Probabilistic Setting of Information-Based Complexity

H. WOŹNIAKOWSKI

Department of Computer Science
Columbia University
New York, N.Y. 10027

Institute of Informatics
University of Warsaw
Warsaw, Poland

April 1986

CUCS-237-86

ABSTRACT. We study the probabilistic (ε,δ)-complexity for linear problems equipped with Gaussian measures. The probabilistic (ε,δ)-complexity, $\mathrm{comp}^{prob}(\varepsilon,\delta)$, is understood as the minimal cost required to compute approximations with error at most ε on a set of measure at least $1-\delta$. We find estimates of $\mathrm{comp}^{prob}(\varepsilon,\delta)$ in terms of eigenvalues of the correlation operator of the Gaussian measure over elements which we want to approximate. In particular, we study the approximation and integration problems. The approximation problem is studied for functions of d variables which are continuous after r times differentiation with respect to each variable. For the Wiener measure placed on rth derivatives, the probabilistic $\mathrm{comp}^{prob}(\varepsilon,\delta)$ is estimated by $\Theta\Big((\sqrt{2\ln(1/\delta)}/\varepsilon)^{1/(r+a)}\Big(\ln(\sqrt{2\ln(1/\delta)}/\varepsilon)\Big)^{(d-1)(r+1)/(r+a)}\Big)$ where $a=1$ for the lower bound and $a=0.5$ for the upper bound. The integration problem is studied for the same class of functions with $d=1$. In this case, $\mathrm{comp}^{prob}(\varepsilon,\delta)=\Theta\Big((\sqrt{2\ln(1/\delta)}/\varepsilon)^{1/(r+1)}\Big)$.

1. INTRODUCTION

Information-based complexity is the study of the intrinsic difficulty of approximately solved problems. It deals with the *minimal cost* which is required to compute approximations with *error* at most ε, where $\varepsilon \geq 0$. Different settings of information-based complexity are obtained depending on how the error and the cost are defined.

In the *worst case setting*, the error and cost are defined by a hardest element. In the *average case setting*, the error and cost are similarly defined with the exception that the hardest element is replaced by the "average" one, the average being with respect to a given probability measure.

In the *probabilistic setting*, we relax the worst case requirement that the error be at most ε for *all* elements. Instead, we require that the error be at most ε for a set of elements of measure at least $1-\delta$. Here δ is a given parameter, $\delta \in [0,1]$. Thus, in the probabilistic setting we agree that the error can be arbitrarily bad for a set of measure at most δ.

The cost in the probabilistic setting can be defined in various ways. It can be defined as in either the average case or worst case settings. One can also disregard a set of measure

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

at most δ for which the error exceeds ε and then define the cost as in the average case or worst case settings.

The worst and average case settings of information-based complexity have been studied in many papers. A recent survey of these two settings may be found in Woźniakowski (1986).

The probabilistic setting of information-based complexity has been recently studied in Wasilkowski (1984,1986) and in Lee and Wasilkowski (1986). This setting has been analyzed for linear problems and Gaussian measures. The main result can be explained as follows. Let ν be a probability measure of elements which we want to approximate. During the computation we learn more and more about the elements we want to approximate. Mathematically, this means that we change the measure ν. After n evaluations, we have a conditional measure ν_n. Let B_ε be a ball of center zero and radius ε. For large n, the measure ν_n resembles an atomic measure concentrated over the exact solution and therefore for a positive ε, $\nu_n(B_\varepsilon)$ tends to one. For a positive δ, there exists n for which $\nu_n(B_\varepsilon) \geq 1-\delta$. Then we can compute approximations with error at most ε on a set of measure at least $1-\delta$. Under some assumptions, Wasilkowski (1986) found lower and upper bounds on the probabilistic (ε,δ)-complexity in terms of the smallest n for which $\nu_n(B_\varepsilon) \geq 1-\delta$.

In this paper we specialize general estimates on the probabilistic (ε,δ)-complexity for two cases. In the first case, we approximate elements Sf for f from a separable Banach space F equipped with a Gaussian measure μ. Here S is a continuous linear operator into a separable Hilbert space. We assume that arbitrary linear continuous functionals on f can be computed. We find estimates on the probabilistic (ε,δ)-complexity in terms of eigenvalues of the correlation operator C_ν of the Gaussian measure $\nu = \mu S^{-1}$. In particular, for the approximation problem defined as in Papageorgiou and Wasilkowski (1986) for functions of d variables which are continuous after r times differentiation with respect to each variable, see Section 4, and which are equipped with the Wiener measure placed on rth derivatives, we find that the probabilistic (ε,δ)-complexity is at least proportional to $\left(\sqrt{2\ln(1/\delta)}/\varepsilon\right)^{1/(r+1)}\left(\ln\sqrt{2\ln(1/\delta)}/\varepsilon\right)^{(d-1)}$ and is at most proportional to $\left(\sqrt{2\ln(1/\delta)}/\varepsilon\right)^{1/(r+0.5)}\left(\ln\sqrt{2\ln(1/\delta)}/\varepsilon\right)^{(d-1)(r+1)/(r+0.5)}$. This exhibits the dependence of the complexity on ε, δ, the smoothness parameter r and the dimension d.

In the second case, we assume that S is a continuous linear functional, $S \in F^*$, and that only some functionals from F^* can be computed. The dependence on ε,δ is, as in the first case, through $\sqrt{2\ln(1/\delta)}/\varepsilon$. In particular, for the integration problem studied by Lee and Wasilkowski (1986), for the same class of functions as for the approximation problem with $d=1$, the probabilistic (ε,δ)-complexity is proportional to $\left(\sqrt{2\ln(1/\delta)}/\varepsilon\right)^{1/(r+1)}$.

For both cases, information computed about the function is *nonadaptive*. Thus, although we permit *adaptive* information, it turns out it is not essentially more powerful than nonadaptive information. This follows from Wasilkowski (1986) who proved that even in the general case adaptive information is not more powerful than information which is essentially nonadaptive. We do not pursue this subject here.

In this paper we concentrate on estimates on the probabilistic (ε,δ)-complexity and do not discuss optimal algorithms. We mention here that for the approximation problem, such an algorithm is provided by n terms of the truncated series of Sf in the basis of eigenele-

ments of C_ν, where $n = \Theta\big((\sqrt{2\ln(1/\delta)}/\varepsilon)^{1/(r+0.5)}\big(\ln\sqrt{2\ln(1/\delta)}/\varepsilon\big)^{(d-1)(r+1)/(r+0.5)}\big)$. For the integration problem, the algorithm is given by the integral of the natural spline which interpolates f at n equally spaced points, where $n = \Theta\big((\sqrt{2\ln(1/\delta)}/\varepsilon)^{1/(r+1)}\big)$.

We summarize the content of the paper. In Section 2 we precisely formulate the probabilistic setting and the probabilistic (ε,δ)-complexity. In Section 3 we recall general estimates on the probabilistic (ε,δ)-complexity from Wasilkowski (1986). Sections 4 and 5 deal with the two cases described above. In the final section we discuss different definitions of the probabilistic cost, relations between average and probabilistic settings and some open problems.

2. Probabilistic Setting

In this section we define the probabilistic setting of information-based complexity. This setting consists of problem formulation, information and model of computation.

Let F and G be linear normed spaces over the real field. Consider an operator S, called the *solution operator*,

$$S : F \to G. \tag{2.1}$$

Let μ be a probability measure defined on Borel sets of F. Let ε and δ be given, where $\varepsilon \geq 0$ and $0 \leq \delta \leq 1$. Our problem is: for each f from F, compute an element $U(f)$ from G such that

$$\mu\{f \in F \ : \ \|S(f) - U(f)\| \leq \varepsilon\} \geq 1 - \delta. \tag{2.2}$$

That is, we wish to approximate $S(f)$ to within ε with probability at least $1-\delta$.

How can we compute an approximation $U(f)$? We assume that, in addition to knowing the problem formulation (2.1) and (2.2), we can gather information about f by computing $L_i(f)$ for a number of i. Here L_i is a continuous linear functional, $L_i \in F^*$, where F^* is the dual space of F. Let Λ be a class of permissible functionals L. Thus $\Lambda \subset F^*$. For some problems, one can assume that all continuous linear functionals can be computed, i.e., $\Lambda = F^*$. This assumption may be reasonable, for instance, for the approximation problem $S(f) = f$. For other problems, one has to impose some conditions on Λ. For instance, for the integration problem, $S(f) = \int_0^1 f(t)\,dt$, we usually can compute only function evaluations. Then Λ consists of $L(f) = f(x)$ for various x, and Λ is a proper subset of F^*.

The approximation $U(f)$ is computed by combining such information operations, $U(f) = \varphi\big(L_1(f), L_2(f), \ldots, L_n(f)\big)$ where $L_i \in \Lambda$. We stress that the choice of L_i may adaptively depend on the previous values $L_1(f), L_2(f), \ldots, L_{i-1}(f)$. Also the number n of information operations may adaptively depend on the values $L_1(f), L_2(f), \ldots$. Thus, n may vary with f. The precise definition of U may be found in Wasilkowski (1986).

We assume that we are charged for each computation of $L(f)$ and that each $L(f)$ costs c, where $c > 0$. We also assume that we can perform certain operations in the space G, called combinatory operations, such as the addition $g_1 + g_2$ of two elements g_1, g_2 from

G, and multiplication by scalars αg for $\alpha \in \mathbb{R}$ and $g \in G$. The cost of such operations is taken as unity.

The probabilistic setting is formalized as follows.

Problem : For each f in F, compute an element $U(f)$ from G such that

$$\mu\{f \in F \ : \ \|S(f) - U(f)\| \le \varepsilon\} \ge 1 - \delta.$$

Information :

(i) We know the problem formulation, i.e., the solution operator S, the spaces F and G, the probability measure μ on F, the class Λ and the parameters ε and δ.

(ii) We can compute $L(f)$ for any $L \in \Lambda$ and any $f \in F$.

Model of Computation :

(i) Each information operation $L(f)$ costs c.

(ii) We can perform certain combinatory operations at unit cost. Examples of such operations include addition of two elements from G and multiplication by scalars. ∎

We want to compute $U(f)$ with minimal average cost. The cost of computing $U(f)$, $\mathrm{cost}(U, f)$, is defined as the sum of the cost of information operations and the cost of combinatory operations needed to compute $U(f)$. The *average cost* of U, $U : F \to G$, is defined as

$$\mathrm{cost}^{avg}(U) = \int_F \mathrm{cost}(U, f)\, \mu(df). \tag{2.3}$$

See Section 6 where different definitions of the cost of U are discussed.

The probabilistic (ε, δ)-*complexity* is defined as the minimal average cost of U provided that $U(f)$ approximates $S(f)$ to within ε with probability at least $1 - \delta$. That is,

$$\mathrm{comp}^{prob}(\varepsilon, \delta) = \inf\{\mathrm{cost}^{avg}(U) \ : \ U \text{ satisfies (2.2)}\}. \tag{2.4}$$

3. General Case

In this section we briefly recall estimates of the probabilistic (ε, δ)-complexity from Wasilkowski (1986). Then in Section 4 we specialize them to the case where S is an arbitrary linear operator and there is no restriction on the class Λ, $\Lambda = F^*$. In Section 5 we deal with the case where S is a linear continuous functional and Λ is an arbitrary subset of F^*.

Assume that F is a separable Banach space and G is a separable Hilbert space, both over the real field. The space F is equipped with a Gaussian measure μ of mean zero and correlation operator C_μ, $C_\mu : F^* \to F$. The definition and basic properties of Gaussian measures can be found in Kuo (1975), Parthasarathy (1967) and Vakhania (1981).

We assume that the solution operator S, $S : F \to G$, is a continuous linear operator. Then $\nu = \mu S^{-1}$ is a probability measure on the separable Hilbert space G. The measure

ν tells us about the distribution of solution elements Sf. It is a Gaussian measure with mean zero and correlation operator C_ν, $C_\nu : G \to G$, such that

$$C_\nu g = S(C_\mu(L_g S)), \qquad g \in G, \tag{3.1}$$

where $L_g(h) = (h, g)$ for $h \in G$. Note that $L_g S \in F^*$. The correlation operator C_ν is symmetric, nonnegative definite and has a finite trace.

Let $N = [L_1, \ldots, L_n]$, with $L_i \in \Lambda$. Given N, define $K_1, K_2, \ldots, K_m$ such that

$$K_i \in span(L_1, L_2, \ldots, L_n)$$

$$K_i(C_\mu K_j) = \delta_{i,j} \qquad i, j = 1, 2, \ldots, m.$$

Thus $K_1, K_2, \ldots, K_m$ are linear combinations of $L_1, L_2, \ldots, L_n$ which are orthonormalized in the sense $K_i(C_\mu K_j) = \delta_{i,j}$. Here $m \le n$. If $L_1, L_2, \ldots, L_n$ are linearly independent and C_μ is one-to-one then $m = n$.

Define the Gaussian measure ν_N on the space G with mean zero and correlation operator C_N given by

$$C_N g = C_\nu g - \sum_{j=1}^{m} (g_j, g) g_j, \qquad g \in G, \tag{3.2}$$

where $C_\nu$is given by (3.1) and $g_j = S(C_\mu K_j)$.

Let $B_\varepsilon = \{g \in G : ||g|| \le \varepsilon\}$ be the ball of radius ε in G. Let

$$b(n, \varepsilon) = \sup\{\nu_N(B_\varepsilon) : N = [L_1, L_2, \ldots, L_n],\ L_i \in \Lambda\}. \tag{3.3}$$

Thus $b(n, \varepsilon)$ denotes the maximal measure of the ball of radius ε among Gaussian measures ν_N obtained by different choices of n functionals from Λ.

Define the *probabilistic (ε, δ)-cardinality number* as

$$m(\varepsilon, \delta) = \min\{n : b(n, \varepsilon) \ge 1 - \delta\}. \tag{3.4}$$

Thus $m(\varepsilon, \delta)$ denotes the smallest n for which there exists a Gaussian measure generated by n functionals from Λ for which the ball of radius ε has measure at least $1 - \delta$.

Recall that one evaluation of $L(f)$ costs c. We are ready to present estimates of the probabilistic complexity which follow from Wasilkowski (1986, Section 5 with the error functional $E : G \to \mathbb{R}_+$ given by $E(g) = 1$ if $||g|| > \varepsilon$ and $E(g) = 0$ for $||g|| \le \varepsilon$).

THEOREM 3.1. *The probabilistic (ε, δ)-complexity satisfies the inequalites*

$$c \sup_{1 \le x \le 1/\delta} \min\left\{ m(\varepsilon, x\delta)\,,\ \frac{x-1}{x} m(\varepsilon, \delta) \right\} \le \mathrm{comp}^{prob}(\varepsilon, \delta) \le (c+2) m(\varepsilon, \delta) - 1. \quad \blacksquare \tag{3.5}$$

The probabilistic complexity is estimated in terms of the probabilistic cardinality number. In particular, if

$$m\left(\varepsilon\,,\ \delta \ln \frac{1}{\delta}\right) = m(\varepsilon, \delta)\,(1 + o(1)), \qquad \text{as} \quad \delta \to 0, \tag{3.6}$$

then applying (3.5) with $x = \ln(1/\delta)$ we have

$$\mathrm{comp}^{prob}(\varepsilon, \delta) = (c + a)\, m(\varepsilon, \delta)\,(1 + o(1)), \qquad \text{as} \quad \delta \to 0, \tag{3.7}$$

where $a \in [0, 2]$. Since usually $c \gg 1$, we have tight bounds on $\mathrm{comp}^{prob}(\varepsilon, \delta)$. We summarize this in

COROLLARY 3.1. *The probabilistic (ε, δ)-complexity is approximately equal to*

$$\mathrm{comp}^{prob}(\varepsilon, \delta) \approx c\, m(\varepsilon, \delta)$$

whenever $c \gg 1$ and (3.6) holds. ∎

We now indicate how to achieve the upper bound in Theorem 3.1. Assume that (3.3) and (3.4) are achieved for $N_n^* = [L_1^*, L_2^*, \ldots, L_n^*]$ with $n = m(\varepsilon, \delta)$. Obviously, L_i^* are now linearly independent. They can be orthonormalized by the Gram-Schmidt method by taking $K_j^* = \sum_{i=1}^{j} a_{i,j} L_i^*$, the coefficients $a_{i,j}$ being chosen such that $K_j^*(C_\mu K_i^*) = \delta_{i,j}$ for $i, j = 1, \ldots, n$. This corresponds to the solution of n linear equations whose matrix is nonsingular and lower triangular. Define the elements

$$q_i = \sum_{j=i}^{n} a_{i,j} S(C_\mu K_j^*).$$

Note that q_i do not depend on elements f and they can be precomputed. The approximation $U(f)$ is defined by

$$U(f) = \sum_{i=1}^{n} L_i^*(f) q_i. \tag{3.8}$$

The cost of computing $U(f)$ is equal to $n\,c + 2\,n - 1 = (c+2)\,m(\varepsilon, \delta) - 1$. Furthermore

$$\mu\{ f : \|S(f) - U(f)\| \le \varepsilon\} = \nu_{N_n^*}(B_\varepsilon) \ge 1 - \delta.$$

Thus U solves the problem with cost given by the upper bound of Theorem 3.1.

4. $\Lambda = F^*$ AND S ARBITRARY

In this section we specialize estimates of the probabilistic complexity given in Theorem 3.1 for $\Lambda = F^*$ and an arbitrary continuous linear S. To do this, we express the probabilistic cardinality number in terms of eigenvalues of the covariance operator C_ν, see (3.1).

The operator C_ν is symmetric, nonnegative definite and has a finite trace. Therefore there exists an orthonormal basis $\{\zeta_i\}$ of G which consists of eigenelements of C_ν,

$$C_\nu \zeta_i = \lambda_i \zeta_i, \qquad (\zeta_i, \zeta_j) = \delta_{i,j}, \tag{4.1}$$

where $\lambda_1 \ge \lambda_2 \ge \cdots \ge 0$ for $i = 1, 2, \ldots, \dim G$. If $\dim G$ is finite then we formally set $\zeta_i = 0, \lambda_i = 0$ for $i > \dim G$. The trace of C_ν is the sum of λ_i, $\mathrm{trace}(C_\nu) = \sum_{i=1}^{\infty} \lambda_i$.

We need to find the maximal measure of $\nu_N(B_\varepsilon)$, see (3.3). It is known, see Wasilkowski (1984), that $\nu_N(B_\varepsilon)$ is maximized for $N = N_n^*$, where

$$N_n^*(f) = [(Sf, \zeta_1), (Sf, \zeta_2), \ldots, (Sf, \zeta_n)]. \tag{4.2}$$

The measure $\nu_n := \nu_{N_n^\perp}$ is Gaussian with mean zero and correlation operator C_n given by

$$C_n g = C_\nu g - \sum_{i=1}^{n} \lambda_i (g, \zeta_i)\, \zeta_i, \qquad g \in G. \tag{4.3}$$

Thus $C_n g = 0$ for $g \in span(\zeta_1, \zeta_2, \ldots, \zeta_n)$ and $C_n g = C_\nu g$ for g orthogonal to $span(\zeta_1, \zeta_2, \ldots, \zeta_n)$. Obviously, $C_n = C_n^* \geq 0$, $\mathrm{trace}(C_n) = \sum_{i=n+1}^{\infty} \lambda_i$ which goes to zero as n goes to infinity. If $\dim G$ is finite and $n \geq \dim G$, then $C_n = 0$. The measure ν_n is then atomic, i.e., $\nu_n(B) = 1$ if $0 \in B$, and $\nu_n(B) = 0$ if $0 \notin B$.

We now estimate the measure $\nu_n(B_\varepsilon)$ in terms of eigenvalues λ_i. From Vakhania (1981, p.40) we know that there exists a constant c_1 such that

$$\nu_n(B_\varepsilon) \geq 1 - c_1 \exp\left(-\frac{\varepsilon^2}{2\ \mathrm{trace}(C_n)}\right). \tag{4.4}$$

It is easy to show that one can take $c_1 = 5$. Thus to guarantee that $\nu_n(B_\varepsilon) \geq 1 - \delta$, it suffices to take n such that $\mathrm{trace}(C_n) \leq \varepsilon^2/(2 \ln \frac{5}{\delta})$. Thus the upper bound on the probabilistic cardinality number $m(\varepsilon, \delta)$ is given by

$$m_U(\varepsilon, \delta) = \min\left\{ n : \sum_{i=n+1}^{\infty} \lambda_i \leq \frac{\varepsilon^2}{2 \ln \frac{5}{\delta}} \right\}. \tag{4.5}$$

We now obtain an upper bound on $\nu_n(B_\varepsilon)$. Observe that $B_\varepsilon \subset \{g : |(g, \zeta_{n+1})| \leq \varepsilon\}$. Thus

$$\nu_n(B_\varepsilon) \leq \nu_n\{g : |(g, \zeta_{n+1})| \leq \varepsilon\} = \frac{1}{\sqrt{2\pi\lambda_{n+1}}} \int_{|t| \leq \varepsilon} \exp\left(-\frac{t^2}{2\lambda_{n+1}}\right) dt =: \Phi\left(\frac{\varepsilon}{\sqrt{\lambda_{n+1}}}\right),$$

where Φ is the probability integral. Therefore a lower bound of $m(\varepsilon, \delta)$ is given by

$$m_L(\varepsilon, \delta) = \min\left\{ n : \Phi\left(\frac{\varepsilon}{\sqrt{\lambda_{n+1}}}\right) \geq 1 - \delta \right\}. \tag{4.6}$$

For large x, $\Phi(x) = \left(1 - \sqrt{\frac{2}{\pi}} \frac{1}{x} e^{-x^2/2}\right)\left(1 + o(1)\right)$. Thus, for small δ we have

$$m_L(\varepsilon, \delta) = \min\left\{ n : \lambda_{n+1} \leq \frac{\varepsilon^2}{2 \ln \frac{1}{\delta}} \right\}\left(1 + o(1)\right). \tag{4.7}$$

From this and Theorem 3.1 we have

COROLLARY 4.1. *For $\Lambda = F^*$, the probabilistic (ε, δ) complexity is bounded by*

$$c \sup_{1 \leq x \leq 1/\delta} \min\left\{ m_L(\varepsilon, x\delta), \frac{x-1}{x} m_L(\varepsilon, \delta) \right\} \leq \mathrm{comp}^{prob}(\varepsilon, \delta) \leq (c+2)\, m_U(\varepsilon, \delta) - 1. \blacksquare \tag{4.8}$$

As in Section 3, if m_L satisfies (3.6) then the left hand side of (4.8) is approximately equal to $c\, m_L(\varepsilon,\delta)$.The upper bound in (4.8) is achieved for

$$U(f) = \sum_{i=1}^{m_U(\varepsilon,\delta)} (Sf,\zeta_i)\zeta_i.$$

This corresponds to the truncated series of Sf with respect to eigenelements of the covariance operator C_ν of the measure $\nu = \mu S^{-1}$.

REMARK 4.1.

From Wasilkowski (1986) we know that the average ε-complexity of approximating Sf is given by

$$c\, m^{avg}(\varepsilon) \le comp^{avg}(\varepsilon) \le (c+2)\, m^{avg}(\varepsilon) - 1,$$

where $m^{avg}(\varepsilon) = \min\{\, n : \sum_{i=n+1}^{\infty} \lambda_i \le \varepsilon^2 \}$.

Comparing this to (4.8) we see that the upper bound on the probabilistic (ε,δ)-complexity corresponds to the average ε_1-complexity with $\varepsilon_1 = \varepsilon/\sqrt{2\ln(5/\delta)}$. ∎

We illustrate Corollary 4.1 by assuming that $\lambda_k = (ak)^{-p}$ for some positive a and $p > 1, i = 1,2,\ldots$. Then (4.7) yields

$$m_L(\varepsilon,\delta) = \left(\left\lceil \frac{1}{a\eta_1^{1/p}} \right\rceil - 1 \right)\big(1 + o(1)\big), \tag{4.9}$$

where $\eta_1 = \varepsilon^2/(2\ln(1/\delta))$. From (4.5) we have

$$m_U(\varepsilon,\delta) = \left\lceil \frac{1}{a\big[a(p-1)\eta_2\big]^{1/(p-1)}} \right\rceil - x, \qquad \text{for some} \quad x \in [1,2], \tag{4.10}$$

where $\eta_2 = \varepsilon^2/(2\ln(5/\delta))$.

We now consider the approximation problem defined as in Papageorgiou and Wasilkowski (1986). Let $f : D = [0,1]^d \to \mathbb{R}$ 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 x_i, $j = 1,2,\ldots,d$. Let r be a given nonnegative integer.

Define the Banach space F as the class of functions f for which $f^{(r,r,\ldots,r)}$ is continuous and $f^{(i_1,i_2,\ldots,i_d)}(t) = 0, \forall i_j = 0,1,\ldots,r$ and any t for which one of the components is zero. The space F is equipped with the norm $\|f\| = \sup_{t\in D} |f^{(r,r,\ldots,r)}(t)|$ and with the Wiener measure placed on rth derivatives. That is, $\mu(A) = w(D_r A)$, where $D_r f = f^{(r,r,\ldots,r)}$ and w is the classical Wiener measure.

Let $G = L_2(D)$ and let the solution operator S, $S : F \to G$, be given by the embedding $Sf = f$. The eigenvalues λ_k of the measure $\nu = \mu S^{-1}$ are given by

$$\lambda_k = \left(\frac{(\ln k)^{d-1}}{\pi^d (d-1)!\, k} \right)^{2r+2} \big(1 + o(1)\big), \qquad \text{as} \quad k \to \infty.$$

From (4.7) we have

$$m_L(\varepsilon,\delta) = a_1\left(\frac{\sqrt{2\ln(1/\delta)}}{\varepsilon}\right)^{\frac{1}{r+1}}\left(\ln\frac{\sqrt{2\ln(1/\delta)}}{\varepsilon}\right)^{d-1}, \tag{4.11}$$

where

$$a_1 = \left(\pi^d\,(d-1)!\,(r+1)^{d-1}\right)^{-1}(1+o(1)).$$

From (4.5) we have

$$m_U(\varepsilon,\delta) = a_2\left(\frac{\sqrt{2\ln(1/\delta)}}{\varepsilon}\right)^{\frac{1}{r+\frac{1}{2}}}\left(\ln\frac{\sqrt{2\ln(1/\delta)}}{\varepsilon}\right)^{(d-1)\left(1+\frac{1}{2r+1}\right)} \tag{4.12}$$

where

$$a_2 = \left(\pi^d(d-1)!\right)^{-1}\left((2r+1)\pi^d(d-1)!\right)^{-\frac{1}{2r+1}}(r+\frac{1}{2})^{-(d-1)\left(1+\frac{1}{2r+1}\right)}(1+o(1)).$$

Applying Corollary 4.1 with $x = \ln\frac{1}{\delta}$, we have

COROLLARY 4.2. *For $\Lambda = F^*$, the probabilistic (ε,δ)-complexity for the approximation problem is bounded by*

$$a_1\,c\,x^{\frac{1}{r+1}}\,(\ln x)^{d-1} \le \mathrm{comp}^{prob}(\varepsilon,\delta) \le a_2\,(c+2)\,x^{\frac{1}{r+\frac{1}{2}}}\,(\ln x)^{(d-1)\left(1+\frac{1}{2r+1}\right)}, \tag{4.13}$$

where a_1 and a_2 are given above and $x = \sqrt{2\ln(1/\delta)}/\varepsilon$. ∎

Observe the difference in the exponents of x. For the lower bound we have $(r+1)^{-1}$, for the upper bound we have $(r+\frac{1}{2})^{-1}$. This difference is essential for small r. For the extreme case, $r = 0$, the upper bound exponent is twice as big as the lower bound exponent. For large r, the difference between the upper and lower bounds becomes less significant. We believe that the upper bound of Corollary 4.2 is sharp.

REMARK 4.2.

We stress that $\mathrm{comp}^{prob}(\varepsilon,\delta)$ depends on $\sqrt{2\ln(1/\delta)}/\varepsilon$. Thus, it depends on δ through $\ln(1/\delta)$ and it depends on ε through $1/\varepsilon$. The dependence on ε is therefore much more crucial than the dependence on δ. ∎

REMARK 4.3.

The same approximation problem has been studied in the average case setting by Papageorgiou and Wasilkowski (1986). They proved that the average ε-complexity is given by

$$comp^{avg}(\varepsilon) = \Theta\left(c\left(\frac{1}{\varepsilon}\right)^{\frac{1}{r+\frac{1}{2}}}\left(\ln\frac{1}{\varepsilon}\right)^{(d-1)\left(1+\frac{1}{2r+1}\right)}\right).$$

Comparing this to (4.13) we see that the upper bound on the probabilistic (ε, δ)-complexity corresponds to the average ε_1-complexity with $\varepsilon_1 = \varepsilon/\sqrt{2\ln(1/\delta)}$. For further relations between average case and probabilistic settings see Section 6 (ii). ∎

5. Λ ARBITRARY AND $S \in F^*$

In this section we specialize estimates of the probabilistic complexity for an arbitrary class Λ and a continuous linear functional S. Thus, $G = \mathbb{R}$ and the measure ν is a one dimensional Gaussian measure with mean zero. Its covariance operator C_ν is just a number. It is its variance $\mathrm{var}_\nu = C_\nu = S(C_\mu S)$. As in Lee and Wasilkowski (1986), define the semi-inner product and the semi norm

$$\langle L_1, L_2 \rangle_\mu = L_1(C_\mu L_2), \qquad \|L\|_\mu = \langle L, L \rangle_\mu^{1/2} \tag{5.1}$$

for $L_1, L_2, L \in F^*$.

The maximal measure of the ball $B_\varepsilon = [-\varepsilon, +\varepsilon]$ is obtained for the measure ν_n which is a one dimensional Gaussian measure with mean zero and variance var_n,

$$\mathrm{var}_n = \inf\Big\{ \| S - \sum_{i=1}^n a_i L_i \|_\mu^2 \,:\, a_i \in \mathbb{R},\, L_i \in \Lambda \Big\}, \tag{5.2}$$

see Lee and Wasilkowski (1986, Section 2). Thus, var_n is the square of the approximation error of S by a linear combination of n functionals from Λ. We now have, see (3.3),

$$b(n,\varepsilon) = \nu_n(B_\varepsilon) = \frac{1}{\sqrt{2\pi\ \mathrm{var}_n}} \int_{|t| \le \varepsilon} \exp\left(\frac{-t^2}{2\,\mathrm{var}_n}\right) dt \;=\; \Phi\left(\frac{\varepsilon}{\sqrt{\mathrm{var}_n}}\right).$$

The probabilistic (ε, δ)-cardinality number, see (3.4), is now given by

$$m(\varepsilon\,,\,\delta) = \min\Big\{ n \,:\, \Phi\Big(\frac{\varepsilon}{\sqrt{\mathrm{var}_n}}\Big) \ge 1 - \delta \Big\}. \tag{5.3}$$

For small δ, we have

$$m(\varepsilon\,,\,\delta) = \min\left\{ n \,:\, \mathrm{var}_n \le \frac{\varepsilon^2}{2\ln(1/\delta)} \right\} (1 + o(1)). \tag{5.4}$$

We now find $m(\varepsilon, \delta)$ for the integration problem studied by Lee and Wasilkowski (1986). Consider the class F as in Section 4 with $d = 1$. Let $G = \mathbb{R}$ and $Sf = \int_0^1 f(t)\,dt$.

The class Λ consists of function evaluations, i.e., $L(f) = f(x)$ for some x. From Section 5.4 of Lee and Wasilkowski (1986) we get

$$\mathrm{var}_n = \inf\left\{ \int_0^1 \left(\frac{(1-t)^{r+1}}{(r+1)!} - \sum_{i=1}^n a_i\,(t - t_i)_+^r \right)^2 dt \,:\, a_i \in \mathbb{R},\, t_i \in [0,1] \right\}.$$

Sacks and Ylvisaker (1970) showed that

$$\text{var}_n = \Theta\big(n^{-2(r+1)}\big). \tag{5.5}$$

This is achieved by the equally spaced points $t_i = i/(n+1)$, $i = 1,2,\ldots,n$. From (5.4) and (5.5) we conclude that

$$m(\varepsilon,\delta) = \Theta\left(\left(\frac{\sqrt{2\ln(1/\delta)}}{\varepsilon}\right)^{\frac{1}{r+1}}\right). \tag{5.6}$$

The formula (3.8) is now the integral of the natural spline of degree $2r+1$ which interpolates the function values $f\big(i/(n+1)\big)$ for $i = 1,2,\ldots,n$ and $n = m(\varepsilon,\delta)$. Applying Theorem 3.1 and (3.6) we have

COROLLARY 5.1. *The probabilistic (ε,δ)-complexity for the integration problem is given by*

$$\text{comp}^{prob}(\varepsilon,\delta) = \Theta\left(c\left(\frac{\sqrt{2\ln(1/\delta)}}{\varepsilon}\right)^{\frac{1}{r+1}}\right). \tag{5.7}$$

The integral of the natural spline of degree $2r+1$ which interpolates f at $m(\varepsilon,\delta)$ equally spaced points solves the problem with almost minimal cost, where $m(\varepsilon,\delta)$ is given by (5.6). ∎

REMARK 5.1. From Lee and Wasilkowski (1986) we know that the average ε-complexity of the integration problem is given by

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

Comparing this to (5.7) we note that the probabilistic (ε,δ)-complexity corresponds to the average ε_1-complexity with $\varepsilon_1 = \varepsilon/\sqrt{2\ln(1/\delta)}$, see also Section 6 (ii). ∎

6. FINAL REMARKS

In this section we briefly discuss different definitions of the cost in the probabilistic setting, relations between average case and probabilistic settings as well as some open problems.

(i) *Different Definitions of Cost.*

We defined $\text{cost}^{avg}(U)$ by the average value of $\text{cost}(U,f)$ over all elements from F, see (2.3). One may define the $\text{cost}(U)$ as in the worst case setting. That is,

$$\text{cost}(U) = \sup_{f\in F} \text{cost}(U,f). \tag{6.1}$$

Let $\mathrm{comp}_1^{prob}(\varepsilon,\delta)$ denote the probabilistic (ε,δ)-complexity with cost given by (6.1). Then it is easy to show that

$$c\,m(\varepsilon,\delta) \le \mathrm{comp}_1^{prob}(\varepsilon,\delta) \le (c+2)\,m(\varepsilon,\delta) - 1, \tag{6.2}$$

where $m(\varepsilon,\delta)$ is given by (3.4). For $c \gg 1$, we have

$$\mathrm{comp}_1^{prob}(\varepsilon,\delta) \approx c\,m(\varepsilon,\delta).$$

One can modify the definitions (2.3) and (6.1) by disregarding a set of measure at most δ for which $U(f)$ does not approximate $S(f)$ with error at most ε. That is,

$$\mathrm{cost}^{avg}(U) = \frac{1}{1-\delta}\,\inf\Big\{ \int_A \mathrm{cost}(U,f)\,\mu(df) \;:\; A \text{ such that } \mu(A) \ge 1-\delta \text{ and } \|Sf - U(f)\| \le \varepsilon, \text{ for all } f \in A \Big\}. \tag{6.3}$$

Let $\mathrm{comp}_2^{prob}(\varepsilon,\delta)$ denote the probabilistic (ε,δ)-complexity with cost given by (6.3). For $\delta < \frac{1}{2}$, it is easy to show that

$$\mathrm{comp}^{prob}(\varepsilon,\delta_1) \le \mathrm{comp}_2^{prob}(\varepsilon,\delta) \le \frac{1}{1-\delta_1}\,\mathrm{comp}^{prob}(\varepsilon,\delta).$$

where $1-\delta_1 = (1-2\delta)/(1-\delta)$. For small δ, $\delta_1 = \delta\,(1+o(1))$ and therefore $\mathrm{comp}_2^{prob}(\varepsilon,\delta)$ is practically equal to comp^{prob}.

One may also disregard a set of measure at most δ in the supremum (6.1) as in (6.3). Then it is easy to see that the corresponding (ε,δ)-complexity will be the same as $\mathrm{comp}_1^{prob}(\varepsilon,\delta)$.

(ii) *Average Case and Probabilistic Settings.*

In this paper we indicated that the probabilistic (ε,δ)-complexity, or its upper bound, is related to the average ε_1-complexity with $\varepsilon_1 = \varepsilon/\sqrt{2\ln(1/\delta)}$. This holds for linear problems with Gaussian measures for the two cases studied in Sections 4 and 5. For small δ, ε_1 is smaller than ε. Thus, the probabilistic (ε,δ)-complexity is larger than the average ε-complexity. For arbitrary measures, it is not in general true that the probabilistic (ε,δ)-complexity with small δ is larger then the average ε-complexity. In fact, it may happen that the average ε-complexity is infinity whereas the probabilistic (ε,δ)-complexity is finite for positive ε and δ. If, however, the average ε-complexity is finite then Chebyshev's inequality yields

$$\mathrm{comp}^{prob}(\varepsilon,\delta) \le comp^{avg}(\varepsilon_1) \tag{6.4}$$

with $\varepsilon_1 = \sqrt{\delta}\,\varepsilon$. Relations between average case and probabilistic settings seem to be an interesting subject of the future research.

(iii) *Open Problems.*

We indicate a few open problems in the probabilistic setting. It will be interesting to improve estimates of Section 4. In particular, better estimates of Gaussian measures of balls would be welcome.

The probabilistic setting should also be analyzed for more general linear problems, i.e., for S whose range is not necessarily a Hilbert space, and for nonlinear problems, i.e., for nonlinear operators S. It will also be interesting to study more general measures and error criteria. For instance, it seems worthwhile to analyze the relative error in which (2.2) is replaced by

$$\mu\left\{ f \in F : \frac{||Sf - U(f)||}{||Sf||} \leq \varepsilon \right\} \geq 1 - \delta$$

or by

$$\mu\left\{ f \in F : \frac{||Sf - U(f)||}{||f||} \leq \varepsilon \right\} \geq 1 - \delta.$$

Here $||f||$ is a norm of the space F and S is a linear or nonlinear operator.

ACKNOWLEDGMENTS

I am grateful to A. Papageorgiou, J.F. Traub, G.W. Wasilkowski and A.G. Werschulz for valuable comments on this paper.

REFERENCES

Kuo, H.H., "Gaussian Measures in Banach Spaces, Lectures Notes in Mathematics **463**", Springer Verlag, Berlin, 1975.

Lee, D. and Wasilkowski, G. W., *Approximation of Linear Functionals On a Banach Space With a Gaussian Measure*, J.of Complexity **2** (1986), 12-43.

Papageorgiou, A. and Wasilkowski, G. W., *Average Case Complexity for Multivariate Problems, in progress..*

Parthasarathy, K. R., "Probability Measures On Metric Spaces", Academic Press, New York, 1967.

Sacks, J. and Ylvisaker, D., *Statistical Designs and Integral Approximation*, Proc. of the 12th Biennial Seminar of the Canadian Math. Congress (1970), 115-136.

Vakhania, N. N., "Probability Distributions on Linear Spaces", North Holland, New York, 1981.

Wasilkowski, G. W., *Optimal Algorithms for Linear Problems with Gaussian Measures*, Dept. of Computer Science, Columbia University, 1984. To appear in Rocky Mountain J. of Math.

Wasilkowski, G. W., *Information of Varying Cardinality*, to appear in J. of Complexity **2** (1986).

Woźniakowski, H., *Information-Based Complexity*, to appear in, "Annual Review of Computer Science", Annual Reviews Inc., William Kaufman, Inc., (1986).