JOURNAL OF COMPLEXITY 3, 57–80 (1987)

Average Complexity for Linear Operators over Bounded Domains*

H. WOŹNIAKOWSKI

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

Received September 10, 1986

Suppose one wants to compare worst case and average complexities for approximation of a linear operator. In order to get a fair comparison the complexities have to be obtained for the same domain of the linear operator. In previous papers, average complexity was studied when the domain was the entire space. To avoid trivial results, worst case complexity has been studied for bounded domains, and in particular, for balls of finite radius. In this paper we study the average complexity for approximation of linear operators whose domain is a ball of finite radius q. We prove that the average complexities even for modest q and for $q = +\infty$ are closely related. This and existing results enable us to compare the worst case and average complexities for balls of finite radius. We also analyze the average complexity for the normalized and relative errors. The paper is illustrated by integration of functions of one variable and by approximation of functions of d variables which are equipped with a Wiener measure. © 1987 Academic Press, Inc.

INTRODUCTION

The worst case and average complexities have been analyzed for approximation of linear operators in many papers. A recent survey and an extensive bibliography may be found in Woźniakowski (1986).

The worst case setting has been analyzed, in particular, for elements belonging to a ball B_q of arbitrary radius q, i.e., for elements f for which $\|f\| \leq q$, where $\|\cdot\|$ stands for a norm or seminorm. The worst case complexity is a nondecreasing function of q. For nontrivial problems, it is infinity for $q = +\infty$. This dictates finite q in the worst case setting.

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

0885-064X/87 $3.00

The average setting has been analyzed for Banach spaces equipped with Gaussian measures, i.e., for elements belonging to the ball B_q with $q = +\infty$. See a recent paper of Wasilkowski (1986) and the papers cited there.

Suppose one wishes to compare the complexities of a problem in the worst case and average settings. It is then natural to assume that $f \in B_q$ for a finite q in order to make the worst case complexity finite. It is then necessary to analyze the average setting over the balls B_q instead of the whole space $B_{+\infty}$.

The purpose of this paper is to study the average setting over balls B_q. We assume that the ball B_q is equipped with a truncated Gaussian measure. To preserve essential properties of Gaussian measures, it is necessary to assume that q is not too small. Since the tail of a Gaussian measure goes to zero exponentially fast, we need only a mild assumption on q.

It is intuitively obvious that for large q, the average complexity should be close to the average complexity for the whole space. We prove it in quantative terms even for moderate q. This will be done for approximation of linear operators defined on a separable Banach space into a separable Hilbert space. We find lower and upper bounds on the average complexity for the ball B_q in terms of the average complexity for the whole space. These bounds are quite tight. Under mild assumptions, the average complexities for a finite q and $q = +\infty$ differ by a factor $1 + O(e^{-q^2 b})$, where b depends only on a Gaussian measure of the Banach space.

The above results are for an absolute error. We also analyze the average complexity for normalized and relative errors. Using the results for the balls B_q and the results of Wasilkowski (1986), we find lower and upper bounds on the average complexity for the normalized and relative errors in terms of the average complexity for the absolute error. Under suitable assumptions, we prove that the average complexities for the absolute, normalized, and relative errors are proportional to each other. Tight bounds on the average complexity for the normalized or relative error will require a more refined analysis than has been carried out here.

We outline the contents of this paper. In Section 2 we precisely define the average complexity for the balls B_q and state the main theorem which relates the complexities for finite and infinite q. In Section 3 we analyze the minimal average error of algorithms using given information. The proof of the main theorem is given in Section 4 and is primarily based on estimates of Section 3. Section 5 deals with the comparison between the worst case and the average complexities. We consider three problems: the integration problem defined on the Banach space of r times continuously differentiable scalar functions of one variable equipped with the classical Wiener measures placed on rth derivatives, the approximation problem defined on the Banach space of scalar functions of d variables which are r_i times continuously differentiable with respect to the ith variable and equipped with the classical Wiener measure placed on r_i-partial derivatives, and a problem defined in Hilbert spaces with polynomial or exponential distributions of eigenvalues. The first two problems are analyzed using the results of Lee and Wasilkowski (1986), and Papageorgiou and Wasilkowski (1986), respectively. The final section deals with normalized and relative errors.

2. Average Complexity

Let F be a separable Banach space and let G be a separable Hilbert space, both over the real field. Consider a continuous linear operator S, S: $F \to G$. We wish to approximate $S(f)$ for all f from the ball $B_q = \{f \in F$: $\|f\| \leq q\}$.

We assume that the ball B_q is equipped with a probability measure μ_q defined as follows. Let μ be a Gaussian measure defined on Borel sets of F with mean zero and a correlation operator C_μ, C_μ: $F^* \to F$. The definition and basic properties of Gaussian measures may be found in Kuo (1975) and Vakhania (1981). The probability measure μ_q is defined as a truncation of the Gaussian measure μ to the ball B_q. That is, for any Borel set A of F,

$$\mu_q(A) = \frac{\mu(A \cap B_q)}{\mu(B_q)}. \tag{2.1}$$

Note that for finite q the measure μ_q is not Gaussian. However, for large q, u_q "resembles" μ.

In order to approximate $S(f)$, we need to compute some information about f. Let Λ denote a set of continuous linear functionals L for which $L(f)$ can be computed for all f from F. That is, Λ is a subset of F^*. For instance, if $S(f)$ is given by the integral of a function f, Λ usually consists of function and/or derivative evaluations, $L(f) = f^{(i)}(x)$ for some i and x from the domain of f. For some problems, such as the approximation problem $S(f) = f$, we may assume that an arbitrary functional from F^* can be computed. Then $\Lambda = F^*$.

As in Wasilkowski (1986) we consider adaptive information N defined for arbitrary f as

$$N(f) = [L_1(f), L_2(f; y_1), \ldots, L_{n(f)}(f; y_1, \ldots, y_{n(f)-1})], \tag{2.2}$$

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 ith information evaluation. Here we assume that for fixed $y = [y_1, y_2, \ldots]$, the functional

$$L_{i,y} \equiv L_i(\cdot, y_1, \ldots, y_{i-1}): F \to \Re \tag{2.3}$$

belongs to the class Λ.

The number $n(f)$ of information evaluations is determined as follows. Let $\text{ter}_i: \Re^i \to \{0, 1\}$ be a given Boolean function. Knowing $y_1, \ldots, y_i$ we compute $\text{ter}_i(y_1, \ldots, y_i)$ and if it is 1 we terminate the information evaluations and set $n(f) = i$. If not, we select $L_{i+1,y} = L_{i+1}(\cdot; y_1, \ldots, y_i)$ from the class Λ, compute $y_{i+1} = L_{i+1,y}(f)$ and then the process is repeated. The number $n(f)$ is defined as

$$n(f) = \inf\{i : \text{ter}_i(y_1, \ldots, y_i) = 1\} \tag{2.4}$$

with the convention that $\inf \emptyset = +\infty$.

The essence of (2.2)–(2.4) is that the choice of the next information evaluation as well as their total number depends adaptively on the previously computed information about f. In what follows we assume that N is measurable.

We compute an approximation to $S(f)$ by combining the information $N(f)$. That is, we approximate $S(f)$ by $\phi(N(f))$, where $\phi: N(B_q) \to G$ is measurable. The mapping ϕ is called an algorithm. The average error of the algorithm ϕ is defined as

$$e^{\text{avg}}(\phi, N, q) = \left\{\int_{B_q} \|Sf - \phi(N(f))\|^2 \mu_q(df)\right\}^{1/2} \tag{2.5}$$

For a given nonnegative ε, we want to find N and ϕ such that $e^{\text{avg}}(\phi, N, q) \leq \varepsilon$ and the average cost of computing $\phi(N(f))$ is minimal. The cost of computing $\phi(N(f))$ is defined as follows.

Assume that each information evaluation costs c, $c > 0$, and that we can multiply elements from G by scalars, add two elements from G, and compare real numbers at unit cost. Usually $c \gg 1$.

In order to compute $\phi(N(f))$ we first compute $y = N(f)$ and then $\phi(y)$. Let $\text{cost}(N, f)$ denote the cost of computing $y = N(f)$ and let $\text{cost}(\phi, y)$ denote the cost of computing $\phi(y)$. Observe that $\text{cost}(N, f) \geq cn(f)$. The average cost of (ϕ, N) is defined as

$$\text{cost}^{\text{avg}}(\phi, N, q) = \int_{B_q} \{\text{cost}(N, f) + \text{cost}(\phi, N(f))\}\mu_q(df). \tag{2.6}$$

By the average (ε, q)-complexity we mean the minimal average cost needed to approximate $S(f)$ on the average to within ε,

$$\text{comp}^{\text{avg}}(\varepsilon, q) = \inf\{\text{cost}^{\text{avg}}(\phi, N, q) : \phi, N \text{ such that } e^{\text{avg}}(\phi, N, q) \leq \varepsilon\}. \tag{2.7}$$

For $q = +\infty$ we adapt the notation $\mathrm{comp}^{\mathrm{avg}}(\varepsilon) \equiv \mathrm{comp}^{\mathrm{avg}}(\varepsilon, +\infty)$. In this case, $B_q = F$ and the measure $\mu_q = \mu$ becomes Gaussian. As mentioned in the Introduction the average complexity $\mathrm{comp}^{\mathrm{avg}}(\varepsilon)$ has been studied before in many papers.

Relations between $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, q)$ and $\mathrm{comp}^{\mathrm{avg}}(\varepsilon)$ depend on the measure $\mu(B_q)$ of the ball B_q. More precisely, they depend on how fast $\mu(B_q)$ goes to one as q tends to infinity. For Gaussian measures, $\mu(B_q)$ goes to one exponentially fast. Indeed, from Borell (1975, 1976) we have

$$\mu(B_q) = 1 = e^{-q^2 a^*(1+o(1))} \qquad \text{as } q \to +\infty, \tag{2.8}$$

where

$$a^* = [2 \sup\{L(C_\mu L) : L \in F^*, \|L\| = 1\}]^{-1}. \tag{2.9}$$

The difference $1 - \mu(B_q)$ can be also estimated using Fernique's theorem (see, for instance, Araujo and Giné, 1980, p. 141), which states together with (2.8) that $\int_F e^{\|f\|^2 a}\, \mu(df) < +\infty$ for any number a such that $a < a^*$. Then

$$\begin{aligned} 1 - \mu(B_q) = \int_{\|f\|>q} \mu(df) \le e^{-q^2 a} \int_{\|f\|>q} e^{\|f\|^2 a} \mu(df) \le e^{-q^2 a} \\ \int_F e^{\|f\|^2 a} \mu(df). \end{aligned} \tag{2.10}$$

Let $x = 1 - \mu(B_q)$. In what follows we assume that q is chosen such that

$$1 - x - \sqrt{3x} > 0. \tag{2.11}$$

That is, $\mu(B_q) \ge (\sqrt{21} - 3)/2 \simeq 0.8$. Due to (2.8) and (2.10) we have

$$\begin{aligned} x \le (5 - \sqrt{21})/2 \simeq 0.2, \qquad x \le e^{-q^2 a} \int_F e^{\|f\|^2 a} \mu(df), \\ x = e^{-q^2 a^*(1+o(1))}, \end{aligned} \tag{2.12}$$

for any $a < a^*$.

Remark 2.1. We illustrate (2.12) for a separable Hilbert space F. Let $\{\beta_i\}$ be eigenvalues of the correlation operator C_μ of the Gaussian measure μ. Then for $a < a^* = 1/(2 \max \beta_i)$,

$$\int_F e^{\|f\|^2 a} \mu(df) = \prod_{i=1}^{\infty} (1 - 2a\beta_i)^{-1/2}.$$

For $a < 1/(2\, \mathrm{trace}(C_\mu))$, $\mathrm{trace}(C_\mu) = \Sigma_i \beta_i$, the above integral is no greater than $[1 - 2a\, \mathrm{trace}(C_\mu)]^{-1/2}$. Thus

$$x \in [0, e^{-q^2 a}(1 - 2a \operatorname{trace}(C_\mu))^{-1/2}].$$

Assume that $\beta_i = i^{-p}$ for some $p > 1$. Then $\operatorname{trace}(C_\mu) = \zeta(p)$ is Riemann's zeta function at p and $\zeta(p) \leq 1 + 1/(p - 1)$. For $a = (p - 1)/(4p)$, we get

$$x \in [0, \sqrt{2}e^{-q^2(p-1)/4p}].$$

We are ready to state relations between the two average complexities.

THEOREM 2.1. (i) *Let* $x = 1 - \mu(B_q)$ *and let* q *satisfy* (2.11). *Then*

$$\frac{c}{c+2}\frac{1 - x - \sqrt{3x}}{1 - x}\operatorname{comp}^{\mathrm{avg}}\left(\sqrt{\frac{1 - x}{1 - x - \sqrt{3x}}}\,\varepsilon\right) \leq \operatorname{comp}^{\mathrm{avg}}(\varepsilon, q) \leq \frac{1}{1 - x}\operatorname{comp}^{\mathrm{avg}}(\sqrt{1 - x}\,\varepsilon). \tag{2.13}$$

(ii) *If*

$$\operatorname{comp}^{\mathrm{avg}}((1 + \delta)\varepsilon) = \operatorname{comp}^{\mathrm{avg}}(\varepsilon)(1 + O(\delta)) \quad \text{as } \delta \to 0 \tag{2.14}$$

then

$$\operatorname{comp}^{\mathrm{avg}}(\varepsilon, q) = \operatorname{comp}^{\mathrm{avg}}(\varepsilon)(1 + o(e^{-q^2 a/2}))(1 + O(c^{-1}))$$

for any $a < a^*$ *as* q *and* c *go to infinity.*

The proof of Theorem 2.1 is given in Section 4 and is based on the analysis of Section 3.

3. AVERAGE RADIUS OF INFORMATION

In order to prove Theorem 2.1 we study the minimal average error of algorithms ϕ using adaptive information N given by (2.2). Due to geometrical interpretations, this minimal average error is called the average radius of information,

$$r^{\mathrm{avg}}(N, q) = \inf_{\phi} e(\phi, N, q). \tag{3.1}$$

We stress that the infimum in (3.1) is over *all* measurable mappings ϕ, ϕ: $N(B_q) \to G$.

We find relations between $r^{\mathrm{avg}}(N, q)$ and $r^{\mathrm{avg}}(N) = r^{\mathrm{avg}}(N, +\infty)$. One may hope that both the radii are essentially the same for large q. More precisely, that there exists a function δ: $\Re_+ \to \Re_+$, $\lim_{q\to\infty} \delta(q) = 0$, such that for any adaptive information N we have

$$r^{\text{avg}}(N, q) = (1 + \delta(N, q))r^{\text{avg}}(N), \qquad |\delta(N, q)| \le \delta(q). \tag{3.2}$$

We show this for *nonadaptive* information N, i.e., for $N = [L_1, L_2, \ldots, L_k]$ where $L_i \in \Lambda$, and with $\delta(q) = \sqrt{3(1 - \mu(B_q))} = o(e^{-q^2 a/2})$; see (2.8)–(2.10).

For adaptive information, (3.2) is in general false. It can happen that for any finite q there exists adaptive information N such that $r^{\text{avg}}(N, q) = 0$ and $r^{\text{avg}}(N) > 0$ as illustrated by the following example.

EXAMPLE 3.1. Let $F = G$ and let ζ_i be an orthonormal basis of G. Define $y_1 = L_1(f) = (f, \zeta_1)$ and

$$y_i = L_{i,y}(f) = \begin{cases} (f, \zeta_i) & \text{if } \Sigma_{j=1}^{i-1} y_j^2 \le q^2, \\ 0 & \text{otherwise.} \end{cases}$$

Then $N(f) = [L_1(f), L_{2,y}(f), \ldots]$ is adaptive. Observe that for $\|f\| \le q$, we can recover f exactly from $N(f)$. Therefore the algorithm $\phi(N(f)) = Sf$ is well defined and its average error is zero. Thus $r^{\text{avg}}(N, q) = 0$.

On the other hand, for $|(f, \zeta_1)| > q$, $N(f) = (f, \zeta_1)$. Therefore

$$r^{\text{avg}}(N)^2 \ge \int_{|(f,\zeta_1)|>q} \|Sf - (f, \zeta_1)S\zeta_1\|^2 \mu(df),$$

which is positive for bijective S and C_μ. Thus, (3.2) is false in this case.

For adaptive information N we relax (3.2) by showing that

$$r^{\text{avg}}(N, q) \ge (1 - \delta(q))r^{\text{avg}}(N^*) \tag{3.3}$$

for some information N^* whose cardinality is roughly the same as the cardinality N and whose structure is much simpler than the structure of N. This will be done by applying the results of Wasilkowski (1986).

By cardinality of adaptive information N we mean the average number of information evaluations,

$$\text{card}^{\text{avg}}(N, q) = \int_{B_q} n(f)\mu_q(df). \tag{3.4}$$

As before, we denote $\text{card}^{\text{avg}}(N, +\infty)$ by $\text{card}^{\text{avg}}(N)$.

The function $\delta(q)$ in (3.3) will be given in terms of $x = 1 - \mu(Bq)$ and we shall have $\delta(q) = o(e^{-q^2 a/2})$.

In order to define information N^* in (3.3) we proceed as in Wasilkowski (1986). For a vector $y \in N(B_q) \cap \Re^k$, define

$$N_y(f) = [L_{1,y}(f), L_{2,y}(f), \ldots, L_{k,y}(f)]. \tag{3.5}$$

The information N_y is nonadaptive of cardinality k and is derived from adaptive information N by fixing the values of y. Let $r^{\text{avg}}(N_y)$ denote its average radius. Let

$$r_k = \inf_{y \in N(B_q) \cap \Re^k} r^{\text{avg}}(N_y) \tag{3.6}$$

with the convention that $r_k = 0$ if $N(B_q) \cap \Re^k = \emptyset$. Without loss of generality we assume that there exists a vector y_k such that for $N_k = N_{y_k}$ we have $r_k = r^{\text{avg}}(N_k)$.

We are ready to state relations between $r^{\text{avg}}(N, q)$, $r^{\text{avg}}(N)$, and $r^{\text{avg}}(N^*)$.

THEOREM 3.1. *Let $x = 1 - \mu(B_q)$ with q satisfying* (2.11).

(i) *Let N be nonadaptive. Then*

$$r^{\text{avg}}(N, q) \geq r^{\text{avg}}(N)\sqrt{1 - \sqrt{3x}},$$

$$r^{\text{avg}}(N, q) = (1 + \delta(N, q))r^{\text{avg}}(N), \qquad |\delta(N, q)| \leq \sqrt{3x}.$$

(ii) *Let N be adaptive with $n(f) \equiv k$. Then*

$$r^{\text{avg}}(N, q) \geq r^{\text{avg}}(N_k)\sqrt{1 - \sqrt{3x}}.$$

(iii) *Let N be arbitrary adaptive. Then there exists information N^* such that*

$$r^{\text{avg}}(N, q) \geq r^{\text{avg}}(N)\sqrt{1 - \frac{\sqrt{3x}}{1 - x}},$$

$$\text{card}^{\text{avg}}(N, q) \geq \text{card}^{\text{avg}}(N^*)\,\frac{1 - x - \sqrt{3x}}{1 - x}.$$

Information N^* consists of two nonadaptive information

$$N^*(f) = \begin{cases} N_{k_1}(f) & \text{if } L_1(f) \in A, \\ N_{k_2}(f) & \text{otherwise,} \end{cases}$$

for some indices k_1 and k_2, and a Borel set A of $\Re$.

Proof. Let N be adaptive information given by (2.2). Without loss of generality we can normalize $L_{i,y}$ such that $L_{i,y}(C_\mu L_{j,y}) = \delta_{ij}$, C_μ being a covariance operator of Gaussian measure μ. As in Lee and Wasilkowski (1986) decompose μ as follows. Let $\mu_1 = \mu N^{-1}$ and let $\mu_2(\cdot|y)$ be a conditional probability measure such that $\mu_2(N^{-1}(y)|y) = 1$ and $\mu(B) = \int_{N(F)} \mu_2(B|y)\mu_1(dy)$ for any Borel set of F_1.

Let $y = [y_1, y_2, \ldots, y_k] \in N(F)$. Then $\mu_2(\cdot|y)$ is Gaussian with mean $\sigma(y) = \Sigma_{i=1}^k y_i C_\mu L_{i,y}$ and correlation operator $C_y = C_\mu - \Sigma_{i=1}^k L_{i,y} C_\mu(\cdot) C_\mu L_{i,y}$.

Let $\nu(\cdot|y) = \mu_2(S^{-1}\cdot|y)$. Then $\nu(\cdot|y)$ is Gaussian on the separable Hilbert space G with mean $S\sigma(y)$ and correlation operator $C_{\nu,y}(L) = SC_y(LS)\ \forall L \in G^*$. The operator $C_{\nu,y}$ is self-adjoint, nonnegative definite, and has a finite trace. Let $\lambda_i = \lambda_i(y)$, $\zeta_i = \zeta_i(y)$ be its orthonormal eigenpairs, $C_{\nu,y}\zeta_i = \lambda_i\zeta_i$, $\lambda_1 \geq \lambda_2 \geq \cdots \geq 0$, and $\Sigma_i \lambda_i < +\infty$.

Let ϕ be an arbitrary algorithm using N with a finite average error. Since $\|Sf - \phi(N(f))\|^2 = \Sigma_i (Sf - \phi(N(f)), \zeta_i(N(f)))^2$, its average error can be expressed as

$$u \stackrel{\text{df}}{=} \mu(B_q)e(\phi, N, q)^2 = \int_{N(F)} \int_F \sum_i (Sf - \phi(y), \zeta_i(y))^2 \times (1 - b(f, q))\mu_2(df|y)\mu_1(df), \tag{3.7}$$

where $b(f, q) = 1$ for $\|f\| \geq q$ and $b(f, q) = 0$ otherwise. For $\lambda_i = \lambda_i(y) > 0$, define $\eta_i = \eta_i(y) = (1/\sqrt{\lambda_i})\zeta_i$ and

$$g_i(y) = \int_F (Sf - \phi(y), \eta_i)^2 b(f, q)\mu_2(df|y).$$

Note that

$$\begin{aligned} 0 \leq g_i(y) &\leq \int_F (Sf - \phi(y), \eta_i)^2 \mu_2(df|y) \\ &= \int_G (g - \phi(y), \eta_i)^2 \nu(dg|y) = \frac{1}{\sqrt{2\pi}} \int_{\Re} (t - (\phi(y), \eta_i))^2 \\ &\quad \times e^{-(t-(S\sigma(y),\eta_i))^2/2} dt = 1 + (S\sigma(y) - \phi(y), \eta_i)^2 = 1 + e_i^2, \end{aligned} \tag{3.8}$$

where $e_i = e_i(y) = (S\sigma(y) - \phi(y), \eta_i)$. Furthermore

$$\begin{aligned} g_i^2(y) &\leq \int_F (Sf - \phi(y), \eta_i)^4 \mu_2(df|y) \int_F b(f, q)\mu_2(df|y) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\Re} (t - e_i)^4 e^{-t^2/2} dt (1 - \mu_2(B_q|y)) \\ &= (3 + 6e_i^2 + e_i^4)(1 - \mu_2(B_q|y)). \end{aligned} \tag{3.9}$$

We now rewrite (3.7) as

$$u = \int_{N(F)} \sum_i \lambda_i(y)(1 + e_i^2(y) - g_i(y))\mu_1(dy). \tag{3.10}$$

The series in (3.10) is convergent since it is no greater than $\Sigma_i \lambda_i(y) + \|S\sigma(y) - \phi(y)\|^2$. Let

$$h(y) = \frac{\Sigma_i \lambda_i(y) g_i(y)}{\Sigma_i \lambda_i(y)(1 + e_i^2(y))}.$$

Then (3.8) yields that $h(y) \in [0, 1]$ and (3.9) yields that

$$\begin{aligned} h(y) &\le \sup_i \frac{g_i(y)}{1 + e_i^2(y)} \le \sup_{t \ge 0} \frac{\sqrt{3 + 6t + t^2}}{1 + t} (1 - \mu_2(B_q|y))^{1/2} \\ &= \sqrt{3}(1 - \mu_2(B_q|y))^{1/2}. \end{aligned} \tag{3.11}$$

From this we conclude that

$$\begin{aligned} \int_{N(F)} h(y)\mu_1(dy) &\le \left(\int_{N(F)} h^2(y)\mu_1(dy)\right)^{1/2} \\ &= \sqrt{3}\left(1 - \int_{N(F)} \mu_2(B_q|y)\,\mu_1(dy)\right)^{1/2} \\ &= \sqrt{3}(1 - \mu(B_q))^{1/2} = \sqrt{3x}. \end{aligned} \tag{3.12}$$

Finally we rewrite (3.10) as

$$\begin{aligned} u &= \int_{N(F)} \sum_i \lambda_i(y)(1 + e_i^2(y))(1 - h(y))\mu_1(dy) \\ &\ge \int_{N(F)} \sum_i \lambda_i(y)(1 - h(y))\mu_1(dy). \end{aligned} \tag{3.13}$$

We consider three cases as indicated in Theorem 3.1.

(i) *Nonadaptive information.* For nonadaptive information the correlation operator $C_{\nu,y}$ does not depend on y. Therefore $\lambda_i(y) \equiv \lambda_i$, $\forall y$. It is known (see, for instance, Lee and Wasilkowski, 1986) that

$$r^{\text{avg}}(N) = \sqrt{\sum_i \lambda_i}.$$

(It also follows from (3.13) with $q = +\infty$. Then $g_i \equiv h \equiv 0$ and (3.13) is minimized for $e_i = 0$, i.e., for $\phi(y) = S\sigma(y)$.)

From (3.13) and (3.12) we have

$$\begin{aligned} e^{\text{avg}}(\phi, N, q) &\ge \sqrt{u} \ge r^{\text{avg}}(N)\left(1 - \int_{N(F)} h(y)\mu_1(dy)\right)^{1/2} \\ &\ge r^{\text{avg}}(N)(1 - \sqrt{3x})^{1/2}. \end{aligned}$$

This holds for any ϕ and therefore $r^{\text{avg}}(N, q) \geq r^{\text{avg}}(N)\sqrt{1 - \sqrt{(3x)}}$.

On the other hand, $r^{\text{avg}}(N, q) \leq (1/\sqrt{\mu(B_q)})r^{\text{avg}}(N)$, which with the previous estimate leads to

$$r^{\text{avg}}(N, q) = (1 + \delta(N, q))r^{\text{avg}}(N), \qquad |\delta(N, q)| \leq \sqrt{3x},$$

as claimed.

(ii) *Adaptive information with $n(f) \equiv k$.* For nonadaptive information N_y given by (3.5) we have $r^{\text{avg}}(N_y) = \sqrt{\Sigma_i \lambda_i(y)}$. Then (3.6) yields $\sqrt{\Sigma_i \lambda_i(y)} \geq r_k = r^{\text{avg}}(N_k)$. Applying this to (3.13) we have

$$e(\phi, N, q) \geq \sqrt{u} \geq r^{\text{avg}}(N_k)(1 - \sqrt{3x})^{1/2}.$$

This implies (ii) of Theorem 3.1.

(iii) *Adaptive information with varying $n(f)$.* Let $A_k = N(B_q) \cap \Re^k$. Then

$$\text{card}^{\text{avg}}(N, q) = \frac{1}{\mu(B_q)} \int_{B_q} n(f)\mu(df) = \frac{1}{1 - x} \sum_{k=1}^{\infty} k\mu_1(A_k).$$

and $\Sigma_{k=1}^{\infty} \mu_1(A_k) = \mu(B_q)$. We rewrite (3.13) using (3.6) as

$$\begin{aligned} u &\geq \sum_{k=1}^{\infty} \int_{A_k} \sum_{i=1}^{\infty} \lambda_i(y)(1 - h(y))\mu_1(dy) \\ &\geq \sum_{k=1}^{\infty} r_k^2 \int_{A_k} (1 - h(y))\mu_1(dy) = \sum_{k=1}^{\infty} r_k^2 \beta_k, \end{aligned}$$

where $\beta_k = \mu_1(A_k) - \int_{A_k} h(y)\mu_1(dy)$. Obviously, $\beta_k \geq 0$ and $\beta = \Sigma_{k=1}^{\infty} \beta_k = \mu(B_q) - \int_{N(B_q)} h(y)\mu_1(dy) \geq 1 - x - \sqrt{3x} > 0$ due to (2.11). Let $a_k = \beta_k/\beta$. Then $\Sigma_k a_k = 1$ and

$$\sum_k k\, a_k \leq \frac{1}{\beta} \sum_k k\mu_1(A_k) \leq \frac{1 - x}{1 - x - \sqrt{3x}} \text{card}^{\text{avg}}(N, q) =: b.$$

From this we conclude that

$$(1 - x)r^{\text{avg}}(N, q)^2 \geq \beta \inf\left\{\sum_{k=1}^{\infty} r_k^2 a_k : a_k \geq 0, \sum_k a_k = 1, \sum_k k a_k \geq b\right\}.$$

The above minimization problem was studied recently by Wasilkowski (1986). He proved that there exist a Borel set A of $\Re$ and two indices k_1

and k_2 such that the infimum is attained by the square of the average radius of information N^* such that

$$\text{card}^{\text{avg}}(N^*) \leq \frac{1 - x}{1 - x - \sqrt{3x}} \text{card}^{\text{avg}}(N, q)$$

and $N^*(f) = N_{k_1}(f)$ if $L_1(f) \in A$ and $N^*(f) = N_{k_2}(f)$ otherwise. Thus

$$r^{\text{avg}}(N, q) \geq \sqrt{\frac{\beta}{1 - x}}\, r^{\text{avg}}(N^*) \geq \sqrt{1 - \frac{\sqrt{3x}}{1 - x}}\, r^{\text{avg}}(N^*),$$

as claimed.

4. Proof of Theorem 2.1

Without loss of generality we can assume that the infimum in (2.7) is attained for a pair (ϕ, N). Thus $\text{cost}^{\text{avg}}(\phi, N, q) = \text{comp}^{\text{avg}}(\varepsilon, q)$ and $e^{\text{avg}}(\phi, N, q) \leq \varepsilon$. From (2.6) we have

$$\text{comp}^{\text{avg}}(\varepsilon, q) \geq \int_{B_q} \text{cost}(N, f)\, \mu_q(df) \geq c \int_{B_q} n(f)\, \mu_q(df) = c \, \text{card}^{\text{avg}}(N, q).$$

Obviously, $r^{\text{avg}}(N, q) \leq \varepsilon$. Theorem 3.1(iii) yields that

$$r^{\text{avg}}(N^*) \leq \varepsilon \sqrt{\frac{1 - x}{1 - x - \sqrt{3x}}},$$

$$\begin{aligned} \text{card}^{\text{avg}}(N^*) &\leq \frac{1 - x}{1 - x - \sqrt{3x}} \text{card}^{\text{avg}}(N, q) \\ &\leq \frac{1 - x}{c(1 - x - \sqrt{3x})} \text{comp}^{\text{avg}}(\varepsilon, q). \end{aligned}$$

Information N^* is of the form given in Theorem 3.1(iii). It is known that the algorithm $\phi^*(N^*(f)) = S\sigma(N^*(f))$ has the average error equal to $r^{\text{avg}}(N^*)$, is piecewise linear, and can be computed at average cost equal to

$$(c + 2)\text{card}^{\text{avg}}(N^*) \leq \frac{c + 2}{c} \frac{1 - x}{1 - x - \sqrt{3x}} \text{comp}^{\text{avg}}(\varepsilon, q).$$

Since the pair (ϕ^*, N^*) solves the problem for $\varepsilon_1 = \varepsilon \sqrt{(1 - x)/(1 - x - \sqrt{3x})}$, we have

$$\mathrm{comp}(\varepsilon_1) \le \frac{c+2}{c} \frac{1-x}{1-x-\sqrt{3x}} \mathrm{comp}^{\mathrm{avg}}(\varepsilon, q).$$

as claimed in the left-hand side of (2.13).

To prove the right-hand side of (2.13), take a pair (ϕ, N) for which $e^{\mathrm{avg}}(\phi, N) \le \varepsilon\sqrt{1-x}$ and $\mathrm{cost}^{\mathrm{avg}}(\phi, N) = \mathrm{comp}^{\mathrm{avg}}(\varepsilon\sqrt{1-x})$. Then

$$e^{\mathrm{avg}}(\phi, N, q) \le \frac{1}{\sqrt{1-x}} e^{\mathrm{avg}}(\phi, N) \le \varepsilon.$$

Thus

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, q) \le \frac{1}{1-x} \mathrm{cost}^{\mathrm{avg}}(\phi, N) = \frac{1}{1-x} \mathrm{comp}^{\mathrm{avg}}(\varepsilon\sqrt{1-x}),$$

as claimed. This proves (i) of Theorem 2.1.

The estimate (ii) of Theorem 2.1 easily follows from (2.13) and the fact that $x = o(e^{-q^2 a})$. This completes the proof. ∎

5. Comparison between Worst Case and Average Complexities

The worst case complexity $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, q)$ is defined analogously as the average complexity with the only difference that the integrals in the definitions (2.5) and (2.6) of error and cost are replaced by the supremum over the ball B_q. Clearly, $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, q) \le \mathrm{comp}^{\mathrm{wor}}(\varepsilon, q)$.

In this section we compare the worst case and average complexities for three problems.

5.1 *Integration*

Define the Banach space F as the class of real functions f: $[0, 1] \to \Re$ for which $f(0) = f'(0) = \cdots = f^{(r)}(0) = 0$ and $f^{(r)}$ is continuous. The space F is equipped with the norm $\|f\| = \max_{0\le t\le 1}|f^{(r)}(t)|$ and with the Wiener measure placed on rth derivatives. That is, $\mu(A) = w(D^r A)$, where $D^r f = f^{(r)}$ and w is the classical Wiener measure.

Let $G = \Re$ and let $Sf = \int_0^1 f(t)dt$ be the integral of f. Assume that only function and derivative values can be computed, each of them at cost c. That is, $L \in \Lambda$ iff $L(f) = f^{(j)}(x)$ for some $x \in [0, 1]$ and $j \le r$.

The worst case of this problem has been studied in many papers; see, for instance, Traub and Woźniakowski (1980) and papers cited therein. For $r = 0$, we have

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, q) = \begin{cases} +\infty & \text{if } \varepsilon < q, \\ 0 & \text{if } \varepsilon \ge q. \end{cases}$$

For $r \geq 1$, we have

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, q) = \Theta\left(c\left(\frac{q}{\varepsilon}\right)^{1/r}\right). \tag{5.2}$$

We now turn to the average case, which was studied recently by Lee and Wasilkowski (1986) for $q = +\infty$. They proved that

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r+1)}\right).$$

From Theorem 2.1 we conclude that for large q,

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, q) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r+1)}\right). \tag{5.3}$$

For $r = 0$ and ε less than q, the ratio of the worst case and average case complexity is infinity. The mere continuity of the functions is not enough to solve the problem in the worst cae. The average case can be done with the complexity proportional to $c\varepsilon^{-1}$ due to the extra smoothness of the functions given by the Wiener measure.

For $r \geq 1$, we get from (5.2) and (5.3),

$$\frac{\mathrm{comp}^{\mathrm{wor}}(\varepsilon, q)}{\mathrm{comp}^{\mathrm{avg}}(\varepsilon, q)} = \Theta(q^{1/r}\varepsilon^{1/r(r+1)}). \tag{5.4}$$

This ratio goes to infinity as $\varepsilon \to 0$ or $q \to +\infty$. For example, consider fixed q and $r = 1$. Then the ratio of the two complexities is proportional to $\sqrt{1/\varepsilon}$. That is, the worst case is $\Theta(\sqrt{1/\varepsilon})$ times harder than the average one.

5.2. *Approximation*

As in Papageorgiou and Wasilkowski (1986), consider the approximation problem defined as follows. Let $f\colon D = [0, 1]^d \to \Re$ be a function of d variables. By $f^{(i_1, i_2, \ldots, i_d)}$ we mean i_j times differentiation of f with respect to the j variable, $j = 1, 2, \ldots, d$. Define the Banach space F as the class of functions f for which $f^{(r_1, r_2, \ldots, r_d)}$ is continuous and for which $f^{(i_1, i_2, \ldots, i_d)}(t) = 0$, $\forall\ i_j = 0, 1, \ldots, r_j$ and any t with at least one zero component. The space F is equipped with the norm $\|f\| = \max_{t \in D}|f^{(r_1, \ldots, r_d)}(t)|$ and with the Wiener measure placed on $(r_1, r_2, \ldots, r_d)$ partial derivatives. That is, $\mu(A) = w(D^{r_1, \ldots, r_d}A)$ with $D^{r_1, \ldots, r_d}f = f^{(r_1, \ldots, r_d)}$ and with the classical Wiener measure w.

Let $G = L_2(D)$ and let $Sf = f$ be the embedding operator. Assume that $\Lambda = F^*$, i.e., an arbitrary continuous linear functional can be computed at cost c.

Papageorgiou and Wasilkowski (1986) obtained the worst case complexity and the average complexity for $q = +\infty$. Let $r_{\min} = \min\{r_j: 1 \le j \le d\}$ and let k_d be the number of r_j equal to $r_{\min}$.

If $r_{\min} = 0$ then $\mathrm{comp}^{\mathrm{wor}}(\varepsilon, q) = +\infty$ for small ε. For $r_{\min} \ge 1$, we have

$$\mathrm{comp}^{\mathrm{wor}}(\varepsilon, q) = \Theta\left(c\left(\frac{q}{\varepsilon}\right)^{1/r_{\min}}\left(\ln\frac{q}{\varepsilon}\right)^{k_d-1}\right). \tag{5.5}$$

For the average case,

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r_{\min}+1/2)}\left(\ln\frac{1}{\varepsilon}\right)^{(k_d-1)(1+1/(2r_{\min}+1))}\right).$$

From Theorem 2.1 we conclude that for large q, $\mathrm{comp}^{\mathrm{avg}}(\varepsilon, q) \simeq \mathrm{comp}^{\mathrm{avg}}(\varepsilon)$ and therefore

$$\mathrm{comp}^{\mathrm{avg}}(\varepsilon, q) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r_{\min}+1/2)}\left(\ln\frac{1}{\varepsilon}\right)^{(k_d-1)(1+1/(2r_{\min}+1))}\right). \tag{5.6}$$

For $r_{\min} = 0$ and small ε, the ratio of the worst case and average complexities is infinity. As for the integration problem, the mere continuity of the functions in one of the directions makes it impossible to solve the approximation problem in the worst case. The approximation problem can be solved in the average case with complexity proportional to $c\varepsilon^{-2}(\ln(1/\varepsilon))^{2(k_d-1)}$.

For $r_{\min} > 0$, we get from (5.5) and (5.6) for fixed q and small ε

$$\frac{\mathrm{comp}^{\mathrm{wor}}(\varepsilon, q)}{\mathrm{comp}^{\mathrm{avg}}(\varepsilon, q)} = \Theta\left(\left(\frac{1}{\varepsilon}\right)^{1/r_{\min}(2r_{\min}+1)}\left(\ln\frac{1}{\varepsilon}\right)^{-(k_d-1)/(2r_{\min}+1)}\right). \tag{5.7}$$

The ratio goes to infinity as $\varepsilon \to 0$. For example, consider $r_{\min} = k_d = 1$. Then the ratio is proportional to $(1/\varepsilon)^{1/3}$. That is, the worst case is $\Theta((1/\varepsilon)^{1/3}$ times harder than the average one.

5.3 *Hilbert Space*

In this subsection we assume that F is a separable Hilbert space. We also assume that $K = S^*S: F \to F$ is a compact operator. Let $\Lambda = F^*$, i.e., one can compute inner products (f, h) for arbitrary elements h, each at cost c.

Let (γ_i, η_i) be eigenpairs of the operator K, $K\eta_i = \gamma_i\eta_i$, where $\gamma_1 \ge \gamma_2 \ge \cdots \ge 0$ and $(\eta_i, \xi_j) = \delta_{i,j}$. In order to solve the problem in the worst case, one has to compute $m^{\text{wor}}(\varepsilon, q)$ inner products,

$$m^{\text{wor}}(\varepsilon, q) = \min \left\{ n : \gamma_{n+1} \le \left(\frac{\varepsilon}{q}\right)^2 \right\}. \tag{5.8}$$

The worse case complexity is then bounded by

$$cm^{\text{wor}}(\varepsilon, q) \le \text{comp}^{\text{wor}}(\varepsilon, q) \le (c+2)m^{\text{wor}}(\varepsilon, q) - 1. \tag{5.9}$$

For large c, $\text{comp}^{\text{wor}}(\varepsilon, q) \simeq cm^{\text{wor}}(\varepsilon, q)$. See Traub and Woźniakowski (1980, Chap. 5).

The average case complexity for $q = +\infty$ was found by Wasilkowski (1986). It depends strictly on the eigenvalues of the correlation operator C_ν of the Gaussian measure $\nu = \mu S^{-1}$, which is given by $C_\nu = SC_\mu S^*$. Let (λ_i, ζ_i) be its eigenpairs, $C_\nu\zeta_i = \lambda_i\zeta_i$, where $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$, and $(\zeta_i, \zeta_j) = \delta_{i,j}$. Of course, $\Sigma_i\lambda_i < +\infty$. In order to solve the average case problem, one has to compute $m^{\text{avg}}(\varepsilon)$ inner products,

$$m^{\text{avg}}(\varepsilon) = \min \left\{ n: \sum_{i=n+1}^{\infty} \lambda_i \le \varepsilon^2 \right\}. \tag{5.10}$$

The average complexity is then bounded by

$$cm^{\text{avg}}(\varepsilon) \le \text{comp}^{\text{avg}}(\varepsilon) \le (c+2)m^{\text{avg}}(\varepsilon) - 1. \tag{5.11}$$

For large c, $\text{comp}^{\text{avg}}(\varepsilon) \simeq cm^{\text{avg}}(\varepsilon)$.

In general, there is no relation between the worst case and average complexities since the eigenvalues $\{\gamma_i\}$ need not be related to the eigenvalues $\{\lambda_i\}$.

To relate the two complexities, assume that the covariance operator C_μ has the same eigenelements as the operator K and that the corresponding eigenvalues are properly ordered, i.e.,

$$\begin{aligned} K\eta_i &= \gamma_i\eta_i, \qquad \gamma_1 \ge \gamma_2 \ge \cdots \ge 0, \\ C_\mu\eta_i &= \beta_i\eta_i, \qquad \beta_1 \ge \beta_2 \ge \cdots \ge 0. \end{aligned} \tag{5.12}$$

This assumption means that the directions of large eigenvalues of K correspond to large weights of the Gaussian measure μ.

From (5.12) we conclude that the nonzero eigenvalues of C_ν are given by $\lambda_i = \gamma_i\beta_i$. Indeed, $C_\nu\zeta_i = \lambda_i\zeta_i$ implies that $C_\mu^{1/2}KC_\mu^{1/2}z_i = \lambda_i z_i$ with $z_i =$

$C_\mu^{1/2}S^*\zeta_i \neq 0$. The operator $C_\mu^{1/2}\mathrm{K}C_\mu^{1/2}$ has eigenvalues $\gamma_i\beta_i$ and therefore $\lambda_i = \gamma_i\beta_i$ as claimed. Using this, we have

$$m^{\text{avg}}(\varepsilon) = \min\left\{n: \sum_{i=n+1}^{\infty} \gamma_i\beta_i \leq \varepsilon^2\right\}. \tag{5.13}$$

We consider two cases of eigenvalues $\{\gamma_i\}$ and $\{\beta_i\}$.

(i) *Polynomial case*. Assume that $\gamma_i = i^{-u}$ for $u > 0$ and $\beta_i = i^{-v}$ for $v > 1$. Then (5.9) yields for large c,

$$\text{comp}^{\text{wor}}(\varepsilon, q) \simeq c\left(\frac{q}{\varepsilon}\right)^{2/u}.$$

From (5.10) with $\lambda_i = i^{-(u+v)}$ and from Theorem 2.1 we conclude that for large c and q, and small ε,

$$\text{comp}^{\text{avg}}(\varepsilon, q) \simeq c(u + v - 1)^{-1/(u+v-1)}\varepsilon^{-2/(u+v-1)}.$$

Thus,

$$\frac{\text{comp}^{\text{wor}}(\varepsilon, q)}{\text{comp}^{\text{avg}}(\varepsilon, q)} \simeq q^{2/u}(u + v - 1)^{1/(u+v-1)}\left(\frac{1}{\varepsilon}\right)^{2(v-1)/u(u+v-1)} \tag{5.14}$$

The ratio goes to infinity as $e \to 0$. The speed of convergence depends on the magnitudes of $v - 1$ and u. For v close to one or for large u, the ratio goes to infinity slowly.

(ii) *Exponential case*. Assume that $\gamma_i = \rho_1^i$ and $\beta_i = \rho_2^i$, where $\rho_1, \rho_2 \in (0, 1)$. Then for large c and q, and small ε

$$\text{comp}^{\text{wor}}(\varepsilon) \simeq c\,\frac{\ln(q/\varepsilon)^2}{\ln(1/\rho_1)},$$

and due to Theorem 2.1,

$$\text{comp}^{\text{avg}}(\varepsilon, q) \simeq c\,\frac{\ln(1/\varepsilon^2) + \ln(1/(1 - \rho_1\rho_2))}{\ln(1/\rho_1\rho_2)}.$$

Thus,

$$\frac{\text{comp}^{\text{wor}}(\varepsilon, q)}{\text{comp}^{\text{avg}}(\varepsilon, q)} \simeq 1 + \frac{\ln(1/\rho_2)}{\ln(1/\rho_1)}.$$

Hence, the worst case is roughly $1 + \ln \rho_2/\ln \rho_1$ times harder than the average one.

6. Normalized and Relative Errors

In this section we indicate briefly how Theorem 2.1 can be used to find bounds on the average complexity for normalized and relative errors.

We first deal with the normalized error, where the distance between the element Sf and the value of the algorithm $\phi(N(f))$ is defined by $\|Sf - \phi(N(f))\|/\|f\|$. The average normalized error is then defined as

$$e^{\text{nor}}(\phi, N) = \left\{\int_F \left(\frac{\|Sf - \phi(N(f))\|}{\|f\|}\right)^2 \mu(df)\right\}^{1/2}. \tag{6.1}$$

Let $\text{comp}^{\text{nor}}(\varepsilon)$ denote the average complexity for the normalized error defined as in (2.7) with $e^{\text{avg}}(\phi, N, q)$ replaced by $e^{\text{nor}}(\phi, N)$.

As in Section 2, let $x = 1 - \mu(B_q)$ with q satisfying (2.11), and let c denote the cost of one information evaluation.

THEOREM 6.1. (i) *Lower Bound:*

$$\text{comp}^{\text{nor}}(\varepsilon) \geq \frac{c}{c+2}(1 - x - \sqrt{3x})\,\text{comp}^{\text{avg}}\left(\frac{q\varepsilon}{\sqrt{1 - x - \sqrt{3x}}}\right).$$

(ii) *Upper Bound: Assume that* $\dim C_\mu(F^*) \geq 3$. Let $p \in (1, \frac{1}{2}\dim C_\mu(F^*))$. Then

$$\text{comp}^{\text{nor}}(\varepsilon) \leq \frac{c+2}{c} \inf_{0 \leq t \leq 1} \max\left\{\frac{1}{1-t}\,\text{comp}^{\text{avg}}\left(\frac{\varepsilon}{c_p}\right), \text{comp}^{\text{avg}}\left(\frac{\varepsilon}{c_p}\sqrt{t}\right)\right\},$$

where $c_p = (\int_F \|f\|^{-2p}\mu(df))^{1/2p}(1/\sqrt{2\pi}\int_{\Re} |t|^{2p/(p-1)}\, e^{-t^2/2}dt)^{(p-1)/2p}$ is finite.

(iii) If $\text{comp}^{\text{avg}}(b\varepsilon) = \Theta(\text{comp}^{\text{avg}}(\varepsilon))$ for all positive b as ε goes to zero, then

$$\text{comp}^{\text{nor}}(\varepsilon) = \Theta(\text{comp}^{\text{avg}}(\varepsilon)).$$

Proof. (i) Take a pair (ϕ, N) such that $e^{\text{nor}}(\phi, N) \leq \varepsilon$ and $\text{cost}^{\text{avg}}(\phi, N) = \text{comp}^{\text{nor}}(\varepsilon)$. Observe that for any positive q,

$$\begin{aligned}
\int_F \left(\frac{\|Sf - \phi(N(f))\|}{\|f\|}\right)^2 \mu(df) &\geq \int_{B_q} \left(\frac{\|Sf - \phi(N(f))\|}{\|f\|}\right)^2 \mu(df) \\
&\geq q^{-2}\mu(B_q)\frac{1}{\mu(B_q)}\int_{B_q} \|Sf - \phi(N(f))\|^2\mu(df) \\
&= q^{-2}(1 - x)e^{\text{avg}}(\phi, N, q)^2.
\end{aligned}$$

This proves that $r^{\text{avg}}(N, q) \le q\varepsilon/\sqrt{1-x}$. Furthermore, $\text{cost}^{\text{avg}}(\phi, N, q) \le (1/(1-x))\,\text{cost}^{\text{avg}}(\phi, N) = (1/(1-x))\,\text{comp}^{\text{nor}}(\varepsilon)$. Thus, the pair (ϕ, N) solves the problem for the ball B_q with $\varepsilon_1 = q\varepsilon/\sqrt{1-x}$. Therefore,

$$\text{comp}^{\text{avg}}(\varepsilon_1, q) \le \frac{1}{1-x}\,\text{comp}^{\text{nor}}(\varepsilon).$$

Applying (2.13) of Theorem 2.1, we get (i).

(ii) In order to prove the upper bound we need two lemmas.

LEMMA 6.1. *Let $p > 0$. If $\dim C_\mu(F^*) > 2p$ then $\int_F \|f\|^{-2p}\mu(df)$ is finite.*

Proof. Take a finite k such that $k \le \dim C_\mu(F^*)$ and $k > 2p$. Then there exist $L_1, L_2, \ldots, L_k$ from F^* such that $L_i(C_\mu L_j) = \delta_{ij}$, $i, j = 1, 2, \ldots, k$. Define $Qf = \sum_{i=1}^k L_i(f)C_\mu L_i$. Then $\|f\| \ge \|Qf\|/\|Q\|$ and

$$u = \int_F \|f\|^{-2p}\mu(df) \le \int_F \left(\frac{\|Qf\|}{\|Q\|}\right)^{-2p}\mu(df) = \|Q\|^{2p}\int_X \|x\|^{-2p}\lambda(dx),$$

where $X = \text{span}(C_\mu L_1, \ldots, C_\mu L_k)$ and $\lambda = \mu Q^{-1}$ is Gaussian with mean zero and the identity correlation operator. Since X is finite dimensional, $\|\cdot\|$ for $x \in X$ is equivalent to the second norm, i.e., there exists a positive constant b such that $\|\sum_{i=1}^k x_i C_\mu L_i\| \ge b\sqrt{\sum_{i=1}^k x_i^2} = b\|x\|_2$. Thus

$$\begin{aligned}\int_X \|x\|^{-2p}\lambda(dx) &\le (2\pi)^{-k/2}b^{-2p}\int_{\Re^k}\|t\|_2^{-2p}e^{-\|t\|_2^2/2}\,dt\\ &\le (2\pi)^{-k/2}b^{-2p}\int_{\|t\|_2\le 1}\|t\|_2^{-2p}\,dt + b^{-2p}.\end{aligned}$$

The last integral $v = \int_{\|t\|_2\le 1}\|t\|^{-2p}\,dt$ is finite iff $k > 2p$ (see Gradshteyn and Ryzhik, 1965, 4.642). For $k > 2p$, $v = 2\sqrt{\pi^k}/(\Gamma(k/2)(k-2p))$. ■

LEMMA 6.2. *Let* $\dim C_\mu(F^*) \ge 3$ *and* $p \in (1, \frac{1}{2}\dim C_\mu(F^*))$. *For nonadaptive information N and the μ-spline algorithm ϕ, $\phi(N(f)) = S\sigma(N(f))$, we have*

$$e^{\text{nor}}(\phi, N) \le c_p r^{\text{avg}}(N)$$

with c_p given in (ii) *of Theorem* 6.1.

Proof. We use the notation of Section 3. We have

$$e^{\text{nor}}(\phi, N)^2 = \sum_i \int_F \frac{(Sf - \phi(N(f)), \zeta_i(N(f)))^2}{\|f\|^2}\,\mu(df)$$

$$\le \left\{\int_F \|f\|^{-2p}\mu(df)\right\}^{1/p} \sum_i \left\{\int_F |(Sf - \phi(N(f)), \zeta_i(N(f)))|^{2p/(p-1)}\mu(df)\right\}^{(p-1)/p}$$

$$= c_p^2 \sum_i \left\{\int_{N(F)} \int_F |(Sf - S\sigma(y), \zeta_i)|^{2p/(p-1)}\mu_2(df|y)\mu_1(dy)\right\}^{(p-1)/p}$$

$$= c_p^2 \sum_i \lambda_i = c_p^2\, r^{\text{avg}}(N)^2,$$

as claimed. ■

We are now ready to prove (ii) of Theorem 6.1. Take information N for which $c\ \text{card}^{\text{avg}}(N) \le \text{comp}^{\text{avg}}(\varepsilon)$ and $r^{\text{avg}}(N) \le \varepsilon$. From Theorem 4.1 of Wasilkowski (1986), there exists permissible information N^* such taht $\text{card}^{\text{avg}}(N^*) \le \text{card}^{\text{avg}}(N)$, $r^{\text{avg}}(N^*) \le r^{\text{avg}}(N)$ and N^* has the form

$$N(f) = \begin{cases} N_1(f) & \text{if } L_1(f) \in A, \\ N_2(f) & \text{otherwise,} \end{cases}$$

where N_1 and N_2 are both nonadaptive, L_1 is the first functional of N_1 and N_2, and the set A from $\mathfrak{R}$ has measure t. Let i_j be the cardinality of N_j. Then $i_1 \le \text{card}^{\text{avg}}(N^*) \le i_2$ and $i_1 t + i_2(1 - t) = \text{card}^{\text{avg}}(N^*)$. Let r_i be the average radius of N_i. Then $r_2 \le r^{\text{avg}}(N^*) \le r_1$ and $\sqrt{tr_1^2 + (t - t)r_2^2} = r^{\text{avg}}(N^*)$. Assume for a moment that $t > 0$.

Now take nonadaptive information N_1. Its average radius is no greater than $r^{\text{avg}}(N^*)/\sqrt{t}$. Apply Lemma 6.2 for information N_1. (Note that c_p is finite due to Lemma 6.1.) Let ϕ_1 denote the μ-spline algorithm using N_1. Then

$$e^{\text{nor}}(\phi_1, N_1) \le c_p\varepsilon/\sqrt{t}.$$

Thus the pair (ϕ_1, N_1) solves the problem for $\varepsilon_1 = \varepsilon c_p/\sqrt{t}$ with the average cost $(c + 2)i_1 - 1 \le (c + 2)\,\text{card}^{\text{avg}}(N^*) \le ((c + 2)/c)\,\text{comp}^{\text{avg}}(\varepsilon)$. Therefore

$$\text{comp}^{\text{nor}}(\varepsilon_1) \le \frac{c + 2}{c}\,\text{comp}^{\text{avg}}\left(\frac{\varepsilon_1}{c_p}\sqrt{t}\right). \tag{6.2}$$

Observe that (6.2) is trivially true for $t = 0$.

Assume now that $t < 1$. Consider nonadaptive information N_2. Since $r_2 \le r^{\text{avg}}(N^*)$, Lemma 6.2 yields for the μ-spline algorithm ϕ_2 using N_2,

$$e^{\text{nor}}(\phi_2, N_2) \le c_p\varepsilon.$$

The pair (ϕ_2, N_2) solves the problem for $\varepsilon_2 = c_p\varepsilon$ with the average cost

$$(c+2)i_2 - 1 \le (c+2)\,\mathrm{card}^{\mathrm{avg}}(N^*)/(1-t) \le \frac{c+2}{c}\frac{1}{1-t}\,\mathrm{comp}^{\mathrm{avg}}(\varepsilon).$$

Therefore

$$\mathrm{comp}^{\mathrm{nor}}(\varepsilon_2) \le \frac{c+2}{c}\frac{1}{1-t}\,\mathrm{comp}^{\mathrm{avg}}\left(\frac{\varepsilon_2}{c_p}\right). \tag{6.3}$$

Note that (6.3) holds trivially for $t = 1$.

We establish (6.2) and (6.3) for arbitrary ε. Therefore (6.2) holds for any ε_1 and (6.3) for any ε_2. From this we conclude that

$$\begin{aligned}\mathrm{comp}^{\mathrm{nor}}(\varepsilon) &\le \frac{c+2}{c}\sup_{0\le t\le 1}\min\left\{\frac{1}{1-t}\,\mathrm{comp}^{\mathrm{avg}}\left(\frac{\varepsilon}{c_p}\right),\ \mathrm{comp}^{\mathrm{avg}}\left(\frac{\varepsilon}{c_p}\sqrt{t}\right)\right\}\\ &= \frac{c+2}{c}\inf_{0\le t\le 1}\max\left\{\frac{1}{1-t}\,\mathrm{comp}^{\mathrm{avg}}\left(\frac{\varepsilon}{c_p}\right),\ \mathrm{comp}^{\mathrm{avg}}\left(\frac{\varepsilon}{c_p}\sqrt{t}\right)\right\},\end{aligned}$$

as claimed.

(iii) This follows easily from (i) and (ii). ∎

We now turn to the relative error, where the distance between Sf and $\phi(N(f))$ is defined by $\|Sf - \phi(N(f))\|/\|Sf\|$. The average relative error is then defined as

$$e^{\mathrm{rel}}(\phi, N) = \left\{\int_F \left(\frac{\|Sf - \phi(N(f))\|}{\|f\|}\right)^2 \mu(df)\right\}^{1/2}. \tag{6.4}$$

Let $\mathrm{comp}^{\mathrm{rel}}(\varepsilon)$ denote the average complexity for the relative error defined as in (2.7) with $e^{\mathrm{avg}}(\phi, N, q)$ replaced by $e^{\mathrm{rel}}(\phi, N)$.

We use the same notation as in Theorem 6.1. Recall that $\nu = \mu S^{-1}$ is a Gaussian measure on the Hilbert space G with mean zero and correlation operator C_ν.

THEOREM 6.2. (i) *Lower bound:*

$$\mathrm{comp}^{\mathrm{rel}}(\varepsilon) \ge \frac{c}{c+2}(1 - x - \sqrt{3x})\,\mathrm{comp}^{\mathrm{avg}}\left(\frac{\varepsilon\|S\|q}{\sqrt{1-x}}\right).$$

(ii) *Upper bound: Assume that* $\dim C_\nu(G) \ge 3$. Let $p \in (1, \frac{1}{2}\dim C_\nu(G))$. Then

$$\mathrm{comp}^{\mathrm{rel}}(\varepsilon) \le \frac{c+2}{c}\inf_{0\le t\le 1}\max\left\{\frac{1}{1-t}\,\mathrm{comp}^{\mathrm{avg}}\left(\frac{\varepsilon}{d_p}\right),\ \mathrm{comp}^{\mathrm{avg}}\left(\frac{\varepsilon}{d_p}\sqrt{t}\right)\right\},$$

where $d_p = (\int_G \|g\|^{-2p}\nu(dg))^{1/2p}(1/\sqrt{2\pi}\int_{\Re}|t|^{2p/(p-1)}e^{-t^2/2}(dt))^{(p-1)/2p}$ is finite.

(iii) If $\mathrm{comp}^{\mathrm{avg}}(b\varepsilon) = \Theta(\mathrm{comp}^{\mathrm{avg}}(\varepsilon))$ for all positive b as ε goes to zero, then

$$\mathrm{comp}^{\mathrm{rel}}(\varepsilon) = \Theta(\mathrm{comp}^{\mathrm{avg}}(\varepsilon)).$$

Proof. The proof is essentially the same as the proof of Theorem 6.1. To get (i) it is enough to note that $\|Sf\|^{-2} \geq \|S\|^{-2}\|f\|^{-2}$. To get (ii), it is enough to note that Lemma 6.2 holds for the relative error with c_p replaced by d_p. The number d_p is finite due to Lemma 6.1. ∎

Theorems 6.1 and 6.2 relate the average complexity for the normalized or relative error to the average complexity for the absolute error. The bounds obtained permit us to find only lower and upper estimates on $\mathrm{comp}^{\mathrm{nor}}(\varepsilon)$ and $\mathrm{comp}^{\mathrm{rel}}(\varepsilon)$. These estimates, in general, need not to be sharp. Tight bounds on $\mathrm{comp}^{\mathrm{nor}}(\varepsilon)$ and $\mathrm{comp}^{\mathrm{rel}}(\varepsilon)$ will require a more refined analysis than has been carried out here.

We now comment on the assumptions $\dim C_\mu(F^*) \geq 3$ and $\dim C_\nu(G) \geq 3$ in (ii) of Theorems 6.1 and 6.2. The assumption $\dim C_\mu(F^*) \geq 3$ states that the Gaussian measure μ is concentrated on a subspace of dimension at least three. If F is a Hilbert space, this means that C_μ has at least three nonzero eigenvalues.

It seems to us that with a proper definition of the space F, we should always consider only such Gaussian measures for which $\dim C_\mu(F^*) = \dim F$ and therefore for all interesting problems, $\dim C_\mu(F^*)$ is large or infinite. Therefore, $\dim C_\mu(F^*) \geq 3$ is not restrictive.

We now discuss the assumption $\dim C_\nu(G) \geq 3$. This assumption has a different flavor than the previous one. Even if one considers F and a Gaussian measure μ for which $\dim C_\mu(F^*) = +\infty$, it can happen that $\dim C_\mu(G) \leq 2$. Indeed, if S is a continuous linear functional or a two-dimensional operator then its range G is at most of dimension 2 and $\dim C_\nu(G) \leq 2$. As in Jackowski and Woźniakowski (1986), it is easy to prove that $\dim C_\nu(G) \leq 2$ implies $\mathrm{comp}^{\mathrm{rel}}(\varepsilon) = +\infty$ for $\varepsilon < 1$.

The assumption $\dim C_\nu(G) \geq 3$ means that we consider the operator S with at least three-dimensional range and with a nongenerate Gaussian measure on that range.

We illustrate Theorems 6.1 and 6.2 by the four examples of Section 5.

Integration. Consider the integration problem of Section 5.1. Since the assumption of (iii) holds, and $\dim C_\nu(F^*) = +\infty$, $\dim C_\nu(G) = 1$, we have

$$\mathrm{comp}^{\mathrm{nor}}(\varepsilon) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r+1)}\right)$$

$$\mathrm{comp}^{\mathrm{rel}}(\varepsilon) = +\infty, \qquad \varepsilon < 1.$$

Approximation. Consider the approximation problem of Section 5.2. The assumption of (iii) holds and dim $C_\mu(F^*) = \dim C_\nu(G) = +\infty$. Hence

$$\text{comp}^{\text{nor}}(\varepsilon) = \Theta(\text{comp}^{\text{rel}}(\varepsilon)) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{1/(r_{\min}+1/2)}\left(\ln\frac{1}{\varepsilon}\right)^{(k_d-1)(1+1/(2r_{\min}+1))}\right).$$

Hilbert Case with Polynomial Distribution. Consider the Hilbert case problem of Section 5.3 with a polynomial distribution of eigenvalues. Then (iii) holds, and $C_\mu(F^*)$ and $C_\nu(G)$ are infinite dimensional. Hence

$$\text{comp}^{\text{nor}}(\varepsilon) = \Theta(\text{comp}^{\text{rel}}(\varepsilon)) = \Theta(c\varepsilon^{-2/(u+v-1)}).$$

Hilbert Case with Exponential Distribution. Consider the previous problem with an exponential distribution of eigenvalues. Then (iii) holds in a stronger form,

$$\text{comp}^{\text{avg}}(b\varepsilon) = \text{comp}^{\text{avg}}(\varepsilon)(1 + o(1))$$

as $\varepsilon \to 0$ and $c \to +\infty$. Therefore for small ε and large c we have

$$\text{comp}^{\text{nor}}(\varepsilon) \cong \text{comp}^{\text{rel}}(\varepsilon) \cong c\,\frac{\ln(1/\varepsilon^2)}{\ln(1/\rho_1\rho_2)}.$$

In this particular case, we have almost exactly obtained the average complexity $\text{comp}^{\text{nor}}(\varepsilon)$ and $\text{comp}^{\text{rel}}(\varepsilon)$.

Acknowledgments

I am grateful for valuable comments from G. W. Wasilkowski and A. G. Werschulz.

References

Araujo, A. and Giné, E. (1980), "The Central Limit Theorem for Real and Banach Valued Random Variables," Wiley, New York.

Borell, C. (1975), The Brunn–Minkowski inequality in Gauss space, *Invent. Math.* **30,** 207–216.

Borell, C. (1976), Gaussian Radon measures on locally convex spaces, *Math. Scand.* **38,** 265–284.

Gradshteyn, I. S., and Ryzhik, I. W. (1965), "Table of Integrals, Series and Products," Academic Press, New York.

Jackowski, T., and Woźniakowski, H. (1986), "Complexity of Approximation with Relative Error Criterion in Worst, Average, and Probabilistic Settings," Department of Computer Science Report, Columbia University.

KUO, M. M. (1975), "Gaussian Measures in Banach Spaces," Lecture Notes in Mathematics, Vol. 463, Springer-Verlag, Berlin.

LEE, D., AND WASILKOWSKI, G. W. (1986), Approximation of linear functionals on a Banach space with a Gaussian measure, *J. Complexity* **2,** 12–43.

PAPAGEORGIOU, A., AND WASILKOWSKI, G. W. (1986), Average complexity of multivariate problems, in progress.

TRAUB, J. F., AND WOŹNIAKOWSKI, H. (1980), "A General Theory of Optimal Algorithms," Academic Press, New York.

VAKHANIA, N. N. (1981), "Probability Distributions on Linear Spaces," North-Holland, New York.

WASILKOWSKI, G. W. (1986), Information of varying cardinality, *J. Complexity* **2,** 204–228.

WOŹNIAKOWSKI, H. (1986), Information-based complexity, *in* "Annual Review of Computer Science," 319–380, Annual Reviews/Kauffman, Palo Alto, CA.