Approximation and Optimization on the Wiener Space

KLAUS RITTER

Mathematisches Institut, Bismarckstrasse 1 f, D-8520 Erlangen, West Germany

Received February 15, 1990

We study adaptive and nonadaptive methods for $L_q$-approximation and global optimization based on $n$ function evaluations from a Wiener space sample. We derive (asymptotically) optimal methods with respect to an average error. The error of optimal methods converges to zero with the following rates: $n^{-1/2}$ for $L_q$-approximation if $1 \leq q < m$, $(\ln n/n)^{1/2}$ if $q = \infty$, and $n^{-1/2}$ for nonadaptive methods for global optimization. We show that adaption helps for global optimization.

© 1990 Academic Press, Inc.

1. INTRODUCTION

In this paper we consider the following two problems on the space

$C = \{f: [0, 1] \to \mathbb{R} \mid f \text{ continuous, } f(0) = 0\}$.

The approximation problem consists of recovering an unknown function $f \in C$ and the global optimization problem consists of finding a point $\hat{x} \in [0, 1]$ where an unknown function $f$ attains its maximum. For both problems a finite number of function evaluations is assumed to be the only information $N(f)$ about $f$. Since this information is partial, any method for solving one of the above problems causes an error. To compare different methods we use an average error with respect to the Wiener measure on the space $C$. A general treatment for solving problems when only partial information is available can be found in Traub et al. (1988). Our research is part of the average case setting studied there.

We define two classes of information operators $N: C \to \mathbb{R}^n$. In general $N(f) = (f(x_1), \ldots, f(x_n))$ may be computed sequentially; i.e., the choice of $x_k$ may depend on the previously computed values $f(x_1), \ldots,
$f(x_{k-1})$. These operators are called *adaptive* information operators and they are defined by measurable mappings

$$x_k: \mathbb{R}^{k-1} \rightarrow [0, 1]$$

in the following way. If $y_1 = f(x_1)$ and

$$y_k = f(x_k(y_1, \ldots, y_{k-1}))$$

for $k = 2, \ldots, n$, we put

$$N(f) = (y_1, \ldots, y_n).$$

Without loss of generality we can assume that

$$0 \neq x_k(y_1, \ldots, y_{k-1}) \neq x_l(y_1, \ldots, y_{l-1}) \neq 0$$

for $y \in \mathbb{R}^{a-1}$ and $1 \leq k < l \leq n$. The number $n \in \mathbb{N}$ is called the cardinality of $N$ and the class of adaptive information operators of cardinality $n$ is denoted by $\mathcal{N}_n^{ad}$.

The subclass $\mathcal{N}_n^{non}$ consists of those $N \in \mathcal{N}_n^{ad}$ with fixed nodes $x_k$; i.e., the mappings $x_k$ are constant. These operators are called *nonadaptive* and they allow parallel evaluation of $f$.

To specify the approximation problem completely we fix $1 \leq q \leq \infty$ and consider the embedding $C \rightarrow L_q = L_q([0, 1])$. This mapping is to be approximated by any composition $\phi \circ N$, where $N \in \mathcal{N}_n^{ad}$ and $\phi: \mathbb{R}^n \rightarrow L_q$ is a measurable mapping, called an algorithm. The interpretation is as follows: if we get the information $y = N(f)$ we choose $\tilde{f} = \phi(y)$ as an approximation to $f$. This causes the individual error

$$\|f - \tilde{f}\|_q = \left(\int_0^1 |f(t) - \tilde{f}(t)|^q \, dt\right)^{1/q}$$

in the case $1 \leq q < \infty$ and

$$\|f - \tilde{f}\|_\infty = \max_{0 \leq t \leq 1} |f(t) - \tilde{f}(t)|$$

if $q = \infty$.

Now let $1 \leq p < \infty$ and let $w$ be the Wiener measure on the space $C$ (see, e.g., Billingsley (1968)). Then the $p$-average error of $\phi$ and $N$ is defined by

$$e_p(\phi, N, L_q) = \left(\int_C \|f - \phi \circ N(f)\|_q^p w(df)\right)^{1/p}$$
and the bound

\[ r_p(N, L_d) = \inf_{\phi} e_p(\phi, N, L_d) \]

is called the \textit{p-average radius} of \( N \).

Any method to solve the global optimization problem is given by \( N \in N^u_n \) and a measurable algorithm \( \phi: \mathbb{R}^n \to [0, 1] \). Knowing \( y = N(f) \) we guess that \( f \) attains its maximum at the point \( \hat{x} = \phi(y) \). We study the individual error

\[ \max_{0 \leq t \leq 1} f(t) - f(\hat{x}), \]

which leads to the \textit{p-average error}

\[ e_p(\phi, N, \text{Opt}) = \left( \int_C (\max_{0 \leq t \leq 1} f(t) - f(\phi \circ N(f)))^p w(df) \right)^{1/p} \]

and the \textit{p-average radius}

\[ r_p(N, \text{Opt}) = \inf_{\phi} e_p(\phi, N, \text{Opt}). \]

In the special case \( p = q = 2 \) the approximation problem on the Wiener space was investigated by several authors. Suldin (1960) studied non-adaptive methods using linear algorithms and Lee (1986) considered adaptive information operators and arbitrary algorithms. The multivariate approximation problem for functions \( f: [0, 1]^d \to \mathbb{R} \) was analyzed by Papageorgiou and Wasilkowski (1990). They also considered functions with higher regularity by placing the Wiener measure on the partial derivatives. Speckman (1979) studied the approximation problem for the general class of autoregressive Gaussian processes. He considered the case \( 1 \leq p = q < \infty \) and he also announced results for the case \( q = \infty \).

The approximation problem is a special instance of the so-called linear problems with Gaussian measures (see Traub et al. (1988) for a survey). Some general results of this theory are cited in the next sections. The average case analysis for the nonlinear problem of global optimization seems to be new (see Traub et al. (1988, p. 296)).

Methods for the global optimization, which are based on a statistical model like the Wiener space, are discussed by Mockus (1989) and Törn and Žilinskas (1989). Among other things they compare different kinds of methods and they give applications to practical problems.
2. STATEMENT OF THE RESULTS

Let \( N \in \mathcal{N}_n^{\text{ad}} \) be given by functions \( x_k : \mathbb{R}^{k-1} \to [0, 1] \). To simplify the notation we define

\[ x_0 = t_0 = y_0 = 0, \]

and for any nonadaptive information operator we always assume

\[ 0 < x_1 < \cdots < x_n \leq 1. \]

Our first goal is to determine algorithms with minimal \( p \)-average error among all algorithms using \( N \). These algorithms are called \( p \)-optimal for the respective problem and they are defined by the condition

\[ e_p(\phi, N, \cdot) = r_p(N, \cdot). \]

We introduce linear operators \( m(\cdot; t_1, \ldots, t_n) : \mathbb{R}^n \to C \) for \( t_1, \ldots, t_n \in [0, 1] \) mutually different. If \( t_1 < \cdots < t_n \) we put

\[
m(y; t_1, \ldots, t_n)(t) = \begin{cases} 
  y_{k-1} + (t - t_{k-1}) \cdot \frac{y_k - y_{k-1}}{t_k - t_{k-1}}, & \text{if } t_{k-1} \leq t \leq t_k \text{ for } 1 \leq k \leq n, \\
  y_n, & \text{if } t \leq t_1.
\end{cases}
\]

In general we put \( m((y_1, \ldots, y_n); t_1, \ldots, t_n) = m((y_{t(1)}, \ldots, y_{t(n)}); t_{t(1)}, \ldots, t_{t(n)}) \), where \( \tau \) is the permutation of \( \{1, \ldots, n\} \) with \( t_{t(1)} < \cdots < t_{t(n)} \). Now we consider the algorithms

\[ \phi_o(y) = m(y; x_1, \ldots, x_n(y_1, \ldots, y_{n-1})) \]

and

\[ \phi_o(y) = \min\{x_k(y_1, \ldots, y_{k-1}) \mid y_k = \max_{t=0,\ldots,n} y_t\}, \]

where we take the minimum of the set of nodes with maximal function value only to achieve the uniqueness of \( \phi_o \) for all \( y \in \mathbb{R}^n \). From Lee and Wasilkowski (1986) we know that \( \phi_o \) is \( p \)-optimal for the approximation problem for all \( 1 \leq p < \infty \) and \( 1 \leq q \leq \infty \). For the global optimization problem \( \phi_o \) is \( p \)-optimal in many cases, but not in general.

Suppose that \( p = 1 \) or that \( p > 1 \) and \( N \) uses the node \( x = 1 \). Then the algorithm \( \phi_o \) is \( p \)-optimal for the global optimization problem.
The algorithms $\phi_a$ and $\phi_b$ are very simple and similar, because the respective problem is solved for the affine linear interpolation of the information. This interpolation is the conditional expectation given $N(f) = y$, as it turns out in the next section. For linear problems with Gaussian measures, solving the problem for the conditional expectation always yields $p$-optimal algorithms, and these algorithms are called \textit{spline algorithms} (see Lee and Wasilkowski (1986)).

Next we ask for nonadaptive information operators with minimal $p$-average radius in the class $\mathcal{N}^{\text{non}}_n$. These operators are called \textit{$p$-optimal in $\mathcal{N}^{\text{non}}_n$} and they are defined by the condition

$$r_p(N, \cdot) = \inf_{N \in \mathcal{N}^{\text{non}}_n} r_p(N, \cdot).$$

For the approximation problem with $q = \infty$ or $1 \leq q < \infty$ and $p \geq \max(2q/(2 + q), 1)$ we prove that $p$-optimal information operators use \textit{equidistant nodes}, i.e., these nodes satisfy

$$x_1 = x_2 - x_1 = \cdots x_n - x_{n-1}.$$

We conjecture that this statement holds for the approximation problem with arbitrary $p$ and $q$. In particular for $1 \leq p = q < \infty$ the $p$-optimal information operator is uniquely determined by the additional requirement $x_1/(1 - x_n) = a_p$, where $a_p$ is a constant depending only on $p$ and not on $n$ (see (5)). For the global optimization problem, however, $p$-optimal information operators do not use equidistant nodes in general.

If we cannot determine $p$-optimal information operators, we ask for a sequence $N_n \in \mathcal{N}^{\text{non}}_n$ of operators such that $r_p(N_n, \cdot)$ is weakly or strongly equivalent to $\inf_{N \in \mathcal{N}^{\text{non}}_n} r_p(N, \cdot)$. Recall that the weak equivalence of sequences $a_n, b_n > 0$, denoted by $a_n \approx b_n$, is defined by

$$c_1 \leq a_n/b_n \leq c_2$$

with constants $c_1, c_2 > 0$, and the strong equivalence, denoted by $a_n = b_n$, is defined by

$$\lim_{n \to \infty} a_n/b_n = 1.$$

Hence we call any such sequence of information operators \textit{weakly or strongly asymptotically $p$-optimal in $\mathcal{N}^{\text{non}}_n$} = $\bigcup_{n=1}^{\infty} \mathcal{N}^{\text{non}}_n$. In particular we consider the operators

$$N^*_n(f) = (f(1/n), \ldots, f(k/n), \ldots, f(1)).$$
The asymptotic behavior of \( \inf_{N \in \mathcal{N}^{\text{num}}} r_p(N, \cdot) \) itself is a reasonable quantity to describe the difficulty of a problem like approximation or global optimization.

Let \( 1 \leq q < \infty \), let \( p \geq \max(2q/(2 + q), 1) \), and define the constants \( c_p \) and \( d_p \) by (3) and (4). Then the sequence \( N^*_n \) is strongly asymptotically \( p \)-optimal in \( \mathcal{N}^{\text{num}} \) for the \( L_q \)-approximation and we have

\[
r_p(N^*_n, L_p) = (c_p d_p)^{1/p} \cdot n^{-1/2},
\]

if \( p = q \) (see Speckman (1979) for \( p = q \) and Lee (1986) for \( p = q = 2 \)). For arbitrary \( p \) this sequence is weakly asymptotically \( p \)-optimal in \( \mathcal{N}^{\text{num}} \) for the same problem with

\[
r_p(N^*_n, L_p) \approx n^{-1/2}.
\]

Let \( q = \infty \) and \( 1 \leq p < \infty \). Then the sequence \( N^*_n \) is strongly asymptotically \( p \)-optimal in \( \mathcal{N}^{\text{num}} \) for the \( L_\infty \)-approximation with

\[
r_p(N^*_n, L_\infty) \approx (\ln n/(2n))^{1/2}.
\]

Let \( 1 \leq p < \infty \). Then the sequence \( N^*_n \) is weakly asymptotically \( p \)-optimal in \( \mathcal{N}^{\text{num}} \) for the global optimization with

\[
r_p(N^*_n, \text{Opt}) \approx n^{-1/2}.
\]

The analysis for the \( L_q \)-approximation problem is done by different methods in the cases \( 1 \leq q < \infty \) and \( q = \infty \). In the first case it based on the covariance function of the random element \( f \rightarrow f - \phi_{\mu} \circ N(f) \), while we use the distribution of the random variables \( f \rightarrow \max_{x_{1}, \ldots, x_{k}} |f(t) - \phi_{\mu} \circ N(f)(t)| \) in the second case.

Another linear problem which has been analyzed is the integration problem, i.e., computing an approximation to \( \int_{0}^{1} f(t) \, dt \). For \( p = 2 \), Lee (1986) proved that the nodes \( x_k = 2k/(2n + 1) \) define the unique \( p \)-optimal information operator \( N_n \in \mathcal{N}^{\text{num}} \) with \( r_2(N_n, \text{Int}) = (3^{1/2}(2n + 1))^{-1} \). Since the integral is a continuous linear functional, we have \( r_p(N, \text{Int}) = c_p^{1/p} \cdot r_2(N, \text{Int}) \) (see Traub et al. (1988, p. 291)), and the same information is \( p \)-optimal for any \( p \). Moreover the asymptotic behavior of \( \inf_{N \in \mathcal{N}^{\text{num}}} r_p(N, \text{Int}) \) depends slightly on \( p \), whereas the asymptotic behavior of \( \inf_{N \in \mathcal{N}^{\text{num}}} r_p(N, L_\infty) \) is independent of \( p \).

Finally we compare nonadaptive methods and adaptive methods, which are frequently used in practice for the global optimization. Lee and Wasilkowski (1986) proved that adaptation does not help for linear problems with Gaussian measures. In particular this means that for arbitrary \( 1 \leq p \)
\(< \infty, 1 \leq q \leq \infty, \text{ and } N \in \mathcal{N}_n^{ad} \text{ there exists a nonadaptive } \tilde{N} \in \mathcal{N}_n^{non} \text{ with } r_p(\tilde{N}, L_q) \leq r_p(N, L_q). \text{ An analog for the global optimization does not hold.}

Let \( p = 1 \text{ and } n > 1. \) Then we have

\[ \inf\{r_1(N, \text{Opt}) \mid N \in \mathcal{N}_n^{ad}\} < \inf\{r_1(N, \text{Opt}) \mid N \in \mathcal{N}_n^{non}\}. \]

It would be very interesting to quantify the improvement which is due to adaption for the global optimization problem.

### 3. The Regular Conditional Probability

We provide a short discussion on the regular conditional probabilities that are fundamental for our analysis. Since \( N \in \mathcal{N}_n^{ad} \) is surjective and measurable there exists a family \((w(\cdot \mid y))_{y \in \mathbb{R}^n}\) of probability measures on \( \mathbb{C} \) such that

1. \( w(N^{-1}[y] \mid y) = 1 \) for \( Nw \) almost all \( y \in \mathbb{R}^n \),
2. \( w(B \mid \cdot) \) is measurable for any Borel set \( B \subseteq \mathbb{C} \),
3. \( w(B) = \int_{\mathbb{R}^n} w(B \mid y)Nw(dy) \) for any Borel set \( B \subseteq \mathbb{C} \).

This family is uniquely defined \( Nw \) a.e. and it is called the regular conditional probability with respect to \( N \) (see Parthasarathy (1967, p. 147)). Properties 2 and 3 imply that

\[ \int_{\mathbb{C}} H(f)w(df) = \int_{\mathbb{R}^n} \int_{\mathbb{C}} H(f)w(df \mid y)Nw(dy) \]

holds for any \( w \)-integrable \( H : \mathbb{C} \rightarrow \mathbb{R} \).

The Wiener measure \( w \) and, as it turns out, the regular conditional probabilities \( w(\cdot \mid y) \) are Gaussian measures on \( \mathbb{C} \); i.e., any random vector \( f \mapsto (f(t_1), \ldots, f(t_n)) \) is normally distributed with respect to \( w \) and \( w(\cdot \mid y) \). Any Gaussian measure \( \mu \) on \( \mathbb{C} \) is uniquely determined by its mean \( m \in \mathbb{C} \), given by

\[ m(t) = \int_{\mathbb{C}} f(t)\mu(df) \]

and its covariance function

\[ R(s, t) = \int_{\mathbb{C}} (f(s) - m(s)) \cdot (f(t) - m(t))\mu(df) \]

(see Billingsley (1968, p. 64)).

Let \( t_1, \ldots, t_n \in [0, 1] \) be mutually different. Besides the interpolation \( m(\cdot; t_1, \ldots, t_n) \) (recall (1)), we introduce symmetric functions \( R(t_1, \ldots, t_n) \).
. . . , t_n): [0, 1]^2 → [0, 1] to characterize the measures w(·|y). Assume \( t_1 < \cdots < t_n \) first. Then we define

\[
R(t_1, \ldots, t_n)(s, t) = \begin{cases} 
\frac{(t_k - t)(s - t_{k-1})}{t_k - t_{k-1}}, & \text{if } t_{k-1} \leq s \leq t \leq t_k \text{ for } 1 \leq k \leq n, \\
0, & \text{otherwise,}
\end{cases}
\]

(2)

for \( 0 \leq s \leq t \leq 1 \) and \( R(t_1, \ldots, t_n)(s, t) = R(t_1, \ldots, t_n)(t, s) \) for \( 0 \leq t < s \leq 1 \). In general let \( \tau \) be the permutation of \( \{1, \ldots, n\} \) with \( t_{\tau(1)} < \cdots < t_{\tau(n)} \) and put \( R(t_1, \ldots, t_n) = R(t_{\tau(1)}, \ldots, t_{\tau(n)}) \).

Now we are able to state the following proposition, which is well known.

**Proposition 1.** Let \( N \in \mathcal{N}^{ad}_n \) be defined by functions \( x_k: \mathbb{R}^{k-1} \to [0, 1] \) and let \( w_y \) be the Gaussian measure on \( C \) with mean \( m(y; x_1, \ldots, x_n(y_1, \ldots, y_{n-1})) \) and covariance function \( R(x_1, \ldots, x_n(y_1, \ldots, y_{n-1})) \). Then the family \( (w_y)_{y \in \mathbb{R}^n} \) is a version of the regular conditional probability with respect to \( N \).

We sketch a proof for the case \( N \in \mathcal{N}^{adm}_n \) with nodes \( x_1, \ldots, x_n \). Define \( \Delta(f) = f - m(N(f); x_1, \ldots, x_n) \) and \( \mu = \Delta w \). Then \( \mu \) is the Gaussian measure on \( C \) with mean 0 and covariance function \( R = R(x_1, \ldots, x_n) \) and further \( \mu(\ker N) = 1 \) holds. Let \( w_y \) be the translation of \( \mu \) by \( m(y) = m(y|x_1, \ldots, x_n) \). Clearly \( w_y \) is Gaussian with \( w_y(N^{-1}\{y\}) = 1 \), mean \( m(y) \), and covariance function \( R \). Since \( \Delta \) and \( m(N(\cdot); x_1, \ldots, x_n) \) are independent with respect to \( w \), the measure \( w \) is the convolution of \( \mu \) and \( m(N(\cdot); x_1, \ldots, x_n)w \). Therefore properties 2 and 3 hold for \( w(\cdot|y) = w_y \). The case \( N \in \mathcal{N}^{ad}_n \) can be proved by induction (see Traub et al. (1988, p. 474)).

In the following we always assume that the regular conditional probability is given by the family \( (w_y)_{y \in \mathbb{R}^n} \). If \( 0 = x_0 < x_1 < \cdots < x_n = 1 \) are the nodes belonging to the information \( y = N(f) \), then Proposition 1 yields the independence of the random elements \( f \mapsto f \cdot 1_{[x_{k-1}, x_k]} \) for \( k = 1, \ldots, n \) and \( f \mapsto f \cdot 1_{[\cdot, 1]} \) with respect to \( w(\cdot|y) \).

Let \( T > 0, a, b \in \mathbb{R} \), and consider the Gaussian measure on \( C([0, T]) \) with mean \( m(t) = a + t \cdot (b - a)/T \) and covariance function \( R(s, t) = \min(s, t) - st/T \). This measure is the distribution of the so-called Brownian bridge with \( f(0) = a \) and \( f(T) = b \). Obviously any measure \( w(\cdot|y) \) is the distribution of a suitable connection of \( n \) Brownian bridges and a Brownian motion, which are independent.
4. $L_q$-Approximation, $1 \leq q < \infty$

Besides other linear problems on the Wiener space, Lee (1986) studied the $L_2$-approximation with $p = 2$. He derived a formula for the $p$-average radius of an arbitrary nonadaptive information operator and thereby he obtained operators which are $p$-optimal in $N^\text{nom}_n$. Speckman (1979) investigated the case $1 \leq p = q < \infty$ for an autoregressive Gaussian process and he derived asymptotic estimates for $e_p(\phi_p, N, L_p)$ where $N \in N^\text{nom}_n$. In the particular case of the Wiener process it is easy to compute this $p$-average error exactly. For this we need the constants

\begin{equation}
    c_p = (2\pi)^{-1/2} \int_\mathbb{R} |z|^p \exp(-z^2/2) \, dz, \tag{3}
\end{equation}

\begin{equation}
    d_p = \int_0^1 (z(1 - z))^{p/2} \, dz, \tag{4}
\end{equation}

and

\begin{equation}
    a_p = \frac{2}{(d_p(p + 2))^{2/p}}. \tag{5}
\end{equation}

**Theorem 1.** Let $1 \leq p = q < \infty$ and let $N_n \in N^\text{nom}_n$ be defined by fixed nodes

\[ x_k = k \cdot a_p/(1 + na_p). \]

Then the information operator $N_n$ is $p$-optimal in $N^\text{nom}_n$ for the $L_p$-approximation problem with

\[ r_p(N_n, L_p) = (c_p d_p)^{1/p} \cdot (a_p/(1 + na_p))^{1/2}. \]

Further the sequence $N_n^*$ of information operators is strongly asymptotically $p$-optimal in $N^\text{nom}_n$ for the same problem with

\[ r_p(N_n^*, L_p) = (c_p d_p)^{1/p} \cdot n^{-1/2}. \]

Let $1 \leq p < \infty$ and $1 \leq q < \infty$ be arbitrary. Then the sequence $N_n^*$ is weakly asymptotically $p$-optimal in $N^\text{nom}_n$ for the $L_q$-approximation problem with

\[ r_p(N_n^*, L_q) \approx n^{-1/2}. \]

**Proof.** For $N \in N^\text{nom}_n$ let $\phi_p(y) = m(y; x_1, \ldots, x_n)$ be the corresponding spline algorithm. Using the $p$-optimality of $\phi_u$ (see Lee and Wasilkowski (1986)) and Proposition 1 we obtain
where $\mu$ is the Gaussian measure on $C$ with mean 0 and covariance function $R = R(x_1, \ldots, x_n)$ given by (2).

If $p = q$ we get

$$r_p(N, L_p) = \left( \int_{C} \|f - \phi_0 \circ N(f)\|_q^p w(df) \right)^{1/p}$$

$$= \left( \int_{C} \|f - \phi_0(y)\|_q^p w(df | y) Nw(dy) \right)^{1/p}$$

$$= \left( \int_{C} \|f\|_q^p \mu(df) \right)^{1/p}.$$

Since

$$\int_0^1 R(t, t)^{p/2} dt = \sum_{k=1}^n \int_{x_{k-1}}^{x_k} \left( (x_k - t)(t - x_{k-1})/(x_k - x_{k-1}) \right)^{p/2} dt$$

$$+ \int_{x_n}^1 (t - x_n)^{p/2} dt$$

$$= d_p \sum_{k=1}^n (x_k - x_{k-1})^{p/2+1} + 2/(p + 2) \cdot (1 - x_n)^{p/2+1},$$

the $p$-average radius of $N$ for the $L_p$-approximation problem is given by

$$r_p(N, L_p) = (c_p d_p)^{1/p} \cdot \left( \sum_{k=1}^n (x_k - x_{k-1})^{p/2+1} + 2/(d_p(p + 2)) \right)^{1/p} \cdot (1 - x_n)^{p/2+1}$$

(see Lee (1986) for the case $p = 2$). For fixed $x_n \leq 1$ the strict convexity of $x \mapsto x^{p/2+1}$ implies that (6) attains its unique minimum at $x_k = k/n \cdot x_n$. 
Because
\[ n \cdot (n/n)^{p/2+1} + a_p^{p/2} \cdot (1 - x_n)^{p/2+1} \]
becomes minimal if and only if \( x_n = na_p/(1 + na_p) \), we conclude that \( N_n \) is \( p \)-optimal in \( N_n^{\text{nom}} \) for the \( L_p \)-approximation. By evaluating (6) for \( N = N_n \) and \( N = N_n^{*} \) we obtain the \( p \)-average radii stated in the theorem, and \( \lim_{n \to \infty} r_p(N_n, L_p)/r_p(N_n^{*}, L_p) = 1 \) follows immediately.

For arbitrary \( p \) and \( q \) we define \( p_1 = \min(p, q) \) and \( p_2 = \max(p, q) \). Obviously
\[ r_{p_1}(N, L_{p_1}) \leq r_p(N, L_q) \leq r_{p_2}(N, L_{p_2}) \]
holds, and therefore the result for \( p = q \) can be used to prove the weak asymptotic \( p \)-optimality of \( N_n^{*} \) in \( N_n^{\text{nom}} \) with \( r_p(N_n^{*}, L_q) \approx n^{-1/2} \).

The operator \( N_n \) depends on \( p \) through \( a_p \) by the condition
\[ a_p = (x_k - x_{k-1})/(1 - x_n) \]
for \( k = 1, \ldots, n \). Hence for \( p = q \) the constant \( a_p \) is the optimal ratio between the length of the subintervals \( [x_{k-1}, x_k] \), where the conditional probabilities are given by Brownian bridges, and the length of the subinterval \( [x_n, 1] \), where the conditional probabilities are given by Brownian motion. We compute the special values \( a_1 = (16/(3\pi))^2 \approx 2.882, a_3 = 3 \) (see Lee (1986)) and \( \lim_{p \to \infty} a_p = 4 \).

Observe that \( p \)-optimal information operators use equidistant nodes in the case \( p = q \). This result also holds for \( p \geq \max(2q/(2 + q), 1) \) and it can be proved by a simple convexity argument applied to the formula
\[ r_p(N, L_q) = \left( \int_{C^{n+1}} \left( \sum_{k=1}^{n+1} (x_k - x_{k-1})^{q/2+1} \int_0^1 |f_i(t)|^{q} \, dt \right)^{p/q} \right)^{1/p} \]
\[ \mu_0^p \otimes w(d(f_1, \ldots, f_{n+1}))^{1/p}. \]

Here \( p \) and \( q \) are arbitrary, \( x_{n+1} = 1 \), and \( \mu_0 \) denotes the Gaussian measure on \( C \) with mean 0 and covariance function \( R(s, t) = \min(s, t) - st \). Suppose that \( p \)-optimal nodes for the \( L_q \)-approximation are necessarily equidistant for some \( p \) and \( q \). Then we conclude that the sequence \( N_n^{*} \) is strongly asymptotically \( p \)-optimal for the \( L_q \)-approximation in exactly the same way as in the proof of Theorem 2.
5. $L_\infty$-APPROXIMATION

The analysis of the $L_\infty$-approximation problem is more complicated than that for the $L_q$-approximation problem with $1 \leq q < \infty$. But again we can restrict our considerations to spline algorithms using nonadaptive information operators.

In this section let $\mu_0$ denote the Gaussian measure on $C$ with mean 0 and covariance function $R(s, t) = \min(s, t) - st$. We need a lemma concerning the distribution of the $L_\infty$-norm with respect to $\mu_0$, the measure associated with the Brownian bridge with $f(0) = f(1) = 0$. This distribution is characterized by its distribution function

$$F(u) = \mu_0[\|f\|_\infty \leq u] = 1 + 2 \sum_{j=1}^{\infty} (-1)^j \exp(-2j^2u^2)$$

for $u > 0$ (see Billingsley (1968, p. 85)).

**Lemma 1.** Suppose that the nodes $0 < x_1 < \cdots < x_n \leq 1$ are not equidistant; then

$$\prod_{k=1}^{n} F(u/(x_k - x_{k-1})^{1/2}) < F(u/(x_n/n)^{1/2})^n$$

holds for any $u > 0$.

**Proof.** Consider the function $G(u) = F(u^{-1/2})$. We show that $\ln \circ G$ is a strictly concave function to conclude

$$\prod_{k=1}^{n} F(u/(x_k - x_{k-1})^{1/2}) = \prod_{k=1}^{n} G((x_k - x_{k-1})/u^2) < G(x_n/(nu^2))^n$$

$$= F(u/(x_n/n)^{1/2})^n.$$ 

The distribution function $F$ and therefore the function $G$, too, can be expressed in terms of the $\vartheta$ function

$$\vartheta(z, v) = 1 + 2 \sum_{j=1}^{\infty} \cos(2jz)v^j,$$

which is defined for $z, v \in \mathbb{C}$ with $|v| < 1$. The $\vartheta$ function admits the product representation
\[ \vartheta(z, v) = \prod_{j=1}^{\infty} (1 - v^{2j})(1 + 2 \cos(2z)v^{2j-1} + v^{4j-2}) \]  
(8)

(see Erdélyi et al. (1953, p. 357)), and hence we get

\[
\ln \circ G(u) = \ln \circ \vartheta(\pi/2, \exp(-2/u))
\]

\[
= \ln \left( \prod_{j=1}^{\infty} (1 - \exp(-4j/u)) \cdot (1 - \exp(-(4j - 2)/u))^2 \right)
\]

\[
= \sum_{j=1}^{\infty} \ln(1 - \exp(-4j/u)) + 2 \sum_{j=1}^{\infty} \ln(1 - \exp(-(4j - 2)/u)).
\]

Using the Taylor expansion \( \ln(1 - x) = -\sum_{j=1}^{\infty} x^j/j \) for \( 0 < x < 1 \) we obtain

\[
\ln \circ G(u) = -\sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{1}{l} \exp(-4jl/u) - 2 \sum_{j=1}^{\infty} \sum_{l=1}^{\infty} \frac{1}{l} \exp(-(4j - 2)l/u)
\]

\[
= -\sum_{j=1}^{\infty} \frac{1}{l} \cdot \frac{\exp(-4l/u)}{1 - \exp(-4l/u)} - 2 \sum_{j=1}^{\infty} \frac{1}{l} \cdot \exp(2l/u) \cdot \frac{\exp(-4l/u)}{1 - \exp(-4l/u)}
\]

\[
= -\sum_{j=1}^{\infty} \frac{1}{l} \cdot \frac{1 + 2 \exp(2l/u)}{\exp(4l/u) - 1}. \tag{9}
\]

Consider the function

\[ g(u) = \frac{1 + 2 \exp(1/u)}{\exp(2/u) - 1}, \]

where \( u > 0 \), which has the second derivative

\[ g''(u) = \frac{2 \exp(1/u)}{u^3(\exp(2/u) - 1)^3} \cdot h(u) \]

with

\[ h(u) = 2(-\exp(4/u) - \exp(3/u) + \exp(1/u) + 1) \]

\[ + 1/u(\exp(4/u) + 2 \exp(3/u) + 6 \exp(2/u) + 2 \exp(1/u) + 1). \]

A Taylor expansion of \( u \mapsto h(1/u) \) yields \( h > 0 \) on \( ]0, \infty[ \). Therefore \( g \) is strictly convex and by (9) we see that \( \ln \circ G \) is strictly concave.
THEOREM 2. Consider the $L_p$-approximation problem and let $1 \leq p < \infty$. Any information operator which is $p$-optimal in $N_n^{\text{non}}$ uses equidistant nodes. The sequence $N_n^*$ of information operators is strongly asymptotically $p$-optimal in $N_n^{\text{non}}$ with

$$r_p(N_n^*, L_\infty) = (\ln n/(2n))^{1/p}.$$ 

Proof. Let $N \in N_n^{\text{non}}$ be given by nodes $x_1, \ldots, x_n$, and let $\mu$ be the Gaussian measure on $C$ with mean 0 and covariance function $R = R(x_1, \ldots, x_n)$ defined by (2). Since the spline algorithm is $p$-optimal, the same argument as that in the proof of Theorem 1 gives

$$r_p(N, L_\infty) = \left( \int_C \|f\|_p^p \, \mu(df) \right)^{1/p}.$$ 

We define random variables

$$M_k(f) = \max_{x_k \in x_k} |f(t)|$$

for $k = 1, \ldots, n + 1$, where $x_{n+1} = 1$. These variables are independent with respect to $\mu$ and their distribution functions are given by

$$\mu\{M_k(f) \leq u\} = \mu_0(\|f\|_\infty \leq u / (x_k - x_{k-1})^{1/2}) = F(u / (x_k - x_{k-1})^{1/2})$$

for $k = 1, \ldots, n$ and

$$\mu\{M_{n+1}(f) \leq u\} = w(\|f\|_\infty \leq u / (1 - x_n)^{1/2})$$

if $x_n < 1$. We obtain the formula

$$r_p(N, L_\infty) = \left( \int_C \max(M_1(f), \ldots, M_{n+1}(f))^p \mu(df) \right)^{1/p}$$

$$= \left( \int_{C^n} \max(\|f_1\|_\infty, \ldots, (1 - x_n)^{1/2}\|f_{n+1}\|_\infty) \right)^p$$

$$\mu_0^p \otimes w(d(f_1, \ldots, f_{n+1}))^{1/p}$$

(10)

for the $p$-average radius of an arbitrary nonadaptive information operator.

In terms of distribution functions we get
and Lemma 1 implies
\[ r_p(\tilde{N}, L_\infty) < r_p(N, L_\infty), \]
if \( N \) does not use equidistant nodes and \( \tilde{N} \in \mathcal{N}_n^\text{non} \) is defined by nodes \( \tilde{x}_k = k \cdot x_n/n. \)

Next consider the information operator \( N_n^*(f) = (f(1/n), \ldots, f(1)) \) and assume \( n \geq 2 \). By (11) its \( p \)-average radius satisfies
\[
 r_p(N_n^* L_\infty \cdot (n/\ln n)^{1/2} = \left( \int_0^\infty \left( 1 - F(u^{1/p} \cdot (\ln n)^{1/2}) \right) du \right)^{1/p}. \tag{12}
\]

To compute the pointwise limit of the integrand we use the estimate
\[
 1 - 2 \exp(-2u^2) \leq F(u) \leq 1 - \exp(-2u^2),
\]
which follows from (7), (8), and \( F(u) = \vartheta(\pi/2, \exp(-2u^2)) \). If \( u < 2^{-p/2} \), we have
\[
 F(u^{1/p} \cdot (\ln n)^{1/2})^n \leq (1 - \exp(-2u^{2p} \cdot \ln n))^n = (1 - n^{-2u^{2p}})^n,
\]
which implies
\[
 \lim_{n \to \infty} F(u^{1/p} \cdot (\ln n)^{1/2})^n = 0. \tag{13}
\]

If \( u > 2^{-p/2} \), we have
\[
 F(u^{1/p} \cdot (\ln n)^{1/2})^n \geq (1 - 2 \exp(-2u^{2p} \cdot \ln n))^n \geq 1 - 2n^{1-2u^{2p}},
\]
which implies
\[
 \lim_{n \to \infty} F(u^{1/p} \cdot (\ln n)^{1/2})^n = 1 \tag{14}
\]
and
\[
 F(u^{1/p} \cdot (\ln n)^{1/2})^n \geq 1 - 4 \exp(-2 \ln 2 \cdot u^{2p}).
\]
By combining (12), (13), and (14) we get
\[ \lim_{n \to \infty} r_p(N_n^*, L_z) \cdot (n/\ln n)^{1/2} = 2^{-1/2} \] (15)

by means of Lebesgue’s convergence theorem.

Since the \( p \)-average radius of \( N(f) = (f(x_1/n), \ldots, f(k \cdot x_n/n), \ldots, f(x)) \) depends continuously on \( x \in [0, 1] \) and since the restriction to operators of this kind is possible, there exists a sequence of information operators \( N_n \) such that \( N_n \) is \( \mu \)-optimal in \( \mathcal{N}_p^\mu \). Assume \( N_n(f) = (f(x_n^{(n)}/n), \ldots, f(x)) \). Because of (10) and \( \lim_{n \to \infty} r_p(N_n, L_z) = 0 \) we get \( \lim_{n \to \infty} x_n^{(n)} = 1 \) and by (10) we also have

\[ r_p(N_n, L_z) \geq (x_n^{(n)}/n)^{1/2} \left( \int_{C_n} \max(\|f_1\|_{L_p}, \ldots, \|f_n\|_{L_p})^p \mu_0(d(f_1, \ldots, f_n)) \right)^{1/p}. \]

Hence we obtain \( \lim_{n \to \infty} r_p(N_n, L_z)/r_p(N_n^*, L_z) = 1. \)

For the \( L_\infty \)-approximation the above proof shows that it is sufficient to consider information operators using equidistant nodes. In the case \( p \geq 2 \) a simpler argument yields this fact. Fix \( 0 \leq x_n \leq 1 \). Then the function \( (x_1, \ldots, x_n-1) \mapsto \max(x_1^{p/2} c_1, \ldots, (x_1 - x_n-1)^{p/2} c_n, d) \) is convex on the simplex \( \{x_i, \ldots, x_n-1| 0 \leq x_1 \leq \cdots \leq x_n-1 \leq x_n \} \) for any \( c_1, \ldots, c_n, d \geq 0 \). Because of (10) we conclude that \( r_p(\cdot, L_\infty)^p \) defines a convex function on this set, too. Further the radius only depends on the lengths of the subintervals \( [x_{k-1}, x_k] \subseteq [0, x_n] \) and it is independent of the order of these intervals. Therefore we do not increase the radius if we take the operator \( N(f) = (f(x_1/n), \ldots, f(x)) \) instead of \( N(f) = (f(x_1), \ldots, f(x_n)) \).

6. **Global Optimization**

In this section let \( \Phi \) denote the distribution function of the standard normal distribution, i.e.,

\[ \Phi(u) = (2\pi)^{-1/2} \int_{-\infty}^{u} \exp(-z^2/2) \, dz. \]

First we state some facts concerning the random variable \( f \mapsto \max_{0 \leq t \leq T} f(t) \), where \( 0 < T \leq 1 \). The common distribution of this random variable and \( f \mapsto f(T) \) with respect to the Wiener measure is characterized by
Consider the Gaussian measure \( \mu \) on \( C([0, T]) \) with mean \( m(t) = a + t \cdot (b - a)/T \) and covariance function \( R(s, t) = \min(s, t) - st/T \); i.e., \( \mu \) is the measure associated with the Brownian bridge with \( f(0) = a \) and \( f(T) = b \). The distribution function of the maximum with respect to \( \mu \) is given by

\[
\mu \{ \max_{0 \leq t \leq T} f(t) \leq u \} = \begin{cases} 
1 - \exp(-2(u - a)(u - b)/T), & \text{if } u \geq \max(a, b), \\
0, & \text{otherwise}.
\end{cases}
\]  

(18)

Proofs of these results are given in Billingsley (1968, Chap. II) for the case \( T = 1 \) and \( a = b = 0 \), but these proofs also apply to arbitrary \( T, a, \) and \( b \).

Let \( N \in N_d \) be defined by functions \( x_k: \mathbb{R}^k \rightarrow [0, 1] \). Then a reasonable algorithm for the optimization problem is given by

\[
\phi_0(y) = \min \{ x_k(y_1, \ldots, y_{k-1}) \mid y_k = \max_{i=0,\ldots,n} y_i \}.
\]

Obviously this algorithm is measurable and it solves the optimization problem for the mean \( m(y; x_1, \ldots, x_n(y_1, \ldots, y_{n-1})) \) of the conditional probability \( w(\cdot | y) \).

Our discussion on the \( p \)-optimality of \( \phi_0 \) is based on a lemma concerning the distribution of

\[
Z(f) = \max_{0 \leq t \leq 1} f(t)
\]

with respect to the Brownian bridge with \( f(0) = 0 \) and \( f(1) = 1 \).

**Lemma 2.** Let \( \mu \) be the Gaussian measure on \( C \) with mean \( m(t) = t \) and covariance function \( R(s, t) = \min(s, t) - st \) and let \( 0 \leq x < 1 \). Then
holds for any \( u > 0 \).

**Proof.** Because the lemma is obvious for \( x = 0 \) we assume \( x > 0 \). Due to (18) we have

\[
\mu_1\{Z(f) - 1 \leq u\} = 1 - \exp(-2u(u + 1))
\]

and

\[
\mu_1\{Z(f) - f(x) \leq u\}
\]

\[
= (2\pi x(1 - x))^{-1/2} \int_{1-u-x}^{\infty} (1 - \exp(-2u + x + hu/x)) \cdot (1 - \exp(-2u(1 + x + h - 1)/(1 - x))) \cdot \exp(-h^2/(2x(1 - x))) dh
\]

for \( u > 0 \), since the regular conditional probability of \( \mu_1 \) with respect to \( f \mapsto f(x) - x \) is given by two independent Brownian bridges with \( f(0) = 0 \), \( f(x) = x + h \) and \( f(x) = x + h, f(1) = 1 \).

Let \( \frac{1}{2} \leq x < 1 \). It is easy to verify that

\[
(1 - \exp(-2(u + x + h)u/x)) \cdot (1 - \exp(-2u(u + x + h - 1)/(1 - x))) \geq (1 - \exp(-2(u + 1 - x + h)u/(1 - x))) \cdot (1 - \exp(-2u(u - x + h)/x))
\]

holds for \( u > 0 \) and \( h \leq -u + x \). Therefore we get

\[
\mu_1\{Z(f) - f(1 - x) \leq u\} \leq \mu_1\{Z(f) - f(x) \leq u\}
\]

and a restriction to \( \frac{1}{2} \leq x < 1 \) is possible.

In this case we have the estimate

\[
\mu_1\{Z(f) - f(x) \leq u\}
\]

\[
< (2\pi x(1 - x))^{-1/2} \int_{1-u-x}^{\infty} (1 - \exp(-2u + x + hu/x)) \cdot \exp(-h^2/(2x(1 - x))) dh
\]

\[
= 1 - \Phi((1 - u - x)/(x(1 - x))^{1/2})
\]

\[
\exp(2u(u + 1)) \cdot (1 - \Phi((1 + u - x - 2ux)/(x(1 - x))^{1/2}))
\]

\[
= \mu_1\{Z(f) - 1 \leq u\} - \Phi(((1 - x)/x)^{1/2} - u/(x(1 - x))^{1/2})
\]

\[
+ \exp(-2u(u + 1)) \cdot \Phi(((1 - x)/x)^{1/2} - (2x - 1)u/(x(1 - x))^{1/2}).
\]
Hence it is enough to prove

$$\Phi(((1 - x)/x)^{1/2} - u/(x(1 - x))^{1/2})$$

$$\geq \exp(-2u(u + 1)) \cdot \Phi(((1 - x)/x)^{1/2} - (2x - 1)u/(x(1 - x))^{1/2}),$$

which follows from

$$\Phi(-u/(x(1 - x))^{1/2}) \leq \exp(-2u(u + 1)) \cdot \Phi(-u(d^2 - 4)^{1/2})$$

because $c \mapsto \Phi(c - c_1)/\Phi(c - c_2)$ is an increasing function for $c_1 \geq c_2$.

Substituting $d = (x(1 - x))^{-1/2} \geq 2$ we consider the function

$$G(u, d) = \Phi(-ud) - \exp(-2u(u + 1)) \cdot \Phi(-u(d^2 - 4)^{1/2}).$$

Since $G(u, \cdot)$ is increasing on $[2, d_u]$ and decreasing on $[d_u, \infty[$ with a suitably chosen $d_u \geq 2$ and since $G(u, 2) \geq 0$ and $\lim_{u \to \infty} G(u, d) = 0$, we conclude $G \geq 0$.

**Theorem 3.** Assume that $p = 1$ or that $p > 1$ and $1 \in \{x_1, \ldots, y_{n-1}\}$ holds $Nw$ a.e. Then the algorithm $\phi_n$ is $p$-optimal for the global optimization problem.

**Proof.** Let $\phi : \mathbb{R}^n \to [0, 1]$ be an arbitrary algorithm; then

$$e_1(\phi, N, \text{Opt}) = \int_C Z(f)w(df) - \int_{\mathbb{R}^n} \int_C f(\phi(y))w(df|y)Nw(dy)$$

$$= \frac{2}{\pi}^{1/2} - \int_{\mathbb{R}^n} m(y; x_1, \ldots, x_n(y_1, \ldots, y_{n-1}))$$

$$\geq \frac{2}{\pi}^{1/2} - \int_{\mathbb{R}^n} \max(0, y_1, \ldots, y_n)Nw(dy)$$

$$= e_1(\phi_n, N, \text{Opt}),$$

and the theorem is established for $p = 1$.

For arbitrary $p$ we have

$$e_p(\phi, N, \text{Opt}) = \left(\int_{\mathbb{R}^n} \int_C (Z(f) - f(\phi(y)))^p w(df|y)Nw(dy)\right)^{1/p}.$$
for any $x \in ]x_{k-1}, x_k[ = A_k$. Because $f \mapsto (\max_{i \in A_i} f(t), f(x))$ and $f \mapsto \max_{t \in [0, 1]} f(t)$ are independent with respect to $w(\cdot | y)$ it is sufficient to show that

$$
\int_c (\max(c, \max_{i \in A_i} f(t)) - \max(y_{k-1}, y_k)) \rho w(df | y) < \int_c (\max(c, \max_{i \in A_i} f(t)) - f(x)) \rho w(df | y)
$$

for any $c \geq \max(0, y_1, \ldots, y_n)$ and $x \in A_k$. The proof of this inequality can be reduced to the following situation. Let $\mu_1$ be the Gaussian measure on $\mathbb{C}$ as defined in Lemma 2. Then we show that

$$
\int_c (\max(c, Z(f)) - 1)^p \mu_1(df) < \int_c (\max(c, Z(f)) - f(x))^p \mu_1(df)
$$

holds for any $c \geq 1$ and $0 \leq x < 1$. Observe that this inequality holds for $p = 1$ and in general it is equivalent to

$$
\int_0^\infty u^{p-1}(\mu_1\{\max(c, Z(f)) - 1 \leq u\} - \mu_1\{\max(c, Z(f)) - f(x) \leq u\}) \mu_1(df) > 0. \quad (19)
$$

Let

$$
H(u, c) = \mu_1\{\max(c, Z(f)) - 1 < u\} - \mu_1\{\max(c, Z(f)) - f(x) \leq u\} = \mu_1(Z(f) - 1 \leq u),
$$

$$
\leq \mu_1\{Z(f) - 1 \leq u\} - \mu_1\{Z(f) - f(x) \leq u\}.
$$

Therefore we have $H(\cdot, c) \leq 0$ on $[0, c - 1]$ and $H(\cdot, c) > 0$ on $]c - 1, \infty[$ because of Lemma 2. Since (19) holds for $p = 1$ we conclude that it holds for any $p \geq 1$.

The following example shows that the algorithm $\phi_p$ is not $p$-optimal in general. Consider fixed nodes $0 < x_1 < \cdots < x_n < 1$ and fixed function values $y_k = f(x_k)$ for $k = 1, \ldots, n - 1$, and assume $y_n > \max(0, y_1, \ldots, y_{n-1})$. Then we have $\phi_p(y) = x_n$, but it is better to choose $(1 + x_n)/2$, if $y_n$ is sufficiently large and $p = 2$. This is due to the fact that

$$
h(x) = \int_c (Z(f) - f(x))^2 w(df)
$$

attains its unique minimum on $[0, 1]$ at $x = \frac{1}{2}$. 


Now we study nonadaptive information operators, which are defined by fixed nodes \(x_1, \ldots, x_n\). First we consider the case \(p = 1\). If \(\nu\) denotes the \(n\)-dimensional standard normal distribution, Theorem 3 gives

\[
   r_1(N, \text{Opt}) = (2/\pi)^{1/2} - \int_{\mathbb{R}^n} \max(0, y_1, \ldots, y_n) \, Nw(dy)
\]

\[
   = (2/\pi)^{1/2} - \int_{\mathbb{R}^n} \max \left( \sum_{i=1}^{k} (x_i - x_{i-1})^{\frac{1}{2}} \cdot y_i \right) \nu(dy).
\]

Therefore \(x_n = 1\) is a necessary condition for \(N\) to be \(p\)-optimal in \(N_{n}^{\text{non}}\) with \(p = 1\). In the special case \(x_k = k/n\) the last integral is the expectation of the maximal positive part of partial sums of i.i.d. random variables, each of them normally distributed with mean 0 and variance 1/n. We use a formula if Kac (1954) for this expectation, which yields

\[
r_1(N^n_{n^*}, \text{Opt}) = (2/\pi)^{1/2} - (2\pi n)^{-1/2} \sum_{k=1}^{n} k^{-1/2}.
\]

Since

\[
   2(n + 1)^{1/2} - 2 \leq \sum_{k=1}^{n} k^{-1/2} \leq 2n^{1/2} - 1,
\]

we conclude

\[
r_1(N^n_{n^*}, \text{Opt}) \approx n^{-1/2}.
\]

For \(n = 2\) an elementary computation shows that \(N^n_{2^*}\) is \(p\)-optimal. For \(n = 3\) and \(N(f) = (f(x_1), f(x_2), f(1))\) we compute

\[
r_1(N, \text{Opt}) \cdot (2\pi)^{3/2}
\]

\[
   = 7\pi/2 - \pi/2(x_2^{1/2} + (1 - x_1)^{1/2} + 2x_1^{1/2} + (x_2 - x_1)^{1/2} + 2(1 - x_2)^{1/2})
\]

\[
   + \arctan((x_1(1 - x_2)/(x_2 - x_1))^{1/2}) - x_1^{1/2} \arctan(((1 - x_2)/(x_2 - x_1))^{1/2})
\]

\[
   - (1 - x_2)^{1/2} \arctan((x_1/(x_2 - x_1))^{1/2}).
\]

A numerical evaluation of this formula shows \(r_1(N, \text{Opt}) < r_1(N^n_{3^*}, \text{Opt})\) for \(N(f) = (f(\frac{x_1}{2}), f(\frac{x_2}{2}), f(1))\). Hence the operator \(N^n_{3^*}\) is not \(p\)-optimal in \(N_{n}^{\text{non}}\) in general, but we prove the weak asymptotic \(p\)-optimality of the sequence \(N^n_{n^*}\) in the following.
For $0 \leq t_1 < t_2 \leq 1$ we define the random variables

$$Y(f) = \max_{t_1 \leq t \leq t_2} f(t) - \max_{0 \leq t \leq 1} f(t),$$

$$Y_1(f) = \max_{0 \leq t \leq t_1} f(t) - f(t_1),$$

$$Y_2(f) = \max_{t_1 \leq t \leq t_2} f(t) - f(t_1),$$

$$Y_3(f) = f(t_2) - f(t_1),$$

$$Y_4(f) = \max_{t_1 \leq t \leq 1} f(t) - f(t_2),$$

which have the following properties.

1. $Y = Y_2 - \max(Y_1, Y_3 + Y_4)$.
2. $Y_1, (Y_2, Y_3),$ and $Y_4$ are independent with respect to the Wiener measure.
3. $Y_1$ is distributed like $f \mapsto t_1^{1/2} \cdot |f(1)|$ and $Y_4$ is distributed like $f \mapsto (1 - t_2)^{1/2} \cdot |f(1)|$.
4. The common distribution of $Y_2$ and $Y_3$ has the density

$$h(y_2, y_3) = \begin{cases} 
\frac{(2/\pi)^{1/2} 2y_2 - y_3}{(t_2 - t_1)^{3/2}} \exp \left( - \frac{(2y_2 - y_3)^2}{2(t_2 - t_1)} \right), & \text{if } y_2 \geq \max(0, y_3), \\
0, & \text{otherwise.}
\end{cases}$$

The Markov property of the Brownian motion yields property 2, and together with (17) we get the distribution of $Y_4$. Since the Brownian motion is invariant under time inversion on $[0, t_1]$, the random variable $Y_1$ is distributed like $f \mapsto \max_{0 \leq t \leq t_1} f(t)$. Hence 3 is established completely. We obtain the density $h$ by differentiation of (16).

For $f \in C$ we consider the location of the global maxima and we define

$$B = \{ f \in C \mid f(t) = Z(f) \text{ with } t_1 \leq t \leq t_2 \}$$

and

$$B' = \{ f \in C \mid f(t) = Z(f) \text{ with } 0 \leq t \leq t_1 \text{ or } t_2 \leq t \leq 1 \}.$$
and

$$\int_B Y_2(f)w(df) \leq c_2 \cdot \begin{cases} (t_2 - t_1)^{p/2 + 1/(t_1(1 - t_2))^{1/2}}, & \text{if } t_1 > 0 \text{ and } t_2 < 1, \\ t_2^{(p+1)/2}/(1 - t_2)^{1/2}, & \text{if } t_1 = 0 \text{ and } t_2 < 1, \\ (1 - t_1)^{(p+1)/2}/t_1^{1/2}, & \text{if } t_1 > 0 \text{ and } t_2 = 1, \end{cases}$$

hold for any $0 \leq t_1 < t_2 < 1$. The sets $B$ and $B'$ satisfy $w(B \cap B') = 0$.

**Proof.** Let $\sigma = (t_2 - t_1)^{1/2}$. Then we have

$$\left(\pi/2\right)^{3/2} \int_B Y(f)w(df) = \left(\pi/2\right)^{3/2} \int_C Y^+(f)w(df)$$

$$= \left(\pi/2\right)^{1/2} \int_{\mathbb{R}^2} \left( y_2 - \max(t_1^{1/2}y_1, y_3 + (1 - t_2)^{1/2}y_4) \right)^+ \cdot \exp\left(-\left(y_1^2 + y_3^2\right)/2\right) \cdot h(y_2, y_3) \, dy$$

$$\geq \left(\pi/2\right)^{1/2} \int_{\mathbb{R}^2} \left( y_2 - \max(y_1, y_3 + y_4) \right)^+ \cdot \exp\left(-\left(y_1^2 + y_3^2\right)/2\right) \cdot h(y_2, y_3) \, dy$$

$$\geq \int_{\mathbb{R}^2} \int_0^{y_2} \int_{y_1 - y_4}^{y_2 - y_4} \left( y_2 - y_3 - y_4 \right) \left( 2y_2 - y_3 \right) / \sigma^3 \cdot \exp\left(-\left(y_2 - y_3\right)^2/(2\sigma^2)\right) \, dy_3 \cdot \exp\left(-y_1^2/2\right) \, dy_1 \, d(y_2, y_4),$$

if we only consider the set $\{y_2 \geq y_3 + y_4 \geq y_1\}$.

Fix $y_2, y_4 \geq 0$. Since the integral with respect to $y_3$ is a positive and decreasing function of $y_1$ we obtain

$$\int_0^{y_2} \int_{y_1 - y_4}^{y_2 - y_4} \left( y_2 - y_3 - y_4 \right) \left( 2y_2 - y_3 \right) / \sigma^3 \cdot \exp\left(-\left(y_2 - y_3\right)^2/(2\sigma^2)\right) \, dy_3 \cdot \exp\left(-y_1^2/2\right) \, dy_1$$

$$\geq \left(2\pi\right)^{1/2} \cdot \left(\Psi(\sqrt{y_2/2}) - \frac{1}{2}\right) \cdot \int_{y_2/2 - y_4}^{y_2 - y_4} \left( y_2 - y_3 - y_4 \right) \left( 2y_2 - y_3 \right) / \sigma^3 \cdot \exp\left(-\left(y_2 - y_3\right)^2/(2\sigma^2)\right) \, dy_3$$

$$= 2\pi \cdot \left(\Psi(\sqrt{y_2/2}) - \frac{1}{2}\right) \cdot \left(\Psi(-y_2 + y_4)/\sigma - \Phi(-y_2 + y_4)/\sigma - y_2/(2\sigma) - y_2/(2\sigma) \cdot \Phi'(\sqrt{-y_2 + y_4}/\sigma - y_2/(2\sigma))$$

by partial integration. Because of the asymptotic behavior of $\Phi$ there exists $d > 0$ with
\[ \Phi(-a) - \Phi(-a - b) - b \cdot \Phi'(-a - b) \geq 1/(2\pi) \cdot \exp(-a^2/2)/a \]

for all \( a \geq 2d \) and \( b > d/2 \). Hence

\[
\int_{y_1}^{y_2} \int_{y_3}^{y_4} (y_2 - y_3 - y_4)(2y_2 - y_3)/\sigma^3 
\cdot \exp(-(2y_2 - y_3)^2/(2\sigma^2)) \, dy_3 \exp(-y_1^2/2) \, dy_1
\geq (\Phi(y_2/2) - \frac{1}{2}) \cdot \sigma/(y_2 + y_4) \cdot \exp(-(y_2 + y_4)^2/(2\sigma^2))
\]

holds for \( y_2, y_4 \geq \sigma d \) and from (21) we have

\[
(\pi/2)^{3/2} \int_B Y(f)w(df) \geq \sigma
\]

\[
\geq \sigma 
\cdot \int_{[\sigma,\infty]} (\Phi(y_2/2) - \frac{1}{2})/(y_2 + y_4) \cdot \exp(-(y_2 + y_4)^2/(2\sigma^2)) 
\cdot \exp(-y_2^2/2) \, dy_2 \, dy_4
\geq \sigma^2 \cdot (\Phi(\sigma d/2) - \frac{1}{2}) 
\cdot \int_{[\sigma,\infty]} 1/(y_2 + y_4) \cdot \exp(-(y_2 + y_4)^2/2) \cdot \exp(-y_4^2/2) \, dy_2 \, dy_4.
\]

Therefore we conclude

\[
\int_B Y(f)w(df) \geq c_1 \sigma^\xi
\]

with a suitably chosen \( c_1 > 0 \).

Now we prove the second inequality for the case \( t_1 > 0 \) and \( t_2 < 1 \)

Observe that \( B = \{Y_2 \geq Y_1, Y_2 - Y_3 \geq Y_4\} \), and therefore

\[
(\pi/2)^{3/2} \int_B Y_2^p(f)w(df)
= (\pi/2)^{1/2} \int_{R_2} \int_{y_3/2}^{y_2} \exp(-y_1^2/2) \, dy_1 \cdot \int_{0}^{y_2-y_3}(1-t_2)y_2^2 \exp(-y_4^2/2) \, dy_4
\cdot y_2^p \cdot h(y_2, y_3) \, dy_2 \, dy_3
= \sigma^p \int_{-\infty}^{\infty} \int_{-\infty}^{y_2} \int_{y_3}^{y_2} \exp(-y_1^2/2) \, dy_1 \cdot \int_{0}^{\sigma(y_2-y_3)/(1-t_2)y_2^2} \exp(-y_4^2/2) \, dy_4
\cdot y_2^p \cdot (2y_2 - y_3) \cdot \exp(-(2y_2 - y_3)^2/2) \, dy_3 \, dy_2
\leq \sigma^{p+2}/(t_1(1 - t_2))^{1/2} \cdot c_2
\]
with
\[ c_2 = \int_0^\infty \int_0^{y_2} y_2^{p+1}(2y_2 - y_3)^2 \cdot \exp(-(2y_2 - y_3)^2/2) \, dy_3 \, dy_2 < \infty. \]

The other cases can be proved analogously.

Since
\[ B \cap B' \subset \{Y_1 = Y_2\} \cup \{Y_2 - Y_3 = Y_4\} \]

and since the common distribution of the random variables \( Y_i \) is absolutely continuous with respect to the Lebesgue measure we get \( w(B \cap B') = 0. \]

Due to Lemma 3 the global maximum of \( f \in C \) is uniquely defined \( w \), a.e. and its location in \([0, 1]\) is distributed according to the arcsine law (see Billingsley (1968, p. 86)).

**Theorem 4.** Let \( 1 \leq p < \infty \). Then the sequence of \( N_\infty^* \) of information operators is weakly asymptotically \( p \)-optimal in \( N^* \) for the global optimization with
\[ r_p(N_\infty^*, Opt) \approx n^{-1/2}. \]

**Proof.** First we derive a lower bound for the \( p \)-average radius of an arbitrary information operator in the case \( p = 1 \). Without loss of generality we may assume \( x_n = 1 \). Let
\[ B_k = \{ f \in C \mid Z(f) = f(t) \text{ with } x_{k-1} \leq t \leq x_k \} \]

and assume \( n \geq 3 \). We use Theorem 3 and Lemma 3 with \( t_1 = x_{k-1} \) and \( t_2 = x_k \) to conclude
\[ r_1(N, Opt) = e_1(\phi, N, Opt) = \sum_{k=1}^n \int_{B_k} (Z(f) - \max_{l=0,\ldots,n} f(x_l))w(df) \]
\[ \geq \sum_{k=1}^n \int_{B_k} (\max_{x_{k-1} \leq t \leq x_k} f(t) - \max_{0 \leq t \leq x_{k-1}} f(t))w(df) \]
\[ \geq c_1 \sum_{k=1}^n (x_k - x_{k-1})^{3/2} \geq c_1/n^{1/2}, \]

and by combining this inequality with (20) we get the theorem for \( p = 1 \).
Now we derive an upper bound for $r_p(N_n^*, \text{Opt})$ in the case $p \geq 1$ and $n \geq 4$. Again we apply Theorem 3 and Lemma 3 with $t_1 = x_{k-1} = (k - 1)/n$ and $t_2 = x_t = k/n$. We obtain

$$r_p(N_n^*, \text{Opt}) = e_p(\phi_n, N_n^*, \text{Opt}) = \left(\sum_{k=1}^{n} \int_{B_k} (Z(f) - \max_{l=0, \ldots, n} f(x_l))^{p} w(df)\right)^{1/p}$$

$$\leq \left(\sum_{k=1}^{n} \int_{B_k} \left( \max_{x_{k-1} \leq t \leq x_k} f(t) - f(x_{k-1})\right)^{p} w(df)\right)^{1/p}$$

$$\leq \left(c_2 \cdot n^{-p/2} \cdot \left(2/(n - 1)^{1/2} + \sum_{k=2}^{n-1} ((k - 1)(n - k))^{-1/2}\right)^{1/p}\right)$$

$$\leq c_2^{1/p} n^{-1/2} \cdot \left(2 + \sum_{k=1}^{n-1} (k(n - 1 - k))^{-1/2}\right)^{1/p}.$$ 

Take $j \geq 2$ with $n = 2j$ or $n = 2j + 1$. Then we have

$$\sum_{k=1}^{n-2} (k(n - 1 - k))^{-1/2} \leq 2\sum_{k=1}^{j-1} (k(n - 1 - k))^{-1/2} + (j(n - 1 - j))^{-1/2}$$

$$\leq 2/(n - 2)^{1/2} + 2 \int_{1}^{j-1} (k(n - 1 - k))^{-1/2} dk + 1$$

$$\leq 3 + \pi.$$ 

Hence the above sum is uniformly bounded. $\blacksquare$

In the last part of this section we compare adaptive and nonadaptive methods for the global optimization problem. We restrict our considerations to $p = 1$, because Theorem 3 gives a simple characterization of optimality in this case: an information operator is $p$-optimal in $N_n^\text{ad}$ if and only if it maximizes the expectation of

$$\max(0, f(x_1), \ldots, f(x_n)(f(x_1), \ldots, f(x_{n-1})))$$

with respect to the Wiener measure. The Wiener space approach may be considered as a special statistical model of an objective function. In a general model methods $\phi_n \circ N$ which are $p$-optimal for $p = 1$ are called Bayesian methods. They can be characterized by a system of recurrent equations of dynamic programming (see Mockus (1989)), but this does not yield an explicit solution in the case of the Wiener measure.
Theorem 5. Adaption helps for the global optimization problem with \( p = 1 \), i.e.,

\[
\inf \{ r_1(N, \text{Opt}) \mid N \in \mathcal{N}_n^{\text{ad}} \} < \inf \{ r_1(N, \text{Opt}) \mid N \in \mathcal{N}_n^{\text{nom}} \}
\]

holds for \( n > 1 \).

Sketch of the Proof. Because the \( p \)-average radius of a nonadaptive information operator depends continuously on its nodes, there exists a \( p \)-optimal information operator in \( \mathcal{N}_n^{\text{nom}} \). This operator \( N(f) = (f(x_1), \ldots, f(x_{n-1}), f(1)) \) can be improved by first evaluating at \( n - 1 \) of its nodes and then choosing the last node adaptively.

Let \( a, b, c \in \mathbb{R}, c \geq 0 \), and define

\[
h(x) = \int_c \max(f(x), c) \mu(df)
\]

for \( 0 \leq x \leq 1 \), where \( \mu \) denotes the distribution of the Brownian bridge with \( f(0) = a \) and \( f(1) = b \). Since the location of the maximum of \( h \) depends on the parameters \( a, b, \) and \( c \), we conclude that the optimal node \( x_n \) depends on the previously computed function values. \( \blacksquare \)

If we apply the above contribution to a sequence \( N_n \in \mathcal{N}_n^{\text{nom}} \) of information operators, where \( N_n \) is \( p \)-optimal in \( \mathcal{N}_n^{\text{nom}} \), we obtain a sequence \( \tilde{N}_n \in \mathcal{N}_n^{\text{ad}} \) of adaptive information operators with strictly smaller \( p \)-average radius for any \( n > 1 \) but

\[
r_1(\tilde{N}_n, \text{Opt}) = r_1(N_n, \text{Opt}).
\]

Therefore this construction does not answer the question of whether adaptive methods for global optimization yield a better rate of convergence of the \( p \)-average radius than \( n^{-1/2} \). A further investigation of this question seems to be interesting.

References


