Complexity of Approximation with Relative Error Criterion in Worst, Average, and Probabilistic Settings*

T. JACKOWSKI AND H. WOŹNIAKOWSKI

Department of Computer Science, Columbia University, New York, New York 10027; and Institute of Informatics, University of Warsaw, Warsaw, Poland

The complexity of approximating a continuous linear functional defined on a separable Banach space equipped with a Gaussian measure is studied. The quality of the approximation is measured by a relative error criterion. The complexity is studied in the worst case, average case, and probabilistic settings. In the worst and average case settings, the complexity is infinite. In the probabilistic setting, the complexity is finite under a mild assumption. Tight lower and upper complexity bounds are established and an almost optimal algorithm is constructed. We briefly indicate how some of the results generalize for linear operators. In particular, in the worst case setting the complexity remains infinite, whereas in the average case setting the complexity becomes finite if the dimension of the range of a linear operator is at least two. © 1987 by Academic Press, Inc.

I. INTRODUCTION

Complexity is understood as the minimal cost required for computing approximations with a preassigned error. The complexity of approximating continuous linear functionals (and operators) has been recently analyzed in a number of papers. A survey may be found in Woźniakowski (1986a).

Different error criteria have been considered in the worst case, average case, and probabilistic settings. Examples include absolute, normalized, and more general definitions of error (see, e.g., Lee and Wasilkowski, 1986).

In this paper we study relative error in the worst case, average case, and probabilistic settings. The relative error is often used in practice to measure the quality of an approximation.

We show that the complexity is infinite in the worst case and average

* This research was supported in part by the National Science Foundation under Contract DCR-82-14327.

0085-064X/87 $3.00
Copyright © 1987 by Academic Press, Inc.
All rights of reproduction in any form reserved.
case settings. That is, it is impossible to compute an approximation with relative error less than one with finite cost. A precise statement is given in Section 2.

In the probabilistic setting, the complexity if finite under a mild assumption presented in Section 2. We provide tight complexity bounds as well as an algorithm which solves the problem with almost minimal cost. For example, for the integration problem defined on $k$ times continuously differentiable functions equipped with the Wiener measure placed on $k$th derivatives, the complexity is proportional to $(\varepsilon \delta)^{-1/(k+1)}$. Here $\varepsilon$ is a bound on the relative error for functions belonging to a set of measure $1 - \delta$. The algorithm based on the integral of the natural spline of degree $2k + 1$ which interpolates the function at $n$ equally spaced points, $n = \Theta((\varepsilon \delta)^{-1/(k+1)})$, solves the problem with almost minimal cost.

Although the paper is focused on linear functionals, we briefly mentioned how some results generalize for linear operators. The same proof technique yields that in the worst case setting, the complexity of approximating linear operators with relative error less than one remains infinite. In the average case setting, the complexity becomes finite if the range of a linear operator is at least two dimensional. A precise statement is given in Section 8. Thus, unlike linear functionals, linear operators can be approximated under the relative error criterion on the average. The detailed study of the average case and probabilistic settings for approximation of linear operators with relative error will be reported in the future.

We also discuss a modification of relative error, where the distance between two elements $a$ and $b$ of a normed linear space is given by $\|a - b\|/(\|a\| + \eta)$ for a (small) positive $\eta$. Then the complexity under this error criterion becomes finite in all the settings under mild assumptions and is related to the complexity under the absolute error criterion as reported in Section 8.

We outline the contents of this paper. The main results are stated in Section 2. Section 3 contains additional facts explaining how approximations are computed. The worst case result is proved in Section 4. Section 5 contains properties of Gaussian measures needed to analyze the average case and probabilistic settings. The average case result is proved in Section 6. The major technical section of the paper is Section 7, where the probabilistic setting is analyzed. In the final section we briefly discuss modifications of relative error and extensions to approximation of linear operators.

2. Complexity Results

Let $S$, called a solution functional, be a continuous linear functional

$$S : F \rightarrow \mathbb{R},$$
where $F$ is a separable Banach space over the real field. We approximate $S f$ by $U(f)$, $f \in F$, and the quality of approximation is measured by the relative error criterion,

$$\frac{|S f - U(f)|}{|S f|}.$$  \hspace{1cm} (2.1)

We use the convention that $0/0 = 0$.

We shall analyze the relative error in three settings: worst case, average case, and probabilistic.

The relative error in the worst case setting is defined as

$$e^w(U) = \sup_{f \in F} \frac{|S f - U(f)|}{|S f|}. \hspace{1cm} (2.2)$$

In the average case setting we assume that the space $F$ is equipped with a Gaussian measure $\mu$ of mean zero and a one-to-one correlation operator $C_\mu$, where $C_\mu : F^* \to F$. The definition and basic properties of Gaussian measures may be found in Kuo (1975) and Vakhania (1981). The relative error is now defined as

$$e^{\text{avg}}(U) = \int_F \frac{|S f - U(f)|}{|S f|} \mu(df). \hspace{1cm} (2.3)$$

In the probabilistic setting, the relative error is defined as in the worst case setting, but disregarding a set of measure at most $\delta$, where $\delta \in (0, 1)$,

$$e^{\text{prob}}(U) = \inf_{A : \mu(A) \leq \delta} \sup_{f \in F \setminus A} \frac{|S f - U(f)|}{|S f|}. \hspace{1cm} (2.4)$$

We want to find a functional $U$ such that

$$e(U) \leq \varepsilon. \hspace{1cm} (2.5)$$

Here $e(\cdot)$ stands for the relative error in one of the three settings and $\varepsilon$ is a given nonnegative number. We assume that $\varepsilon < 1$, since otherwise $U = 0$ satisfies (2.5).

To find $U(f)$ for $f \in F$, we need to compute some information about $f$. We assume that we can compute values of certain continuous linear functionals at $f$. For instance, if $F$ is a class of smooth functions, then we can assume that $f(x)$ or some derivative $f^{(n)}(x)$ can be computed for $x$ belonging to the domain of $f$. In general, let $\Lambda$, where $\Lambda \subseteq F^*$, denote the class of permissible continuous linear functionals. That is, $L \in \Lambda$ iff $L(f)$ can be
computed for each $f$ from $F$. We call $L(f)$ a permissible information evaluation.

The approximation $U(f)$ is computed by combining a finite or infinite number of permissible information evaluations. That is,

$$U(f) = \phi(L_1(f), \ldots, L_n(f)),$$

where $\phi$ is a mapping into $\mathbb{R}$ and $L_i$ belongs to $\Lambda$. The mapping $\phi$ is called an algorithm. We stress that the number $n$ of evaluations as well as the functionals $L_i$, can be chosen adaptively; i.e., they can depend on the previously computed $L_1(f), \ldots, L_{i-1}(f)$. The precise definition is given in Section 3.

We want to compute $U(f)$ satisfying (2.5) with minimal cost. We assume that the cost of each evaluation $L(f)$ is equal to $c$. We also assume that one can perform certain operations such as arithmetic operations, comparisons, and taking the square roots at cost equal to one. Then $\text{cost}(U, f)$ is defined as the cost of information evaluations plus the cost of operations needed to compute $U(f)$ (see Section 3).

The cost of $U$ is defined depending on the setting. In the worst case and in the probabilistic settings, the global cost is the maximal cost needed to compute $U(f)$,

$$\text{cost}_w(U) = \text{cost}_{\text{prob}}(U) = \sup_{f \in F} \text{cost}(U, f).$$

In the average case setting, the cost is defined as the average cost of computing $U(f)$,

$$\text{cost}_{\text{avg}}(U) = \int_F \text{cost}(U, f) \mu(df).$$

We are ready to define the $\varepsilon$-complexity of computing an approximation to a linear functional under a relative error criterion. As always, complexity is the minimal cost of computing an approximation with a prescribed accuracy. That is,

$$\text{comp}(\varepsilon) = \min \{\text{cost}(U) : U \text{ such that } e(U) \leq \varepsilon\},$$

with the convention that $\min \emptyset = +\infty$. Here $\text{cost}(U)$ and $e(U)$ depend on the setting.

Observe that $\text{comp}(\varepsilon)$ is the worst case $\varepsilon$-complexity if both $e(U)$ and $\text{cost}(U)$ are defined in the worst case setting. Similarly we have the $\varepsilon$-complexity in the average and probabilistic settings. Sometimes we write
comp$^w(\varepsilon)$, comp$^{av}_m(\varepsilon)$, comp$^{prob}(\varepsilon, \delta)$ to emphasize the setting and the dependence on the parameter $\delta$ in the probabilistic setting.

We stress that the $\varepsilon$-complexity depends on the solution functional $S$, the space $F$, and the class $\Lambda$ of permissible information evaluations. In the average case and probabilistic settings, it also depends on the Gaussian measure $\mu$.

We now present the results of this paper. First of all, to make the problem interesting we assume that the solution functional $S$ does not belong to $\Lambda$. That is, $Sf$ cannot be computed. In fact, we make a stronger assumption. Namely, we assume that the solution functional $S$ cannot be represented as a finite linear combination of permissible functionals from $\Lambda$, i.e.,

$$S \notin \text{span}\{\Lambda\}. \quad (2.9)$$

For instance, if $S$ is the integration functional, $Sf = \int_0^1 f(t)dt$, dim $F = +\infty$ and $\Lambda$ consists of function evaluations, then (2.9) holds. On the other hand, if (2.9) is violated, then $Sf$ can be computed exactly with complexity proportional to the smallest $k$ for which (2.9) does not hold.

**Theorem 2.1.** Let (2.9) hold. Then the $\varepsilon$-complexity in the worst and average case settings is infinite,

$$\text{comp}^w(\varepsilon) = \text{comp}^{av}_m(\varepsilon) = +\infty \quad \forall \varepsilon < 1. \quad (2.10)$$

Thus it is impossible to approximate a linear continuous functional with relative error less than one at finite cost in the worst and average case settings.

For the probabilistic setting, the $\varepsilon$-complexity is finite under a mild assumption on the class $\Lambda$. To present bounds on the $\varepsilon$-complexity, define as in Lee and Wasilkowski (1986), $\|L\|_{\mu} = \sqrt{L(C_\mu L)}$ for any $L \in F^*$. Let

$$\sigma_n = \inf_{L \in \Lambda} \inf_{a_i \in \mathbb{R}} \|S - \sum_{i=1}^n a_i L_i\|_{\mu}^2 \quad (2.11)$$

denote the square error of the best approximation of the solution functional $S$ by linear combinations of $n$ functionals from the class $\Lambda$. Note that $\sigma_n \leq \sigma_0 = \|S\|_{\mu}^2 = S(C_\mu S)$, and $\sigma_n$ is positive for all $n$.

For given $\varepsilon$ and $\delta$, define

$$m(\varepsilon, \delta) = \min\left\{n : \sqrt{\frac{\sigma_n}{S(C_\mu S) - \sigma_n}} \leq \varepsilon \tan\left(\frac{\pi}{2} \frac{\delta}{\varepsilon}\right)\right\}. \quad (2.12)$$

Observe that $m(\varepsilon, \delta)$ is finite for all $\varepsilon, \delta \in (0,1)$ iff $\sigma_n \to 0$. 

...
We shall prove that we need to compute roughly \( m(\varepsilon, \delta) \) evaluations in order to guarantee (2.5). That's why \( m(\varepsilon, \delta) \) is called the \((\varepsilon, \delta)\)-cardinality number. More precisely we have

**Theorem 2.2.** The \((\varepsilon, \delta)\)-complexity in the probabilistic setting is bounded by

\[
c \cdot m(\varepsilon, \delta) \leq \text{comp}^\text{prob}(\varepsilon, \delta) \leq (c + 2) \cdot m(\varepsilon, \delta) + 5,
\]

where \( c \) is the cost of one information evaluation and

\[
\delta_1 = \min \left\{ 1, \frac{\delta}{1 - \sqrt{(\pi \varepsilon/4) \ln ((1 + \varepsilon)/(1 - \varepsilon))}} \right\}.
\]

Usually, the cost \( c \) is much higher than the cost of one arithmetic operation or comparison, \( c \gg 1 \). Then the upper bound in (2.13) is close to \( c \cdot m(\varepsilon, \delta) \). For small \( \varepsilon \) we have \( \delta_1 = \delta \) and then the bounds in (2.13) are tight.

Theorem 2.2 implies that the \((\varepsilon, \delta)\)-complexity is finite for all \( \varepsilon, \delta \in (0,1) \) iff \( \sigma_n \to 0 \). That is, the \((\varepsilon, \delta)\)-complexity is finite iff the solution functional \( S \) can be approximated in the norm \( \| \cdot \|_\mu \) by functionals from \( \Lambda \) with arbitrary precision.

We now construct a functional \( U^* \) for which \( e^\text{prob}(U^*) \leq \varepsilon \) and whose cost is bounded by the upper bound of (2.13). Let \( n = m(\varepsilon, \delta) \) be finite. For simplicity, assume that the first infimum in (2.11) is attained. Choose \( L_1^*, L_2^*, \ldots, L_n^* \) from \( \Lambda \) for which this holds. Let \( K_i = \sum_{j=1}^{n} a_{i,j} L_j^* \), where numbers \( a_{i,j} \) are chosen such that \( K_i(C \mu K_j) = \delta_{i,j} \) for \( i, j = 1, 2, \ldots, n \). The numbers \( a_{i,j} \) can be computed by, for instance, Gram-Schmidt orthogonalization. Define the numbers

\[
q_i = \sum_{j=1}^{n} a_{i,j} S(C \mu K_j), \quad i = 1, 2, \ldots, n
\]

\[
q_0 = \left( \frac{2}{\varepsilon} \ln \left( \frac{1 + \varepsilon}{1 - \varepsilon} \right) \right) \sigma_n.
\]

Here \( \sigma_n = S(C \mu S) - \sum_{i=1}^{n} [S(C \mu K_i)]^2 \).

The functional \( U^* \) is then defined as follows. Compute:

(i) \( L_1^*(f), L_2^*(f), \ldots, L_n^*(f) \)

(ii) \( a := \sum_{i=1}^{n} L_i^*(f) q_i \)

(iii) \( U^*(f) := (a + \text{sign}(a) \sqrt{a^2 + q_0}) \frac{1 - \varepsilon^2}{2} \).

Here \( \text{sign}(a) = 1 \) for \( a \geq 0 \) and \( \text{sign}(a) = -1 \) for \( a < 0 \).
We shall prove in Section 7 that $e^\mu_{\text{prob}}(U*) \leq \varepsilon$. The cost of $U^*$ is at most $(c + 2)n + 5$ since $q_0, \ldots, q_n$ as well as $(1 - \varepsilon^2)/2$ can be precomputed.

Observe that $U^*$ is a nonlinear functional. Nevertheless, for small $\varepsilon$ the coefficient $q_0$ is close to zero and $U^*$ is almost linear,

$$U^*(f) \approx U_n(f) = \sum_{i=1}^n L_i^*(f)q_i. \quad (2.16)$$

The functional $U^*$ is linear and $U^*_n(f)$ is equal to the mean of the conditional measure of the values $S(f)$ after $L^*_1(f), \ldots, L^*_n(f)$ have been computed. The functional $U^*$ corresponds to the $\mu$-spline algorithm (see Lee and Wasilkowski, 1986).

We now specialize the results of the probabilistic setting to the integration functional, $Sf = \int_0^1 f(x) \, dx$, where $f$ belongs to the class $F = C^k(0, 1)$ of $k$-times continuously differentiable functions such that $f(0) = f'(0) = \cdots = f^{(k)}(0) = 0$. Assume that $\mu$ is the Wiener measure placed on $k$th derivatives and $A$ is the class of function evaluations.

It is known (see Lee and Wasilkowski, 1986) that $\sigma_n = \Theta(n^{-2(k+1)})$. This is achieved for function evaluations at equally placed points, i.e., not much is lost by taking $L^*_n(f) = f(i/n)$. Furthermore, $U^*_n(f)$ defined by (2.16) is now the integral of the interpolating natural spline of degree $2k + 1$. From this we conclude that the probabilistic $(\varepsilon, \delta)$-complexity is given by

$$\text{comp}_{\text{prob}}(\varepsilon, \delta) = \Theta(c(\varepsilon \delta)^{-1/(k+1)})$$

and the algorithm based on the integral of the natural spline of degree $2k + 1$ which interpolates the function at $n = \Theta((\varepsilon \delta)^{-1/(k+1)})$ equally spaced points solves the problem with almost minimal cost.

3. Information

In this section we precisely define information used to compute an approximation $U(f)$ and the cost of computing $U(f)$. Let

$$N(f) = \{L_1(f), L_1(f; y_1), \ldots, L_{m.f}(f; y_1, \ldots, y_{m(f)-1})\}, \quad (3.1)$$

where $y_1 = L_1(f)$ and $y_i = L_i(f; y_1, \ldots, y_{i-1})$, for $i = 2, 3, \ldots, n(f)$. Thus, $y_i$ denotes the $i$th value of information evaluation. Here we assume that for fixed $y_1, y_2, \ldots, y_i$ the functional $L_{m.f}(\cdot; y_1, \ldots, y_{i-1})$ belongs to $A$. The number $n(f)$ denotes the total number of information evaluations computed for the element $f$. It is called cardinality of $N$ at $f$ and is
determined as follows. Let \( \text{ter}_i : \mathcal{Y}^i \to \{0, 1\} \), called a termination function, be a given Boolean function. Knowing \( y_1, \ldots, y_i \), we compute \( \text{ter}_i(y_1, \ldots, y_i) \). If it is 1 we terminate the information evaluations and set \( n(f) = i \). If not, we select \( L_{i+1}(\cdot; y_1, \ldots, y_i) \), compute \( y_{i+1} = L_{i+1}(f; y_1, \ldots, y_i) \) and then the process is repeated. The number \( n(f) \) is defined as

\[
n(f) = \min\{i : \text{ter}_i(y_1, \ldots, y_i) = 1\}
\]  

(3.2)

(see Wasilkowski, 1986).

The essence of (3.1) and (3.2) is that the choice of the \( i \)th information evaluation, as well as the total number of them, adaptively depends on the previously computed values and it may vary with \( f \). The information \( N \) is therefore called adaptive. If \( n(f) = n \) and \( L_i(\cdot; y_1, \ldots, y_{i-1}) = L_i \) then \( N \) is called nonadaptive.

We compute \( U(f) \) by combining the information \( N(f) \). That is, \( U(f) = \phi(N(f)) \), where \( \phi \), called an algorithm, is an arbitrary mapping \( \phi : N(F) \to \mathcal{Y} \). The cost of computing \( U(f) \) consists of two terms. First we compute \( N(f) \) with cost \( (N, f) \). Thus cost \( (N, f) \) is equal to \( c \cdot n(f) \) plus the cost needed to select the functionals \( L_i \) and to compute the termination functions \( \text{ter}_i \). Knowing \( y = N(f) \) we compute \( \phi(y) \). The cost of \( \phi \) is equal to cost(\( \phi, y \)) which is the sum of the cost of the operations needed to compute \( \phi(y) \). Then cost(\( U, f \)) is defined as

\[
\text{cost}(U, f) = \text{cost}(N, f) + \text{cost}(\phi, N(f)).
\]  

(3.3)

The cost of \( U \), cost(\( U \)), is defined by (2.6) or (2.7) depending on the setting.

4. Worst Case Setting

In this section we prove that Theorem 2.1 in the worst case setting. Let \( N \) be adaptive information given by (3.1) such that \( \sup_{f \in F} n(f) < \infty \). Note that (2.9) implies that

\[
\ker N \not\subset \ker S.
\]  

(4.1)

Here, \( \ker N = \{ f \in F : N(f) = 0 \} \) and it is obviously a linear subspace even for adaptive \( N \). For \( U(f) = \phi(N(f)) \) we have

\[
e_w(U) = \sup_{f \in F} \frac{|Sf - U(f)|}{|Sf|} \geq \sup_{f \in \ker N} |Sf| = \sup_{f \in \ker N} \frac{|Sf - \phi(0)|}{|Sf|}.
\]  

(4.2)
Take \( f \in \text{ker} \, N \) and \( f \notin \text{ker} \, S \). Since \( -f \) also belongs to \( \text{ker} \, N \) we have

\[
e^w(U) \geq \max \left\{ \frac{|Sf - \phi(0)|}{|Sf|}, \frac{|S(-f) - \phi(0)|}{|S(-f)|} \right\}
\]

\[
\geq \frac{1}{2} \cdot \frac{|Sf - \phi(0)| + |S(f) + \phi(0)|}{|Sf|} \geq 1. \tag{4.3}
\]

Hence, \( e^w(U) \leq \varepsilon < 1 \) implies that the cardinalities \( n(f) \) are unbounded. Therefore

\[
\text{cost}(U) \geq c \cdot \sup_{f \in F} n(f) = +\infty. \tag{4.4}
\]

This completes the proof. Compare also with Traub and Woźniakowski (1980, p. 196), where the same result is proven for nonadaptive information \( N \) and for an arbitrary linear operator \( S \).

5. Properties of Gaussian Measures

In order to analyze the average case and probabilistic settings we need to recall some basic properties of Gaussian measures which can be found in Lee and Wasilkowski (1986).

As in the Section 2, let \( \mu \) be the Gaussian measure on a separable Banach space \( F \) with mean zero and a one-to-one correlation operator \( C_\mu \), where \( C_\mu : F^* \to F \). Let \( N(f) = [L(f), L_2(f; y_1), \ldots, L_n(f)] \) be adaptive information of the form (3.1). By taking suitable linear combinations we can assume that for any fixed \( y = [y_1, y_2, \ldots] \)

\[
L_{i,y}(\cdot) = L_i(\cdot; y_1, y_2, \ldots, y_i).
\]

Let \( A_k = N(\{ f : n(f) = k \}) \) be the set of vectors \( y \) for which \( k \) evaluations are performed. If \( k < +\infty \) then \( \lambda|_{A_k} \) can be extended to a Gaussian measure on \( \mathbb{R}^k \) with mean 0 and correlation \( I \), the identity matrix. For \( \lambda \)-almost all \( y = [y_1, \ldots, y_k] \in N(F) \), \( \pi(\cdot|y) \) is a Gaussian measure with mean
COMPLEXITY OF APPROXIMATION

\[ m(y) = \sum_{i=1}^{k} y_i C_\mu L_{i,y} \]  \hspace{1cm} (5.3)

and correlation operator

\[ C_y = C_\mu - \sum_{i=1}^{k} L_{i,y}(C_\mu(\cdot))C_\mu L_{i,y}. \]  \hspace{1cm} (5.4)

We shall need one more well-known property of Gaussian measure. Let \( y \) be a Gaussian measure on a separable Banach space \( F \) with mean \( m_y \) and correlation operator \( C_y \), where \( C_y : F^* \to F \). Then for every \( L \in F^* \), the induced measure \( \gamma \circ L^{-1} \) is Gaussian with mean \( L(m_y) \) and variance \( \sigma = L(C_y L) \). If \( \sigma > 0 \) then for any Borel set \( B \subset \mathcal{R} \) we have

\[ \gamma \circ L^{-1}(B) = \frac{1}{\sqrt{2\pi\sigma}} \int_B e^{-\frac{1}{2}(L(m_y) - t)^2 \sigma} \, dt. \]  \hspace{1cm} (5.5)

Otherwise, if \( \sigma = 0 \) then

\[ \gamma \circ L^{-1}(B) = \begin{cases} 1 & L(m_y) \in B, \\ 0 & \text{otherwise}. \end{cases} \]  \hspace{1cm} (5.6)

In particular, for the solution functional \( S \), we find that

\[ \nu(\cdot|y) = \pi(S^{-1} \cdot|y) \]  \hspace{1cm} (5.7)

is a Gaussian measure on \( \mathcal{R} \). Let

\[ w(y) = \{ SC_\mu L_1, SC_\mu L_{2,y}, \ldots, SC_\mu L_{k,y} \}. \]  \hspace{1cm} (5.8)

Then the mean \( a(y) \) of the measure \( \nu(\cdot|y) \) is given by

\[ a(y) = S(m(y)) = \langle y, w(y) \rangle \]  \hspace{1cm} (5.9)

and the variance \( \sigma(y) \) of the measure \( \nu(\cdot|y) \) is given by

\[ \sigma(y) = S(C_y S) = S(C_\mu S) - \langle w(y), w(y) \rangle. \]  \hspace{1cm} (5.10)

Using the definition (2.11) of \( \| \cdot \|_\mu \), (5.10) and (5.1) yield

\[ \sigma(y) = \min_{a_i} \| S - \sum_{i=1}^{k} a_i L_{i,y} \|_\mu^2. \]  \hspace{1cm} (5.11)

Then (2.9) implies \( \sigma(y) > 0 \).
As an application of (5.5), consider a measurable function \( G : \mathbb{R} \to \mathbb{R} \) and a vector \( w \in \mathbb{R}^n \). Then

\[
(2\pi)^{-n/2} \int_{\mathbb{R}^n} G(y, w) e^{-(y, y)/2} dy = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} G(x \cdot \|w\|) e^{-x^2/2} dx. \tag{5.12}
\]

6. Average Case

In this section we prove Theorem 2.1 in the average case setting. We use notation and properties of Gaussian measures from Section 5. Without loss of generality we can assume that \( \mu(\{f : n(f) = +\infty\}) = 0 \) since otherwise the average complexity is infinite. From (2.3) and (5.2) we have

\[
e^{av}(U) = \int_{N(F)} \int_{N^{-1}(y)} \frac{|Sf - U(f)|}{|Sf|} \pi(df|y) \lambda(dy). \tag{6.1}
\]

Then for \( \lambda \)-almost all \( y = [y_1, \ldots, y_k] \in N(F) \), (5.5) with \( a = a(y) \) and \( \sigma = \sigma(y) \), \( \sigma > 0 \), yields

\[
\int_{N^{-1}(y)} \frac{|Sf - \phi(y)|}{|Sf|} \pi(df|y) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^{+\infty} \frac{|x - \phi(y)|}{|x|} e^{-(x-a)^2/2\sigma} dx.
\]

Due to (6.1) this yields

\[
e^{av}(U) = \begin{cases} 1 & \text{if } U(f) = 0 \text{ for } \mu\text{-almost all } f, \\ +\infty & \text{otherwise.} \end{cases} \tag{6.3}
\]

Therefore if \( \varepsilon < 1 \) we cannot solve our problem with finite complexity.

7. Probabilistic Setting

We need a number of lemmas to prove Theorem 2.2. Let \( N \) be adaptive information of the form (3.1) with fixed cardinality \( n(f) = n \) and normalized as in (5.1). We define an algorithm \( \phi^* \) using such information \( N \). Let \( y = N(f) \in \mathbb{R}^n \). From (5.9) and (5.10) we know that \( a = a(y) \) and \( \sigma = \sigma(y) \) are the mean and variance of the measure \( \nu(\cdot|y) \), respectively. Let
\[ \sigma' = \sigma'(y) = \left( \frac{1}{2\varepsilon} \ln \frac{1 + \varepsilon}{1 - \varepsilon} \right) \sigma(y). \] (7.1)

Note that for small \( \varepsilon \), \( \sigma' \approx \sigma \). Define

\[ \phi^*(y) = (a + \text{sign}(a)\sqrt{a^2 + 4\sigma'}) \frac{1 - \varepsilon^2}{2}. \] (7.2)

For small \( \varepsilon \) we have \( \phi^*(y) \approx a \), i.e., \( \phi^*(y) \) is close to the mean of the measure \( \nu(|y|) \).

We now prove an optimality property of the algorithm \( \phi^* \). To do this, define

\[ M(\phi, N) = \mu \left( \left\{ f \in F : \frac{|Sf - \phi(N(f))|}{|Sf|} > \varepsilon \right\} \right) \] (7.3)

for any algorithm \( \phi \) using \( N \). Thus, \( M(\phi, N) \) is the measure of the set for which the relative error is greater than \( \varepsilon \). Observe that \( M(\phi, N) \leq \delta \) iff the probabilistic error \( e_{\text{prob}}(U) \leq \varepsilon \) for the functional \( U = \phi \circ N \). Note that

\[ M(\phi, N) = \int_{\mathbb{R}^n} M(\phi, N, y)\lambda(dy), \] (7.4)

where

\[ M(\phi, N, y) = \mu \left( \left\{ f \in F : \frac{|Sf - \phi(N(f))|}{|Sf|} > \varepsilon \right\} \right) \]

\[ = \nu \left( \left\{ x \in \mathbb{R} : \frac{|x - \phi(y)|}{|x|} > \varepsilon \right\} \right). \] (7.5)

\textbf{Lemma 7.1.} The algorithm \( \phi^* \) minimizes \( M(\phi, N, y) \) for all \( y \in \mathbb{R}^n \), i.e.,

\[ M(\phi^*, N, y) = \inf_{\phi} M(\phi, N, y), \quad \forall y \in \mathbb{R}^n. \] (7.6)

\textbf{Proof.} From (7.5) it follows that we need to find \( u = \phi(y) \) minimizing \( \nu(\{x \in \mathbb{R} : |x - u|/|x| > \varepsilon\}) \). Observe that

\[ \frac{|x - u|}{|x|} > \varepsilon \Leftrightarrow x < u_1 = \frac{u - |u|\varepsilon}{1 - \varepsilon^2} \quad \text{or} \quad x > u_2 = \frac{u + |u|\varepsilon}{1 - \varepsilon^2}. \] (7.7)

Hence
\[ M(\phi, N, y) = 1 - \frac{1}{\sqrt{2\pi\sigma}} \int_{-\infty}^{\infty} e^{-(x-y)^2/2\sigma} \, dx \overset{\text{def}}{=} 1 - \frac{1}{\sqrt{2\pi\sigma}} \cdot H(u). \quad (7.8) \]

We seek \( u = u^* \) which maximizes \( H(u) \). Since \( H \) is differentiable for \( u \neq 0 \) and \( u = 0 \) does not maximize \( H \), we consider the equation \( H'(u) = 0 \) and find that

\[ u^* = \phi^*(y) = (a + \text{sign}(a) \sqrt{a^2 + 4\sigma'}) \frac{1 - \varepsilon^2}{2}. \quad (7.9) \]

This completes the proof. \( \blacksquare \)

We now estimate \( M(\phi^*, N) \) for nonadaptive information.

**Lemma 7.2.** For nonadaptive information \( N = [L_1, L_2, \ldots, L_n] \), \( L_i(C \mu L_j) = \delta_{ij} \), we have

\[ \left( 1 - \frac{\pi}{4} \varepsilon \ln \left( \frac{1 + \varepsilon}{1 - \varepsilon} \right) \frac{1}{\pi} \arctan \left( \frac{\varepsilon}{\sqrt{\sigma}} \| w \| \right) \right)^2 \]

\[ \leq M(\phi^*, N) \leq \frac{2}{\pi} \arctan \left( \frac{\varepsilon}{\sqrt{\sigma}} \| w \| \right), \quad (7.10) \]

where \( w = [SC \mu L_1, \ldots, SC \mu L_n] \). For small \( \varepsilon \), we have

\[ M(\phi^*, N) = \frac{2}{\pi} \arctan \left( \frac{\varepsilon}{\sqrt{\sigma}} \| w \| \right). \quad (7.11) \]

**Proof.** Let \( \gamma_a = u^*/(1 - \varepsilon^2) - a \) with \( a = \langle y, w \rangle \). From (7.8) and (7.9) we have

\[
M(\phi^*, N, y) = 1 - \frac{1}{\sqrt{2\pi}} \cdot \text{sign}(a) \cdot \left[ \int_{-ae/\sqrt{\sigma}}^{ae/\sqrt{\sigma}} e^{-t^2/2} \, dt \right. \\
- \int_{-ae/\sqrt{\sigma}}^{-ae/\sqrt{\sigma} + (ae/\sqrt{\sigma})(1-\varepsilon)} e^{-t^2/2} \, dt + \int_{ae/\sqrt{\sigma}}^{ae/\sqrt{\sigma} + (ae/\sqrt{\sigma})(1+\varepsilon)} e^{-t^2/2} \, dt \\
\left. + \int_{ae/\sqrt{\sigma} + (ae/\sqrt{\sigma})(1-\varepsilon)}^{ae/\sqrt{\sigma} + (ae/\sqrt{\sigma})(1+\varepsilon)} e^{-t^2/2} \, dt \right] \\
= 1 - \frac{1}{\sqrt{2\pi}} \cdot \text{sign}(a) \cdot \left[ \int_{-ae/\sqrt{\sigma}}^{ae/\sqrt{\sigma}} e^{-t^2/2} \, dt \\
- \int_{ae/\sqrt{\sigma} + (ae/\sqrt{\sigma})(1-\varepsilon)}^{ae/\sqrt{\sigma} + (ae/\sqrt{\sigma})(1+\varepsilon)} e^{-t^2/2} \, dt \\
+ \int_{ae/\sqrt{\sigma} + (ae/\sqrt{\sigma})(1-\varepsilon)}^{ae/\sqrt{\sigma} + (ae/\sqrt{\sigma})(1+\varepsilon)} e^{-t^2/2} \, dt \right]. \quad (7.12) 
\]
Note that the absolute value of the second integral is no less than the absolute value of the third one. The absolute value of the fourth integral can be estimated from above by shifting the interval by \((\gamma_0/\sqrt{\sigma})(1 - \varepsilon)\). This yields

\[
M(\phi^*, N, y) \geq 1 - \frac{1}{\sqrt{2\sigma}} \cdot \sign(a) \cdot \left[ \int_{-\infty}^{\infty} e^{-t^2/2} dt + \int_{a\in\sqrt{\sigma}}^{a\in\sqrt{\sigma}+2\varepsilon(a\in\sqrt{\sigma})} e^{-t^2/2} dt \right].
\]  

(7.13)

Because \(|\gamma_0| \leq \sqrt{\sigma'}\), we have

\[
\left| \int_{a\in\sqrt{\sigma}}^{a\in\sqrt{\sigma}+2\varepsilon(a\in\sqrt{\sigma})} e^{-t^2/2} dt \right| \leq 2\varepsilon \sqrt{\frac{\sigma'}{\sigma}} \cdot e^{-a\in\sqrt{\sigma}t^2/2}.
\]  

(7.14)

We now integrate (7.12) with respect to \(y\). Using (5.12), (7.4) and combining (7.13) with (7.14) we obtain

\[
M(\phi^*, N) \geq \frac{2}{\pi} \int_0^\infty \int_{x\in\sqrt{\sigma}||w||} e^{-t^2/2} dt e^{-x^2/2} dx
\]

\[
- e \int_0^\infty \int_{x\in\sqrt{\sigma}||w||} e^{-t^2/2} dt e^{-x^2/2} dx
\]

\[
\left[ \int_{x\in\sqrt{\sigma}||w||} e^{-t^2/2} dx \right.
\]

\[
\cdot \frac{1}{\sqrt{(x^2/\sqrt{\sigma})||w||}^2 + 1}.
\]  

(7.15)

Let \(\tau = (\varepsilon/\sqrt{\sigma})||w||\). Define

\[
f(\tau) = \frac{2}{\pi} \int_0^\infty \left( \int_{x\in\sqrt{\sigma}||w||} e^{-t^2/2} dt \right) e^{-x^2/2} dx.
\]

Obviously \(f(0) = 1\) and \(f'(\tau) = (2/\pi)(1/(1 + \tau^2)).\). This yields \(f(\tau) = 1 - (2/\pi) \arctg(\tau) = (2/\pi) \arctg(\tau).\). Thus

\[
\frac{2}{\pi} \int_0^\infty \int_{x\in(\varepsilon/\sqrt{\sigma})||w||} e^{-t^2/2} dt e^{-x^2/2} dx = \frac{2}{\pi} \arctg \left( \varepsilon \sqrt{\sigma}||w|| \right).
\]  

(7.16)

Since \(\arctg(\tau) = \arcsin(1/\sqrt{1 + \tau^2}) \geq 1/\sqrt{1 + \tau^2}\), (7.14) yields
This proves the left side of (7.10).

We now find an upper bound on $M(\phi^*, N)$. Define the algorithm $\phi(y) = a(1 - c^2)$. Obviously $M(\phi^*, N) \leq M(\phi, N)$. Due to (7.8) we have

$$M(\phi, N, y) = 1 - \frac{1}{\sqrt{2\pi}} \cdot \text{sign}(a) \int_{-ae/\sqrt{\sigma}}^{ae/\sqrt{\sigma}} e^{-t^2/2} dt. \quad (7.18)$$

As in (7.15) and (7.16) we find out that $M(\phi^*, N) = (2/\pi) \arccot \left( \frac{c}{\sqrt{\sigma}} \right)$. Thus, the right side of (7.10) is also proved. Since (7.11) easily follows from (7.10), the proof is complete. \[ \square \]

We now show that adaptive information is not more powerful than nonadaptive information. We prove the following theorem.

**Theorem 7.1.** For every adaptive information $N^a$ of the form (3.1) with fixed cardinality $n(f) \equiv n$, there exists nonadaptive information $N^*$ of cardinality $n$ such that

$$M(\phi^*, N^*) \leq M(\phi^a, N^a), \quad (7.19)$$

where $\phi^*, \phi^a$ are algorithms defined by (7.2) for $N^*, N^a$, respectively; i.e., $\phi^*$ and $\phi^a$ minimize the functions $M(\cdot, N^*)$ and $M(\cdot, N^a)$ among all algorithms using information $N^*$ and $N^a$, respectively.

To prove this theorem we need two lemmas.

**Lemma 7.3.** Let $w: \mathbb{R}^n \to \mathbb{R}^n$ be a measurable mapping such that $\|w(y)\| = 1$ and

$$w(y) = [w_1, w_2(y_1), \ldots, w_n(y_1, \ldots, y_{n-1})]. \quad (7.20)$$

Let $h, G: \mathbb{R} \to \mathbb{R}$ be measurable functions. Then for every $z \in \mathbb{R}$ such that $\|z\| = 1$ we have

$$\int_{y^n} G((y, w(y)))h((y, y)) \, dy = \int_{y^n} G((y, z))h((y, y)) \, dy. \quad (7.21)$$

**Proof.** For every $y = [y_1, \ldots, y_n] \in \mathbb{R}^n$, we define $n \times n$ orthogonal matrices $D_1, D_2, \ldots, D_n$ such that
The matrix $D_i$ is of the form

$$D_i = \begin{bmatrix} I_i \\ \overline{D}_i \end{bmatrix},$$

where $I_i$ is the $(i - 1) \times (i - 1)$ identity matrix and $\overline{D}_i$ is an $(n - i + 1) \times (n - i + 1)$ orthogonal matrix depending only on $y_1, \ldots, y_{i-1}$. To construct $D_i$, take $\overline{D}_i = \overline{D}_1$ as an orthogonal matrix for which the first component of $D_1(z)$ is $w_1$. Assume inductively that $D_{i-1} \circ \cdots \circ D_1(z)$ agrees with the first $(i - 1)$ components of $w(y)$. Thus $D_{i-1} \circ \cdots \circ D_1(z) = \begin{bmatrix} w_1, w_2(y_1), \ldots, \\
 w_{i-1}(y_1, \ldots, y_{i-2}), v \end{bmatrix}^T$. Define $\overline{D}_i$ such that the first component of $\overline{D}_i(v)$ is $w_i(y_1, \ldots, y_{i-1})$. Clearly $\overline{D}_i$ depends only on $y_1, \ldots, y_{i-1}$ and $D_i \circ \cdots \circ D_1(z)$ agrees with $w(y)$ on the first $i$ components.

For $i \in [1, n]$ we have

$$\int_{\mathbb{R}^{n-i+1}} G(\langle y, D_i \circ \cdots \circ D_1(z) \rangle) \, dy_n \cdot dy_{n-1} \cdot \ldots \cdot dy_i$$

$$= \int_{\mathbb{R}^{n-i+1}} G(\langle D_i^T y, D_{i-1} \circ \cdots \circ D_1(z) \rangle) \, dy_n \cdot dy_{n-1} \cdot \ldots \cdot dy_i$$

$$= \int_{\mathbb{R}^{n-i+1}} G(\langle y, D_{i-1} \circ \cdots \circ D_1(z) \rangle) \, dy_n \cdot dy_{n-1} \cdot \ldots \cdot dy_i.$$

The last equality follows from the change of variables $y' = D_i^T y$ and noting that $y'$ agrees with $y$ on the first $(i - 1)$ components and $D_i^T$ depends only on $y_1, \ldots, y_{i-1}$.

Using (7.24) for $i = n, n-1, \ldots, 1$ we get

$$\int_{\mathbb{R}^n} G(\langle y, w(y) \rangle) h(\langle y, y \rangle) \, dy = \int_{\mathbb{R}^n} G(\langle y, D_n \circ \cdots \circ D_1(z) \rangle) h(\langle y, y \rangle) \, dy$$

$$= \int_{\mathbb{R}^{n+1}} \left[ \int_{\mathbb{R}^n} G(\langle D_n^T y, D_{n-1} \circ \cdots \circ D_1(z) \rangle) h(\langle y, y \rangle) \, dy_n \right] \, dy_{n-1} \cdot \ldots \cdot dy_1$$

$$= \int_{\mathbb{R}^{n+1}} \left[ \int_{\mathbb{R}^n} G(\langle y, D_{n-1} \circ \cdots \circ D_1(z) \rangle) h(\langle y, y \rangle) \, dy_n \right] \, dy_{n-1} \cdot \ldots \cdot dy_1$$

$$= \ldots =$$

$$= \int_{\mathbb{R}^n} G(\langle y, z \rangle) h(\langle y, y \rangle) \, dy.$$

This completes the proof. □

**Lemma 7.4.** Let $w: \mathbb{R}^n \to \mathbb{R}^n$ be a measurable mapping of the form (7.20) such that
Let $h : \mathbb{R} \to \mathbb{R}$ be a nonnegative measurable function. Let $G, G : \mathbb{R} \to \mathbb{R}$ be a measurable function such that

$$|x_1| \leq |x_2| \Rightarrow G(x_1) \geq G(x_2) \quad \forall x_1, x_2 \in \mathbb{R}. \quad (7.27)$$

Then for every $z \in \mathbb{R}^n$ such that $\|z\| \geq \sup_{y \in \mathbb{R}^n} \|w(y)\|$ we have

$$\int_{\mathbb{R}^n} G((y, w(y)))h((y, y)) \, dy \geq \int_{\mathbb{R}^n} G((y, z))h((y, y)) \, dy. \quad (7.28)$$

**Proof.** Without loss of generality we can assume that for all $y \in \mathbb{R}$, $w(y) \neq 0$. Let $r = \sup_{y \in \mathbb{R}^n} \|w(y)\|$. Then using (7.27) and Lemma 7.3 we obtain

$$\int_{\mathbb{R}^n} G((y, w(y)))h((y, y)) \, dy = \int_{\mathbb{R}^n} G\left(\|w(y)\| \left\langle y, \frac{w(y)}{\|w(y)\|} \right\rangle \right) h((y, y)) \, dy$$

$$\geq \int_{\mathbb{R}^n} G\left(r, \frac{w(y)}{\|w(y)\|} \right) h((y, y)) \, dy$$

$$= \int_{\mathbb{R}^n} G\left(r, \frac{z}{\|z\|} \right) h((y, y)) \, dy$$

$$\geq \int_{\mathbb{R}^n} G(\|z\| \left\langle y, \frac{z}{\|z\|} \right\rangle) h((y, y)) \, dy$$

$$= \int_{\mathbb{R}^n} G((y, z))h((y, y)) \, dy. \quad (7.29)$$

The proof is complete. ■

We are now ready for the

**Proof of Theorem 7.1.** For information $N^a$ consider $w(y)$, $a(y)$, $\sigma(y)$ defined by (5.8), (5.9), and (5.10), respectively. Without loss of generality we can assume that there exists $y^*$ such that for $w^* = w(y^*)$,

$$\sup_{y \in \mathbb{R}^n} \|w(y)\| = \|w^*\|. \quad (7.30)$$

This is equivalent to $\inf_{y \in \mathbb{R}^n} \|\sigma(y)\| = \sigma(y^*) = \sigma^*$. Consider $M(\phi^a, N^a, y)$ with $\phi^a$ defined by (7.2). It is easy, although tedious, to show that $M(\phi^a, N^a, y)$ is decreasing as a function of $\sigma(y)$. Thus,

$$M(\phi^a, N^a, y) \geq G(a) \overset{\text{df}}{=} 1 - \frac{1}{\sqrt{2\pi\sigma^*}} \int_{-\infty}^{\infty} e^{-(x-a)^2/2\sigma^*} \, dx, \quad (7.31)$$
where $u^*_1, u^*_2$ are defined by (7.7) with $u = (a + \text{sign}(a) \sqrt{a^2 + 4\sigma^2})(1 - \varepsilon^2)/2$ and $\sigma'$ is defined by (7.1) with $\sigma(y) = \sigma^*$. It is easy to show that $G$ satisfies (7.27) of Lemma 7.4.

Define nonadaptive information $N^* = [L_1, L_2(:, y^*_1), \ldots, L_n(:, y^*_n), \ldots, y^*_{n-1}]$ and let $\mathbf{a}^*$ be defined by (7.2) for information $N^*$. Then using (7.31) and Lemma 7.4 we have

$$M(a^*, N^*) \geq (2\pi)^{-n/2} \int_{W^*} G((y, w(y))) e^{-(y, v)^2/2} \, dy$$

(7.32)

The completes the proof of Theorem 7.1.

**Proof of Theorem 2.2.** We first prove the upper bound of (2.13). Let $n = m(\varepsilon, \delta)$. Consider the nonadaptive information $N^*_n = [L^*_1, L^*_2, \ldots, L^*_n]$ and $U^*$ defined by (2.15). It is enough to prove that $e^{\text{prob}(U^*)} \leq \varepsilon$.

Let $N = \{K_1, K_2, \ldots, K_n\}$, where $K_i \in \text{span}(L^*_1, \ldots, L^*_n)$ and $K_i(K_j) = \delta_{i,j}$. Then $U^*$ coincides with (7.9), $U^*(f) = \phi^*(N(f))$ and $\sigma = \sigma_n$. Observe that $e^{\text{prob}(U^*)} \leq \varepsilon$ iff $M(a^*, N) \leq \delta$. From Lemma 7.2 we get

$$M(a^*, N) \leq \frac{2}{\pi} \text{arc ctg} \left( \frac{\varepsilon}{\sqrt{\sigma}} \right),$$

(7.33)

where $\|w\|^2 = \sum_{i=1}^{n} [S(C_i, K_i)]^2$.

From (2.11) we have

$$\sigma_n = \inf_{a \in R} \|S - \sum_{i=1}^{n} a_i K_i\|^2 = S(C\mu S) - \|w\|^2.$$  

(7.34)

This and (2.12) yield

$$\frac{\varepsilon}{\sqrt{\sigma}} \|w\| = \varepsilon \sqrt{S(C\mu S) - \sigma_n} \geq \text{ctg} \left( \frac{\pi}{2} \delta \right).$$

(7.35)

Therefore $(2/\pi)\text{arc ctg}(\varepsilon / \sqrt{\sigma} \|w\|) \leq \delta$, which completes the proof of the upper bound.

To prove the lower bound of (2.13), assume first that $\delta/(1 - \sqrt{\pi/4}) \varepsilon (\ln ((1 + \varepsilon)/(1 - \varepsilon))) \varepsilon \leq 1$. Then $\delta_1 = 1$ and $m(\varepsilon, 1) = 0$. Thus, (2.13) trivially holds. Assume then that $1 - \sqrt{\pi/4} \varepsilon (\ln ((1 + \varepsilon)/(1 - \varepsilon))) > 0$ and $\delta_1 = \delta/(1 - \sqrt{\pi/4}) \varepsilon (\ln ((1 + \varepsilon)/(1 - \varepsilon))) < 1$. Take an arbitrary $U$, where $U(f) = \phi(N(f))$, such that $e^{\text{prob}(U)} \leq \varepsilon$. Thus $M(\phi, N) \leq \delta$. Since the cost of $U$ is defined by the worst case (see (2.6)) we can assume that information $N$ is of fixed and finite cardinality for any $f$, i.e., $n(f) = k$. 


From Theorem 7.1 we can assume that $N$ is nonadaptive. From Lemmas 7.1 and 7.2 we conclude that
\[
\delta \geq M(\phi, N) \geq \left(\frac{\pi}{2} \ln \left(\frac{1 + \epsilon}{1 - \epsilon}\right)\right) \cdot \frac{2}{\pi} \arctan \left(\epsilon \sqrt{\frac{S(C_\mu S) - \sigma}{\sigma}}\right). \quad (7.36)
\]

where $\sigma \geq \sigma_k$. From this we get
\[
\sqrt{\frac{\sigma_k}{S(C_\mu S) - \sigma_k}} \leq \epsilon \cdot \frac{\pi}{2} \delta_1. \quad (7.37)
\]

Hence $k \geq m(\epsilon, \delta_1)$ and $\text{cost}(U) \geq c \cdot m(\epsilon, \delta_1)$. This completes the proof. $\blacksquare$

8. Final Remarks

In this section we briefly discuss modified definitions of relative error as well as extensions for approximations of linear operators.

(i) We analyzed the relative error defined for two numbers $Sf$ and $U(f)$ as $|Sf - U(f)|/|Sf|$. One can also consider a relative error for which the roles of $Sf$ and $U(f)$ are interchanged. That is, the relative error is now given as

\[
\frac{|Sf - U(f)|}{|U(f)|}.
\]

We briefly discuss the complexity with this definition of relative error in three settings.

In the worst case setting, the complexity for $\epsilon < 1$ remains infinite. Indeed, for each data it can happen that $Sf = 0$, and so that the error of an arbitrary $U$ is at least 1.

In the average case setting, the situation is quite different since the complexity is finite if $\sigma_n$ goes to zero as $n$ tends to $+\infty$. For $y = N(f)$, where $N$ is nonadaptive information, consider the algorithm

\[
\phi^*(y) = a(y) + \sqrt{\sigma} \text{sign}(a(y)),
\]

where $a(y)$ is the mean of the conditional measure $\nu(\cdot|y)$ and $\sigma = \sigma(N)$ is its variance. It is possible to show that the error of $U^*(f) = \phi^*(N(f))$ is equal to
\[ \int_F \frac{|Sf - U^*(f)|}{|U^*(f)|} \mu(df) = \frac{\sqrt{\sigma}}{\|w\|} \ln \left( \frac{\|w\|}{\sqrt{\sigma}} \right) \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} |x + 1|e^{-x^2/2} \, dx (1 + o(1)). \]

It is interesting to note that the error of the \( \mu \)-spline algorithm \( \phi^s, \phi^s(y) = a(y) \), is infinite. Thus the "small" difference between the algorithms \( \phi^s \) and \( \phi^*, |\phi^*(y) - \phi^s(y)| = \sqrt{\sigma} \), is essential and makes the error finite.

The error of \( U^* \) is less than \( \varepsilon \) if the information \( N \) is given by \([L_1^*, \ldots, L_n^*] \) with \( L_1^* \) being the solution of (2.11) and \( n^* \) is proportional to the smallest \( n \) for which

\[ \sqrt{\frac{\sigma_n}{S(C_\mu S) - \sigma_n}} \ln \frac{S(C_\mu S) - \sigma_n}{\sigma_n} \leq \varepsilon. \]

Thus the complexity is finite and at most proportional to \( c n^* \). Its actual value is not known.

In the probabilistic setting, the change in the definition of the relative error does not really affect the complexity. Indeed, observe that for \( \varepsilon < 1 \),

\[ \frac{|Sf - U(f)|}{|U(f)|} \leq \varepsilon \text{ implies that } \frac{|Sf - U(f)|}{|Sf|} \leq \frac{\varepsilon}{1 - \varepsilon}, \]

and, vice versa,

\[ \frac{|Sf - U(f)|}{|S(f)|} \leq \varepsilon \text{ implies that } \frac{|Sf - U(f)|}{|U(f)|} \leq \frac{\varepsilon}{1 - \varepsilon}. \]

Therefore, the \( \varepsilon \)-complexity with the new definition of relative error is no smaller than \( \text{comp}^{\text{prob}}(\varepsilon/(1 - \varepsilon)) \), and is no greater than \( \text{comp}^{\text{prob}}(\varepsilon/(1 + \varepsilon)) \). Thus for small \( \varepsilon \), they are practically the same.

(ii) We now discuss a modification of relative error which combines relative and absolute errors together. One may argue that the relative error \( |Sf - U(f)|/|Sf| \) or \( |Sf - U(f)|/|U(f)| \) is not always a reasonable measure of error since for small \( |Sf| \) or \( |U(f)| \) one would rather want the absolute error \( |Sf - U(f)| \) small instead of the relative one. Therefore it seems reasonable to modify the relative error by adding a small positive number \( \eta \) to the denominator. That is, the distance between \( Sf \) and \( U(f) \) is now given by

\[ \frac{|Sf - U(f)|}{|Sf| + \eta}. \]
For $|Sf| \gg \eta$ we are still close to the relative error, whereas for $|Sf| \ll \eta$ we are close to $\eta^{-1}$ times the absolute error. Let $\text{comp}(\epsilon, \eta)$ denote the complexity with this error criterion.

This complexity is related to the complexity with the absolute error criterion. For example, in the worst case setting, $\text{comp}(\epsilon, \eta)$ is roughly equal to the complexity with the absolute error criterion for $(\epsilon \eta)/q$ if the space $F$ is replaced by the ball $\{f \in F : \|f\| \leq q\}$, (see Traub and Woźniakowski, 1980, p. 196).

In the average case setting, it can be shown that for small $\eta$, the complexity $\text{comp}(\epsilon, \eta)$ is roughly at most equal to $cn$, where $n$ is the smallest integer such that

$$\sqrt{\frac{\sigma_n}{S(C_\mu S)} - \sigma_n} \leq \frac{\sqrt{S(C_\mu S) - \sigma_n}}{\eta} \leq \sqrt{\frac{\pi}{2}} \epsilon.$$ 

From Lee and Wasilkowski (1986), we thus conclude that $\text{comp}(\epsilon, \eta)$ is roughly at most equal to the average complexity with the absolute error criterion for $\sqrt{\pi/2} \|S\|_\mu \epsilon/\ln(1/\eta)$.

In the probabilistic setting, $\text{comp}(\epsilon, \eta)$ is clearly no greater than the probabilistic complexity with the absolute error criterion for $\epsilon \eta$. This means that for small $\epsilon$, $\delta$, and $\eta$, the complexity $\text{comp}(\epsilon, \eta)$ is roughly at most equal to $cn$, where $n$ is the minimal integer such that

$$\sqrt{\sigma_n} \leq \frac{\epsilon \eta}{\sqrt{2} \ln(1/\delta)}$$

(see Lee and Wasilkowski, 1986; Woźniakowski, 1986b). Observe the change in the dependence on $\delta$. For the relative error criterion, the probabilistic complexity depends on $\delta$ (see (2.12)), whereas the upper bound on $\text{comp}(\epsilon, \eta)$ depends merely on $1/\sqrt{\ln(1/\delta)}$. The actual values of $\text{comp}(\epsilon, \eta)$ in the average case and probabilistic settings are not known.

(iii) In this paper we focus on approximation of continuous linear functionals. It is, of course, of interest to also study the relative error criterion for linear operators, $S : F \mapsto G$, where $G$ is a normed linear space.

For the worse case setting, the $\epsilon$-complexity remains infinite whenever (2.9) holds. In fact, the proof of Section 4 remains true for an arbitrary space $G$. For the modified relative error as in (ii) of this section, the $\epsilon$-complexity is roughly the same as the complexity with the absolute error criterion for $(\epsilon \eta)/q$ if $F$ is replaced by the ball of radius $q$.

For the average case setting, the situation changes. For a Hilbert space $G = S(F)$ we have

$$\int_F \frac{\|Sf - U(f)\|}{\|Sf\|} \mu(df) \leq \left( \int_F \|Sf - U(f)\| \mu(df) \right)^{1/p} \left( \int_F \frac{1}{\|Sf\|^q} \mu(df) \right)^{1/q},$$
where \(1/p + 1/q = 1\), and \(p \geq 1\). Observe that

\[
\int_F \frac{1}{||Sf||^q} \mu(df) = \int_G \frac{1}{||x||^q} \nu(df) < -\infty
\]

iff \(q < \text{dim}(G)\). Thus if \(\text{dim}(G) \geq 2\), we take \(p > 2\) and \(q < 2 \leq \text{dim}(G)\). This shows that the average relative error is a multiple of the absolute error in the \(L_p\) norm. This absolute error can be arbitrarily small if

\[
\sigma_n = \inf_{L \in L} \inf_{a \in G} \|S - \sum_{i=1}^{n} a_i L_i\|_\mu^2
\]

goes to zero with \(n\) tending to infinity. Here \(a_i L_i\) denotes an operator such that \((a_i L_i)(f) = L_i(f) a_i\) and \(\|S\|_\mu = (\int_F \|Sf\|^2 \mu(df))^{1/2}\). Then the average \(\varepsilon\)-complexity under the relative error criterion is finite. We conjecture that the \(\varepsilon\)-complexity under the relative error criterion is of the same order as the \(\varepsilon\)-complexity with the absolute error (see Woźniakowski (1987), where a similar problem is discussed).

For the probabilistic setting, the \(\varepsilon\)-complexities of the relative and absolute errors also seem to be roughly the same. We hope to study these issues in the future.

**Acknowledgments**

We are grateful for valuable comments from G. W. Wasilkowski and A. G. Werschulz.

**References**


