{1/p},$$

where $z = (z_1, ..., z_m)$, $\varphi = (\varphi_1, ..., \varphi_m)$, and information is of the form

$$N(z) = (z_{i_1}, ..., z_{i_n}), \qquad 1 \leqslant i_j \leqslant m.$$

We obtained that the difficulty of the original problem is bounded from below by the difficulty of approximating a finite dimensional vector $z \sim \mathcal{N}(0, \lambda I_m)$ from n noisy observations of its coordinates. The solution of the latter problem for $p = 1$ is given in Plaskota [9, see Lemma A2 of Appendix];

however, applying the same technique one can solve it for arbitrary $p \geqslant 1$. We have that the minimal error for the finite dimensional problem (and for n a multiple of m) is

$$\sigma \sqrt{\frac{m\lambda}{\sigma^2 m + \lambda n}} \left(\int_{\mathbb{R}^m} \|z\|_\infty^p \, \pi_1^m(dz) \right)^{1/p}.$$

Hence, taking into account the formula for λ and letting $m \asymp (n/\sigma^2)^{1/(2r+2)}$, we finally obtain

$$e_\infty^p(r, \sigma; n) \geqslant e_*^p(r, \sigma; n) \approx \sigma \sqrt{\frac{\gamma_r m}{\sigma^2 m^{2r+2} + \gamma_r n}} \left(\int_{\mathbb{R}^m} \|m\|_\infty^p \, \pi_1^m(dz) \right)^{1/p}$$
$$\approx \sigma \sqrt{\frac{2\gamma_r m \ln m}{\sigma^2 m^{2r+2} + \gamma_r n}} \asymp \left(\frac{\sigma}{\sqrt{n}} \right)^{(2r+1)/(2r+2)} \ln^{1/2} n,$$

as claimed. ∎

Observe that in the case of $r = 0$ and $p = 1$ we have $\lambda = 1/(4m)$. Hence, taking $m \approx \sqrt{n}/(2\sigma)$, we obtain

$$e_\infty^1(0, \sigma; n) \geqslant \sigma \sqrt{\frac{2m \ln m}{4\sigma^2 m^2 + n}} (1 + o(1)) \approx \left(\frac{\sigma \ln n}{4\sqrt{n}} \right)^{1/2}.$$

Comparing this with (3) we can write

$$e_\infty^1(0, \sigma; n) \approx a_n \left(\frac{\sigma \ln n}{\sqrt{n}} \right)^{1/2}$$

where $1/2 \leqslant a_n \leqslant \sqrt{2}$.

Remark 3. The proof of Theorem 2 can also be easily adopted to obtain lower bounds on the minimal errors for $1 \leqslant q < \infty$. Indeed, let without loss of generality $p = q$. Let $\tilde{R}_r$ be the conditional covariance kernel after n noisy observations $N(f) = (f(t_1), ..., f(t_n))$, and let $\lambda_t = \tilde{R}_r(t, t)$. Let $E_\eta(\psi(x))$ denote the expected value of $\psi(x)$ with respect to $x \sim \mathcal{N}(0, \eta)$. Proceeding as in, e.g., Ritter [10], we can write that the minimal error for the information N equals

$$\inf_\varphi e_q^q(r, \sigma; \varphi, N) = \left(\int_0^1 E_{\lambda_t}(|x|^q) \, dt \right)^{1/q} = (E_1(|x|^q))^{1/q} \left(\int_0^1 \lambda_t^{q/2} \, dt \right)^{1/q}.$$

Using Lemma 2 and notation from the proof of Theorem 2 we find that

$$\lambda_t \geqslant \frac{\sigma^2 \gamma_r}{\sigma^2 m^{2r+1} + 4^{-(2r+1)} k_i \gamma_r} (\tilde{t}(1 - \tilde{t}))^{2r+1},$$

where $m \geqslant 1$ and k_i is the number of points t_i in the interval $((i-1)/m, i/m)$. Hence

$$\inf_\varphi e_q^q(r, \sigma; \varphi, N) \geqslant (E_1(|x|^q))^{1/q} \left(\sum_{i=1}^m \left(\frac{\sigma^2 \gamma_r}{\sigma^2 m^{2r+1} + 4^{(-(2r+1))} k_i \gamma_r} \right)^{q/2} \right.$$
$$\left. \times \int_0^1 (\tilde{t}(1-\tilde{t}))^{q(r+1/2)}\, dt \right)^{1/q}$$
$$= \left(E_1(|x|^q) \int_0^1 u((1-u))^{q(r+1/2)}\, du \right)^{1/q}$$
$$\times \left(\frac{1}{m} \sum_{i=1}^m \left(\frac{\sigma^2 \gamma_r}{\sigma^2 m^{2r+1} + 4^{-(2r+1)} k_i \gamma_r} \right)^{q/2} \right)^{1/q}.$$

The last expression is minimized by $k_i = n/m$, $\forall i$. Thus taking $m \asymp (n/\sigma^2)^{1/(2r+2)}$ we finally obtain

$$\inf_\varphi e_q^q(r, \sigma; \varphi, N) = \Omega \left(\left(\frac{1}{m} \sum_{i=1}^m \left(\frac{\sigma^2 \gamma_r}{\sigma^2 m^{2r+2} + 4^{-(2r+1)} (n/m)\, \gamma_r} \right)^{q/2} \right)^{1/q} \right)$$
$$\asymp \left(\frac{\sigma^2 m}{\sigma^2 m^{2r+2} + n} \right)^{1/2} \asymp \left(\frac{\sigma}{\sqrt{n}} \right)^{(2r+1)/(2r+2)},$$

as claimed. See also Plaskota [6] for a similar proof in the special case of $r = 0$ and $p = q = 2$.

4. NON-REPETITIVE OBSERVATIONS AND SMOOTHING SPLINES

In Section 2 we constructed an order optimal approximation based on repetitive observations. In practice one often uses *non*-repetitive observations at equidistant points, i.e., $N_n(f) = (f(1/n), f(2/n), ..., f(1))$. It can be easily seen that if $\sigma > 0$ then the usual piecewise polynomial approximation does not do the job in this case and we are forced to make some smoothing. That is, as an approximation to f one uses the smoothing spline $\varphi(y) = \varphi_{\text{spl}}(y)$ which is given as the minimizer of

$$\sigma^2 \cdot \int_0^1 (f^{(r+1)}(u))^2\, du + \sum_{j=1}^n (y_j - f(j/n))^2$$

over $f \in W_2^{r+1}(0, 1)$, $f^{(i)}(0) = 0$, $0 \leqslant i \leqslant r$. It is known that the smoothing spline approximation is best possible in the sense that it minimizes the error $e_q^p(r, \sigma; \varphi, N_n)$ over all φ using N_n, see, e.g., Plaskota [8, Sect. 3.6].

It is also known that information N_n together with φ_{spl} provides an order optimal approximation for $q<\infty$, see Ritter [11, 12]. A natural question is whether this is also the case for $q=\infty$.

THEOREM 3. *For $\sigma>0$ we have*

$$e^p_\infty(r, \sigma; \varphi_{\mathrm{spl}}, N_n) = O\left(\left(\frac{\sigma}{\sqrt{n}}\right)^{(2r+1)/(2r+2)} \ln^{1/2} n\right),$$

i.e., the smoothing splice approximation φ_{spl} based on the equidistant and non-repetitive observations N_n is order optimal for uniform approximation.

Proof. We only sketch the proof since it makes use of the ideas already applied for proving the upper bound of Section 2.

Since the approximation φ_{spl} is best possible when using N_n, it suffices to show the order optimality of another approximation that uses N_n. This approximation is defined as follows.

Let m and k be chosen as in Theorem 1. We can assume without loss of generality that $mk=n$ and m is a multiple of r. Then $N_n = (N_{n,0}, N_{n,1}, ..., N_{n,k-1})$ with

$$N_{n,j} = (f(1/m - j/n), f(2/m - j/n), ..., f(1-j/n)), \qquad 0 \leqslant j \leqslant k-1.$$

Let $\varphi_{n,j}$ be an approximation that uses information $N_{n,j}$ and is order optimal for exact information; e.g., one can take (with some obvious modifications) the piecewise polynomial approximation of Lemma 1. Then we let

$$\tilde{\varphi}_n(y) = \frac{1}{k} \sum_{j=0}^{k-1} \varphi_{n,j}(y^j),$$

where $y=(y^0, ..., y^{k-1}) \in \mathbb{R}^n$ and $y^j \in \mathbb{R}^m$, $\forall j$. Note that $\tilde{\varphi}_n$ can be viewed as another smoothing procedure that uses the non-repetitive and equidistant observations.

Repeating the corresponding part of the proof of Theorem 1, we find that

$$\begin{aligned} &e^p_\infty(r, \sigma; \tilde{\varphi}_n, N_n) \\ &\quad \leqslant 2^{1-1/p} \left(\int_{F^r} \left\| \frac{1}{k} \sum_{j=0}^{k-1} (f - \varphi_{n,j}(N_{n,j}(f))) \right\|_\infty^p w^r(df) \right. \\ &\qquad \left. + \sigma^p \int_{\mathbb{R}^m} \cdots \int_{\mathbb{R}^m} \left\| \frac{1}{k} \sum_{j=0}^{k-1} \varphi_{n,j}(x^j) \right\|_\infty^p \pi_1^m(dx^0) \cdots \pi_1^m(dx^{k-1}) \right)^{1/p}. \end{aligned}$$

Since all the $\varphi_{n,j}$'s are order optimal for m observations, the first integral above behaves as $m^{-p(r+1/2)} \ln^{p/2} m$. The second (multiple) integral in turn behaves as $k^{-p/2} \ln^{p/2} m$, due to independency of $\varphi_{n,j}(x^j)$ for $0 \leqslant j \leqslant k-1$. Hence, using the formulas for m and k we obtain

$$e_\infty^p(r, \sigma; \tilde{\varphi}_n, N_n) = O\left(\left(\frac{\sigma}{\sqrt{n}}\right)^{(2r+1)/(2r+2)} \ln^{1/2} n\right),$$

as claimed. ▮

Obviously, a similar idea can be used to prove the order optimality of N_n in the case of $q < \infty$.

Note. Lemma 1 is a result of e-mail conversations with Klaus Ritter.

REFERENCES

1. P. Billingsley, "Convergence of Probability Measures," Wiley, New York, 1968.
2. D. L. Donoho, Asymptotic minimax risk for sup-norm loss: Solution via optimal recovery, *Probab. Theory Related Fields* **99** (1994), 145–170.
3. A. P. Korostelev, An asymptotically minimax regression estimator in the uniform norm up to a constant, *Theory Probab. Appl.* **38** (1993), 737–743.
4. V. E. Maiorov and G. W. Wasilkowski, Probabilistic and average linear widths in L_∞-norm with respect to r-fold Wiener measure, *J. Approx. Theory* **84** (1996), 31–40.
5. T. Müller-Gronbach and K. Ritter, Uniform reconstruction of Gaussian processes, preprint, No. A/26-95 Fachbereich Mathematik, Berlin, 1995.
6. L. Plaskota, Function approximation and integration on the Wiener space with noisy data, *J. Complexity* **8** (1992), 301–323.
7. L. Plaskota, Optimal approximation of linear operators based on noisy data on functionals, *J. Approx.* **73** (1993), 93–105.
8. L. Plaskota, "Noisy Information and Computational Complexity," Cambridge Univ. Press, Cambridge, 1996.
9. L. Plaskota, Worst case complexity of problems with random information noise, *J. Complexity* **12** (1996), 416–439.
10. K. Ritter, Approximation and optimization on the Wiener space, *J. Complexity* **6** (1990), 337–364.
11. K. Ritter, "Average Case Analysis of Numerical Problems," Habilitation Thesis, Erlangen, Germany.
12. K. Ritter, Almost optimal differentiation using noisy data, *J. Approx. Theory* **86** (1996), 293–309.
13. P. Speckman, "L_p-Approximation of Autoregressive Gaussian Processes," report University of Oregon, Department of Statistics, Eugene, 1979.
14. J. F. Traub, G. W. Wasilkowski, and H. Woźniakowski, "Information-Based Complexity," Academic Press, New York, 1988.
15. G. W. Wasilkowski, On average complexity of global optimization problems, *Math. Programming* **57** (1992), 313–324.