RECENT DEVELOPMENTS IN INFORMATION-BASED COMPLEXITY

EDWARD W. PACKEL AND HENRYK WOŹNIAKOWSKI

Contents

I. Introduction. The notion of complexity, from both practical and theoretical standpoints, seems destined to be a major theme of research in both computer science and mathematics. As digital computers evolve, we find

ourselves revising and refining our ideas about what it means to solve a problem and what problems can be effectively solved. One might expect the seemingly endless increases in computational speed and memory size to diminish the urgency of such computational matters, but the opposite seems to be the case. Humankind, ever inquisitive and acquisitive, apparently reacts to such technological breakthroughs by asking questions such as "What new problems do we need to solve?", "Can we solve even larger instances of the problem?", and "Can we improve the accuracy of computed solutions?"

As a consequence, mathematics (and theoretical computer science) has been enriched by an important, vigorous, and relatively new subject area, that of computational complexity. Instead of viewing a problem in terms of finding and analyzing a particular algorithm to solve it, computational complexity addresses the inherent computational characteristics of the general problem (as a function of its "size" or the "error" of computed solutions). One seeks upper bounds (which emerge from looking at specific algorithms) and the often more important and difficult lower bounds on the complexity of problems.

While the early emphasis in computational complexity is commonly associated with discrete problems (for example, traveling salesperson, bin packing, and primality testing), there is also substantial literature starting with the work of Sard [1949], Nikolskij [1950], and Kiefer [1953] on the complexity of continuous problems (see Traub and Woźniakowski [1980, pp. 278–280] for a brief history). More recently there has been considerable interest among mathematicians in the complexity of continuous problems, sparked by the work of Shub and Smale [1985, 1986a] and especially Smale [1985], which contains extensive references to other such work.

Many problems arising in the mathematical sciences have the characteristic that information relevant to their solution is either partial or contaminated. For example, since a digital computer can only manipulate a finite set of numbers, any problem whose domain of possible problem elements is infinite-dimensional will of necessity have only partial information. For most problems with partial or contaminated information only approximate solutions are possible. Given the ubiquity of problems for which information is either not fully available, imprecise due to computational limitations, or purposely discarded to expedite a simplified solution, a general approach to treating approximate solution of such problems is clearly desirable.

The issues raised in the preceding paragraphs provide motivation for *information-based complexity*, an expanding research area concerned with the intrinsic difficulty of approximately solving problems based upon information that is partial, contaminated, and priced. In this article we present a mathematically oriented introduction to information-based complexity and a review of some of its current research results and directions. Our selection of topics will not be fully representative, relegating many important applications and technical details to bibliographic references. Instead we stress aspects of the theory that will highlight its elegance and its unifying connections with modern analysis. To drop a few names, we shall encounter problems and techniques from functional analysis, numerical analysis, approximation theory, measure theory, probability theory, statistics, and partial differential equations.

Our primarily theoretical approach will first discuss the worst case setting. This will be followed by a more elaborate treatment of average case complexity, an increasingly important focus of recent work on problems with either partial or complete information. The presentation has been organized to highlight what we regard as the major theorems of the subject. Interspersed throughout the paper and in the concluding section we also list ten open problems from information-based complexity that we think will be of interest to mathematicians.

For more details and references relating to work in information-based complexity up to 1980, see Traub and Woźniakowski [1980], and also Micchelli and Rivlin [1977], where a similar point of view is presented. A framework more general than the one we develop here can be found in Traub, Wasilkowski, and Woźniakowski [1983]. A recent survey of the field with somewhat less emphasis on mathematical abstraction and more on applications and examples is offered in Woźniakowski [1986a].

II. Worst case setting. Three concepts provide the foundation for developing and applying techniques of information-based complexity: a *problem formulation*, specification of the *information*, and a *model of computation*. Each concept has several options or variations, leading to a variety of settings in which problems can be modeled and results developed. We direct our efforts in this part to the traditional *worst case* setting.

A. *Problem formulation, information, and model of computation.* The *problem formulation* states what we want to approximate, for which problem elements we seek this approximation, and what error criterion we plan to use. Formalizing these ideas, let F be a set and let G be a normed linear space over the scalar field K, where $K = \mathbb{R}$ or $K = \mathbb{C}$. The problem we would ideally like to solve is given by a mapping $S: F \rightarrow G$, which we call the *solution operator*. Given $\varepsilon \geqslant 0$, we wish to compute for each $f \in F$ an *ε-approximation $x(f)$* in G for $S(f)$. In the worst case setting we simply require that $x(f)$ satisfy $\|S(f) - x(f)\| \leqslant \varepsilon$ for each problem element f.

In order to approximate $S(f)$ we need to know something about f. A basic assumption of information-based complexity is that we do not have full knowledge of a given problem element f. Formally, knowledge about f is obtained by computations of the form $L(f)$, where $L: F \rightarrow H$ for some set H. The allowable *information* for a problem is given by specifying a collection Λ from which information operations L may be chosen. Generally, for each problem element f, only a finite number of operations L from Λ may be used to supply information $L(f)$ on f. The number of such operations is, of course, crucial to complexity considerations.

If, for each f in F, we use the same n operations from Λ, the resulting *information operator* $N = [L_1, \ldots, L_n]$ is called *nonadaptive* information of cardinality n. Along with its obvious simplicity, such information would be natural and efficient for use in parallel processing. The more general class of *adaptive* information allows the specific choice of operations from Λ and the number $n(f)$ of such operations to vary with f. This means that the choice of L_i may adaptively depend on the previously computed values $L_1(f)$,

$L_2(f), \ldots, L_{i-1}(f)$. In this case the information operator is defined for each f in F by

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

where $y_1 = L_1(f)$, $y_i = L_i(f; y_1, \ldots, y_{i-1})$ for $i = 2, 3, \ldots, n(f) - 1$, and $L_i(\blacksquare; y_1, \ldots, y_{i-1}) \in \Lambda$ for every i. The number $n(f)$ of evaluations is determined by $y_1, y_2, \ldots$. For a precise definition of $n(f)$ see Wasilkowski [1986].

Any analysis of complexity requires a *model of computation* which specifies what operations are permitted and how much each operation "costs." We use what is referred to as the real number model of computation. Thus we assume that

■ Infinite-precision real numbers are used.

■ Each information operation costs c units.

■ Arithmetic operations (such as addition and scalar multiplication in G), comparisons, and evaluation of elementary functions are performed exactly and cost one unit each.

While this model is clearly oversimplified, we use it here to avoid being sidetracked by such admittedly important issues as round-off, machine specificity, and numerical stability. More general and realistic models of computation are certainly desirable for future work in information-based complexity, but we say no more about these here.

B. *An integration example.* To illustrate some of the above ideas we consider the problem of approximating the integral $\int_a^b f(t)\, dt$, where f is chosen from a suitably defined class F of smooth functions. Assuming the existence of a subroutine that computes $f(t)$ for any $t \in [a, b]$, we want to approximate the above integral with minimal cost. Formally we have:

PROBLEM FORMULATION. The solution operator $S : F \to \mathbb{R}$ is defined by

$$S(f) = \int_a^b f(t)\, dt.$$

For each $f \in F$ we wish to compute an ε-approximation. In this case we seek a real number $x(f)$ such that

$$\left| \int_a^b f(t)\, dt - x(f) \right| \leqslant \varepsilon.$$

INFORMATION. Information operations are taken from the class Λ of function evaluations. Thus for any $f \in F$ and any $t \in [a, b]$ we can compute $f(t)$.

MODEL OF COMPUTATION. Each function evaluation costs c, $c > 0$. Arithmetic operations, comparisons, and elementary function evaluations can be performed at a cost of one unit each and with infinite precision (real number model of computation).

Anticipating a forthcoming general definition, we let comp(ε) denote the minimal cost of computing ε-approximations $x(f)$ for the "worst" $f \in F$. In addition to obtaining a value for comp(ε), we would naturally like to know at what points f should be evaluated and how these values should be combined to compute an ε-approximation $x(f)$. In other words (and again anticipating

definitions from a subsequent section), we would like *optimal information* and an *optimal algorithm* for the problem at hand.

We mention here the results for one particular class F consisting of periodic functions on the interval $[0, 2\pi]$ whose $(r - 1)$st derivatives are absolutely continuous and whose rth derivatives are bounded by 1 in the L^∞ norm. Based on results of Bakhvalov [1971] and Motornyj [1973] it turns out that comp$(\varepsilon) = (c + 1)\lceil(2\pi K_r/\varepsilon)\rceil + a_\varepsilon$, where K_r is the Favard constant,

$$K_r = \frac{4}{\pi} \sum_{i=0}^{\infty} \frac{(-1)^{i(r+1)}}{(2i + 1)^{r+1}} \qquad (K_r \in [1, 2\pi]),$$

and $a_\varepsilon = -1$ or 0.

Thus, for the domain F specified above, we know comp(ε) to within the cost of one arithmetic operation. Letting $n = \lceil(2\pi K_r/\varepsilon)^{1/r}\rceil$, an optimal ε-complexity algorithm for estimating the desired integral is given by the composite midpoint rule:

$$x(f) = \frac{2\pi}{n} \sum_{i=1}^{n} f\left(\frac{2\pi(i - 1)}{n}\right).$$

Consequently, values of f at n "equally spaced" points turn out to be optimal ε-complexity information.

We stress that these results depend strongly on the periodicity of functions in the class F. For other specifications of the problem domain F, a variety of different and more complex formulas are required for optimality (see Traub and Woźniakowski [1980], which cites about 100 papers giving optimal formulas and complexity results for various classes F in the worst case setting). We return to a variant of this example when we consider the average case setting.

C. *The radius and diameter of information.* Returning to the general case, let a problem be specified by a solution operator $S: F \to G$ and let N be an information operator on F supplying information $N(f)$ for each f in F. For any such f let $y = N(f)$ and consider the set $S(N^{-1}(y))$ in the normed linear space G. Using the natural definition of the radius of a subset A of G, rad$(A) = \inf_{x \in G} \sup_{a \in A} \|x - a\|$, we define the worst case *radius of information* for the problem S with information N by

$$r(N) = \sup_{y \in N(F)} \{\text{rad}(S(N^{-1}(y)))\}.$$

We now assume for what follows that the infimum in computing rad$(S(N^{-1}(y)))$ is always attained. Then rad$(S(N^{-1}(y)))$ is the radius of the smallest ball in G containing the set of solutions to elements indistinguishable from f using the information operator N. Thus $r(N)$, which plays a crucial role in information-based complexity, gives the intrinsic worst case error or uncertainty present in solving the problem specified by S with information N. In terms of the idea of an ε-approximation introduced earlier, it is immediate that an ε-approximation using the information operator N can be found for every $f \in F$ if and only if N is such that $r(N) \leqslant \varepsilon$.

Another useful way to measure intrinsic uncertainty of a problem and its information operator N is the *diameter of information*, $d(N)$, defined by $d(N) = \sup_{y \in N(F)} \{\text{diam}(S(N^{-1}(y)))\}$, using the standard definition of the diameter of a subset in a normed linear space. Certain results in the theory are more easily developed in terms of the diameter, which is not generally twice the radius, but can easily be shown to satisfy $r(N) \leqslant d(N) \leqslant 2r(N)$.

D. *Algorithms, complexity, and optimality.* Given a problem $S : F \to G$ with information operator N on F, an (idealized) *algorithm* is a function $\Phi : N(F) \to G$. Thus for each $f \in F$, Φ uses what we know about f to approximate $S(f)$ via $\Phi(N(f))$. The worst case error for Φ is then defined by

$$e(\Phi, N) = \sup_{f \in F} \| S(f) - \Phi(N(f)) \|.$$

From the definition of $r(N)$ given above, we immediately obtain:

THEOREM 1. $r(N) = \inf\{ e(\Phi, N) : \Phi$ *an algorithm using information* $N \}$.

Thus the radius of information gives us a sharp lower bound on the error obtained in approximating S by any algorithm using the given information. It is natural, then, to define Φ to be an *optimal error algorithm* for the problem S with information N if $e(\Phi, N) = r(N)$.

Given an algorithm Φ and a model of computation, it is just a matter of bookkeeping to determine the cost of Φ. Given $f \in F$, let cost(N, f) denote the *information cost* of computing $N(f)$. Recalling that c gives the cost of each information operation L involved in computing $N(f)$ and letting $n(f)$ denote the number of such operations, we have cost$(N, f) \geqslant cn(f)$. In the nonadaptive case equality holds; for adaptive information there will be additional cost for the selection of the L_i and, in general, cost$(N, f) > cn(f)$. To use a specific algorithm Φ on $y = N(f)$ there will be a *combinatory cost*, cost(Φ, y), of computing $\Phi(y)$. Using our model of computation, cost$(\Phi, y) = k$, where k is the number of combinatory operations required. The *worst case cost* of an algorithm Φ using information N can now be defined by

$$\text{cost}(\Phi, N) = \sup_{f \in F} \{\text{cost}(N, f) + \text{cost}(\Phi, N(f))\}.$$

We are now ready to define our fundamental notion of complexity in the worst case setting. Given an error tolerance $\varepsilon \geqslant 0$ for a problem $S : F \to G$, we define the *ε-complexity* as the minimal cost of computing an ε-approximation,

$$\text{comp}(\varepsilon) = \inf\{\text{cost}(\Phi, N) : \Phi, N \text{ such that } e(\Phi, N) \leqslant \varepsilon\}.$$

An information operator N and an algorithm Φ using N for which the above infimum is obtained are defined to be *optimal ε-complexity information* and an *optimal ε-complexity algorithm*, respectively. Since ε is regarded as fixed, we generally omit "ε-complexity" in future reference to these ideas.

With the exception of our admittedly specialized model of computation, the definition of complexity given above is very general, and intentionally so. It makes no ad hoc assumptions or restrictions on the kinds of information and algorithms that can compete for the designation of optimal. While this generality may be questioned on practical grounds, it makes any theoretical results we

obtain that much stronger. The definition can be applied with $\varepsilon = 0$ if an exact solution is called for.

Given a problem which can be modeled in the framework we have described, a major focus of "applied" information-based complexity is to find its ε-complexity together with optimal information and an optimal algorithm. In this article we shall be more concerned with general results of the theory rather than application to particular problems.

III. Linear problems in a worst case setting. We now define and develop results for the important class of linear problems. While this narrows our scope significantly, problems of integration, linear differential and integral equations, approximation, and interpolation are linear, as are many problems stemming from linear models in a variety of fields.

A. *Definition and a basic lemma.* To define a linear problem we require that:

■ The problem domain F is a convex balanced subset of a linear space F_1 over K. When convenient we shall assume that F is generated by a linear *restriction operator* $T: F_1 \to X$ where X is a normed linear space and $F = \{ f \in F_1 : \|Tf\| \leq 1 \}$. The requirement that the convex balanced set F is generated by T is not a serious loss of generality (see Traub and Woźniakowski [1980, p. 32]).

■ The solution operator is defined by a *linear* operator $S: F_1 \to G$. Note that while we are only really interested in problem elements from F, it is convenient to be able to work within the linear space F_1.

■ The collection Λ of allowable information operations must be a subset of the space of *linear functionals* $L: F \to K$.

To illustrate the above ideas, we note that our integration example from Part II concerning integration of periodic functions is a linear problem. The domain F is the convex balanced set generated by the linear restriction operator T defined by $Tf = f^{(r)}$. The linearity of the integral ensures that the solution operator S is linear. And the linear functionals L providing information come from the collection Λ of function evaluations (each fixed $t \in [0, 2\pi]$ provides $L_t \in \Lambda$ defined by $L_t(f) = f(t)$).

The following basic result highlights the importance of the kernel of the information operator for linear problems with nonadaptive information.

LEMMA 1. *Given a linear problem defined by a linear operator $S: F_1 \to G$, a nonadaptive linear information operator $N: F_1 \to K^n$, and a linear restriction operator $T: F_1 \to X$ for which $F = \{ f \in F_1 : \|T(f)\| \leq 1 \}$, then*
  (a) $r(N) < \infty \Rightarrow \ker(N) \cap \ker(T) \subseteq \ker(S)$.
  (b) $d(N) = 2 \sup_{h \in \ker(N) \cap F} \{ \|S(h)\| \}$.

PROOF. (a) If there exists h in $\ker(N) \cap \ker(T)$ with $S(h) \neq 0$, then $\alpha h \in \ker(N) \cap F$ for all $\alpha \in K$. It follows that $S(N^{-1})(0)$ is unbounded in G and hence has infinite radius.

(b) Given any y in $N(F)$ and g_1, g_2 in F with $N(g_1) = N(g_2) = y$, set $h = (g_1 - g_2)/2$ and note that $h \in \ker(N) \cap F$ since F is balanced and convex. Thus, $\|S(g_1) - S(g_2)\| = 2\|S(h)\| \leq 2 \sup_{h \in \ker(N) \cap F} \{ \|S(h)\| \}$. Taking the supremum first over $g_1, g_2 \in N^{-1}(y)$ and then over $y \in N(F)$, we get

$d(N) \leqslant 2 \sup_{h \in \ker(N) \cap F} \{\|S(h)\|\}$. The reverse inequality follows by noting that for any $h \in \ker(N) \cap F$ we also have $-h \in \ker(N) \cap F$, so that $2\|S(h)\| = \|S(h) - S(-h)\| \leqslant d(N)$.

Since problems with infinite radius of information are of no great interest to us (there will always be infinite error yet every algorithm is optimal), we henceforth assume $r(N) < \infty$. Thanks to (a) of the Lemma, we will then always have $\ker(N) \cap \ker(T) \subseteq \ker(S)$.

B. *Adaptive vs. nonadaptive information.* It is not surprising that for general problems adaptive information can be significantly more powerful than non-adaptive information. For instance, finding a root for continuous functions on a closed interval $[a, b]$ (with oppositely signed values at the endpoints) by the well-known adaptive *bisection algorithm* yields a radius of information proportional to $1/2^n$, where n is the number of function evaluations (i.e., the cardinality of the information on f). The best nonadaptive methods give a radius proportional to $1/n$. This problem is studied in full generality in Sikorski [1982].

It is both significant and surprising that for *linear* problems in the worst case setting, adaption is not substantially more powerful than nonadaption. Indeed, let N^a be a linear adaptive information operator. For any f in F we then have

$$N^a(f) = \left[L_1(f), L_2(f; y_1), \ldots, L_{n(f)}(f; y_1, \ldots, y_{n(f)-1})\right],$$

where each $L_i(\blacksquare; y_1, \ldots, y_{i-1})$ is a linear functional from Λ. Define a corresponding nonadaptive information operator N^{non} by setting

$$N^{\mathrm{non}}(f) = \left[L_1(f), L_2(f; 0), \ldots, L_{n(0)}(f; 0, 0, \ldots, 0)\right],$$

where $n(0)$ is the cardinality of N^a applied to the problem element 0. Note that N^{non} is linear and nonadaptive since the linear functionals defining it are independent of problem elements f.

THEOREM 2. *For any linear problem in the worst case setting we have*
(a) $d(N^{\mathrm{non}}) \leqslant d(N^a)$
(b) $r(N^{\mathrm{non}}) \leqslant 2r(N^a)$.

PROOF. (a) We use Lemma 1 and the fact that $\ker(N^{\mathrm{non}}) = \ker(N^a)$ to obtain

$$d(N^{\mathrm{non}}) = 2\sup\left\{\|S(h)\| : h \in \ker(N^{\mathrm{non}}) \cap F\right\}$$

$$= 2\sup\left\{\|S(h)\| : h \in \ker(N^a) \cap F\right\}$$

$$= \mathrm{diam}\left(S\left[(N^a)^{-1}(0) \cap F\right]\right) \leqslant \sup_{y \in N(F)}\left\{\mathrm{diam}\left(S\left[(N^a)^{-1}(y) \cap F\right]\right)\right\}$$

$$= d(N^a).$$

(b) This follows from (a) and the fact that $r(N) \leqslant d(N) \leqslant 2r(N)$ for any information operator N.

The first version of this theorem was proved by Bakhvalov [1971] for the case where the solution operator is a linear functional and the information N is of fixed cardinality $n = n(f)$. For a general solution operator and fixed cardinality see Gal and Micchelli [1980] and Traub and Woźniakowski [1980]. The general case is due to Wasilkowski [1986a], who also provides results in the average case setting.

The theorem tells us that, in the worst case, the far more general structure of adaptive information cannot decrease the uncertainty by more than a factor of two as compared to nonadaptive information of the same cardinality. We note that (b) can be strengthened to $r(N^{\mathrm{non}}) = r(N^a)$ in cases where the radius of information is half of its diameter. While the radius and diameter do not always have this relationship, no linear problem (and information) has been found for which the adaptive radius is less than the nonadaptive radius.

OPEN PROBLEM 1. Prove that $r(N^a) = r(N^{\mathrm{non}})$ for all linear problems or find a linear problem with $r(N^a) < r(N^{\mathrm{non}})$.

With the above as justification, we restrict ourselves to nonadaptive information for the duration of our discussion of linear problems in the worst case setting.

C. *The existence of linear optimal error algorithms.* Given a linear problem $S: F \to G$ with nonadaptive information defined by $N: F \to K^n$ with $N = [L_1, \ldots, L_n]$, it is natural to consider *linear* algorithms for approximating S. Accordingly, we require a linear algorithm $\Phi: K^n \to G$ to have the form

$$\Phi(N(f)) = \sum_{i=1}^{n} L_i(f) g_i, \qquad g_i \in G.$$

Since the g_i are independent of f, they can be precomputed. Thus, in addition to their simplicity and ease of implementation, linear algorithms have combinatory cost proportional to n (n scalar multiplications and $n - 1$ additions in G).

Having indicated that linear algorithms have pleasing properties as far as cost is concerned, we now consider the important matter of whether a linear problem will always have an *optimal* error algorithm which is linear—i.e., given S and N linear, does there exist a linear Φ such that $e(\Phi, N) = r(N)$? This question has an interesting history, beginning with the observation that numerous linear problems of practical importance turn out to have linear optimal error algorithms (see Traub and Woźniakowski [1980] for specifics). Below we outline some of the positive and negative results (for the worst case setting) that have recently been obtained (see Packel [1986b] for more detail).

THEOREM 3. *Given F convex and balanced, N linear, and $S: F_1 \to K$ a linear functional, then there exists a linear optimal error algorithm.*

PROOF. The real ($K = \mathbb{R}$) case is due to Smolyak [1965] and the complex case to Osipenko [1976]. The proof for $\mathbb{R}$ uses the convexity of F and a separating hyperplane theorem to produce a linear optimal algorithm. The real case is extended by Micchelli and Rivlin [1977] to the case of "perturbed" information ($N(f)$ is only known to within a certain error bound). This is done by creating a corresponding problem with "exact" information.

THEOREM 4. *Let* $F = \{ f \in F_1 : \|T(f)\| \leqslant 1 \}$ *be generated by a linear restriction operator* $T : F_1 \to H$ *where* H *is a* **Hilbert space** *and let* $T(\ker(N))$ *be closed in* H. *Then a linear problem determined by* S *and* N *on* F *has a linear optimal error algorithm.*

PROOF. Micchelli and Rivlin [1977] treats the special case where T is the identity operator. In Traub and Woźniakowski [1980, Chapter 4] the general result is presented. The argument uses the fact that a closed convex set in Hilbert space has a unique element of smallest norm and makes heavy use of orthogonality arguments. The resulting linear optimal error algorithm turns out to be a *spline* algorithm and bears witness to the close connection between the notion of optimal error algorithms and the theory of approximation by splines (see Atteia [1965], Holmes [1972]).

Any hope of a fully general result was laid to rest by Micchelli in 1978. His example (which can be found in Traub and Woźniakowski [1980, p. 60]) provides a linear problem from $\mathbb{R}^3$ to $\mathbb{R}^2$ endowed with the $\mathbf{1}^4$ norm for which no optimal linear error algorithm exists. Somewhat simpler counterexamples are presented in Packel [1986a] and Packel [1986b].

In each of the examples referred to in the above paragraph it turns out that there exists a linear algorithm Φ whose error is not appreciably larger than the radius of information, i.e., $e(\Phi, N)/r(N) \leqslant 2$. This suggests the following question: "Does there exist a constant d (hopefully "close" to 1) such that

$$\inf\{ e(\Phi, N)/r(N) : \Phi \text{ linear on } N(F) \} \leqslant d$$

for every linear problem?" Werschulz and Woźniakowski [1986] answers this question in the negative by exhibiting a class of linear problems whose radii $r(N)$ are finite but for which the error of any linear algorithm is infinite. Furthermore, the information N can be chosen to make $r(N)$ arbitrarily small. As an added bonus, a special case of these linear problems (and the example which motivated the discovery of the class of counterexamples) is the inversion of a finite Laplace transform, a problem arising in remote sensing (see Twomey [1977]).

Thus there are linear problems whose best linear algorithms do very badly, and such problems are not merely artificial constructs. The now questionable intuition that general linear problems *ought* to have linear optimal error algorithms can be partially resurrected by allowing an extended codomain for the solution operator (and its approximating algorithms). In what follows we use the notation $\mathbf{B}(X)$ to denote the bounded, scalar-valued functions on a topological space X.

THEOREM 5. *Given a linear problem* $S : F \to G$, *there exists*

(i) *A compact Hausdorff space* X *such that* G *is isometrically isomorphic to a subspace of* $\mathbf{B}(X)$.

(ii) *A linear optimal error algorithm* $\Phi : N(F) \to \mathbf{B}(X)$ *satisfying*

$$\|\Phi(N(f)) - {}^{\wedge}(S(f))\| \leqslant r(N)$$

for all f *in* F, *where* ${}^{\wedge}(S(F))$ *denotes the isometric image of* $S(f)$ *in* $\mathbf{B}(X)$.

PROOF. See Packel [1986a] for details. Part (i) follows immediately from a standard corollary to the Banach-Alaoglu theorem stating that any normed

linear space is isometrically isomorphic to a subspace of some $C(X)$, the bounded continuous functions on X. In our case X is the unit ball in the conjugate space of G endowed with the weak* topology and the isometric action is provided by the *Gelfand map* which imbeds G in its second conjugate space. The optimal error algorithm promised in part (ii) is obtained by applying Theorem 3 for each fixed $x \in X$ and showing that the linear algorithm that results when x is varied is bounded on X for each $N(f)$. Rather than having its value restricted to G, this algorithm takes on values in the vastly larger space $B(X)$ containing a copy of G as a subspace.

Thus there is a real but highly impractical sense in which linear problems do have linear optimal error algorithms. We believe the full story has yet to unfold on this general topic.

OPEN PROBLEM 2. Prove that, perhaps with additional conditions on S, linear optimal error algorithms can be found with range restricted to the space $C(X)$.

OPEN PROBLEM 3. Without allowing extended codomain, find some more general conditions than those of Theorem 3 and 4 under which linear problems must have linear optimal error algorithms.

D. *ε-complexity and optimal information for linear problems.* With Theorem 2 as our justification, we consider linear problems with nonadaptive information $N = [L_1, \ldots, L_n]$, where the L_i are chosen from a class Λ of linear functionals on F. For information of a fixed cardinality n, we would like to choose N to minimize both the cost of an algorithm using N and the radius of information for N. Under our real number model of computation (and assuming each linear functional evaluation has constant cost), the cost of a *linear* algorithm depends only on n. Accordingly, we concentrate on minimizing $r(N)$.

For each fixed positive integer n, define the *nth minimal radius of information* $r(n) = r(N_n^*)$ by

$$r(n) = \inf\{r(N): N = [L_1, \ldots, L_n], \ L_i \ \text{in} \Lambda\}.$$

An information operator N_n^*, should one exist, for which the infimum is attained is called *nth optimal error information.* Such information will ensure a minimum intrinsic error $r(n)$ for the given solution operator and fixed cardinality of information n.

To illustrate the above we consider a linear problem $S: F_1 \to G$ where G is a Hilbert space. Let the problem domain $F \subseteq F_1$ be generated by a bijective restriction operator $T: F_1 \to H$, where H is also a Hilbert space (the setting could be generalized somewhat, but we opt for simplicity here). Let Λ be the class of *all* linear functionals on F_1 and assume that the positive operator $A: H \to H$ defined by $A = (ST^{-1})^*ST^{-1}$ is a compact operator. Letting λ_i and x_i denote the ith eigenvalue-orthonormal eigenvector pair for A ($Ax_i = \lambda_i x_i$) ordered so that $\lambda_1 \geqslant \lambda_2 \geqslant \cdots$, the following pleasing result emerges.

THEOREM 6. *In the above Hilbert space setting, nth optimal information is given by* $N_n^*(f) = [\langle Tf, x_1 \rangle, \langle Tf, x_2 \rangle, \ldots, \langle Tf, x_n \rangle]$. *The unique optimal error algorithm using this information is given by*

$$\Phi^{\mathrm{opt}}\big(N_n^*(f)\big) = \sum_{i=1}^{n} \langle Tf, x_i \rangle ST^{-1}x_i,$$

with error given by

$$e\left(\Phi^{\mathrm{opt}}, N_n^*\right) = r(n) = \sqrt{\lambda_{n+1}}\,.$$

PROOF. The full development for a more general result along these lines can be found in Traub and Woźniakowski [1980]. The key to the argument in our case is to expand Tf as its "Fourier series," $Tf = \sum_{i=1}^{\infty} \langle Tf, x_i \rangle x_i$ and to observe that

$$Sf - \Phi^{\mathrm{opt}}\left(N_n^*(f)\right) = ST^{-1}Tf - \Phi^{\mathrm{opt}}\left(N_n^*(f)\right) = \sum_{i=n+1}^{\infty} \langle Tf, x_i \rangle ST^{-1}x_i.$$

By standard orthogonality arguments and the ordering of the λ_i, it follows that N_n^* minimizes the norm of the above quantity. Furthermore, the worst case f is given by $f = T^{-1}x_{n+1}$, yielding a squared error of $\langle ST^{-1}x_{n+1}, ST^{-1}x_{n+1} \rangle = \langle Ax_{n+1}, x_{n+1} \rangle = \lambda_{n+1}\langle x_{n+1}, x_{n+1} \rangle = \lambda_{n+1}$. We note further that, thanks to the compactness of A, the errors approach 0 as the cardinality of information n increases. As indicated in the "proof" of Theorem 4 the optimal error algorithm is an approximation by splines, a lovely connection with the theory of splines which we briefly elaborate upon in Part V.

Returning to the matter of ε-complexity for general linear problems, we will keep our error within ε by using nth optimal error information N_n^* with

$$n = m(\varepsilon) \equiv \min\{\text{cardinality of } N : r(N) \leqslant \varepsilon\}.$$

From this we conclude that it is necessary to compute at least $m(\varepsilon)$ evaluations, each with cost c, in order to compute an ε-approximation. Thus comp$(\varepsilon) \geqslant cm(\varepsilon)$. If, on the other hand, there exists an optimal error algorithm Φ using nth optimal information N_n^* whose combinatory cost is dominated by information cost (i.e., $e(\Phi, N_n^*) = r(N_n^*)$ and

$$\mathrm{cost}\left(\Phi, N_n^*(f)\right) \ll \mathrm{cost}\left(N_n^*, f\right)\right),$$

then comp$(\varepsilon) \approx cm(\varepsilon)$. This holds if there exists a linear optimal error algorithm using N_n^*. Indeed, in this case we have comp$(\varepsilon) \leqslant (c + 2)m(\varepsilon) - 1$. In the common situation where $c \gg 1$, we then get comp$(\varepsilon) \approx cm(\varepsilon)$.

For the linear problems mentioned in §C that do not have linear optimal error algorithms, one can find nonlinear optimal error algorithms whose combinatory cost is essentially less than the information cost whenever $c \gg 1$. Thus comp$(\varepsilon) \approx cm(\varepsilon)$ also holds for such linear problems. We believe that this is true in general and propose the following open problem.

OPEN PROBLEM 4. For the general linear problem (or significant subclasses thereof) and for $c \gg 1$, prove that comp$(\varepsilon) \approx cm(\varepsilon)$.

In the Hilbert space example just considered it follows that $m(\varepsilon) = \min\{n : \sqrt{\lambda_{n+1}} \leqslant \varepsilon\}$. We can thus conclude that the information $N_{m(\varepsilon)}^*$ and the algorithm Φ^{opt} from the example are *almost optimal*, with comp$(\varepsilon) \approx \mathrm{cost}(\Phi^{\mathrm{opt}}, N_{m(\varepsilon)}^*) \approx cm(\varepsilon)$. By choosing the operators S and T so that λ_i goes to zero arbitrarily slowly, we can make $m(\varepsilon)$ blow up arbitrarily fast as ε approaches zero. Thus there exist linear problems with arbitrarily large complexity. It can further be shown that there are no "gaps" in the complexity

functions (Traub and Woźniakowski [1980, Chapter 5]) in the sense that for any increasing function g one can find a linear problem for which $comp(\varepsilon) \approx cg(1/\varepsilon)$ for small positive ε. This provides an interesting contrast with the theory of recursively computable functions, where complexity gaps are known to occur (Borodin [1972]).

IV. Average case setting. We now take a view which requires more optimism (the worst is not expected) and more prior knowledge (a probability distribution on problem elements can be assumed). In the *average case setting* we replace worst case error and cost by corresponding average case formulations. We sidestep significant practical issues dealing with the appropriateness of an average case model and how a specific probability distribution might be obtained for a problem. We note from the references below, however, that an average case approach to algorithmic analysis has gained considerable attention in theoretical computer science and is attracting increasing attention from mathematicians as well. The reader will find our treatment of this material, much of which is very recent, to be somewhat more technical and detailed than the worst case development.

A. *Historical summary.* Many important discrete and continuous problems defined on finite-dimensional spaces have been analyzed "on the average." These problems have been studied assuming complete information and atomic or weighted Lebesgue measure. A partial list of papers include Karp [1976, 1979, 1980], Karp and Luby [1983, 1985], and Rabin [1976, 1983, 1986] for discrete problems; Blum and Shub [1986] for the evaluation of rational functions; Renegar [1984], Shub and Smale [1985, 1986a], and Smale [1981, 1985] for polynomial zero finding; and Adler and Megiddo [1985], Adler, Karp, and Shamir [1983], and Smale [1983a, 1983b, 1985] for linear programming.

As far as we know, the first paper dealing with the average case in what can be regarded as an information-based setting is due to Suldin [1959, 1960], who studied the integration problem for the class of continuous functions equipped with the classical Wiener measure. Larkin, in a series of pioneering papers commencing with Larkin [1972], studied the approximation of linear problems using a Gaussian measure. Both Suldin and Larkin restricted themselves to linear algorithms using nonadaptive information.

There is also an interesting stream of work in the statistical literature dealing with the approximation of linear functionals defined on function spaces. Linear algorithms are primarily studied and information, assumed to be nonadaptive, consists of function and derivative evaluations. In such cases, full knowledge of a measure on the problem elements is not needed. It is enough to know the mean and correlation operator of the measure. A partial list of references includes Kimeldorf and Wahba [1970a, 1970b], Sacks and Ylvisaker [1970], Wahba [1971, 1978], and Ylvisaker [1975]. Relations between Bayesian statistics and average case information-based complexity are explored in Kadane and Wasilkowski [1985].

The average case setting with adaptive linear information and unrestricted classes of algorithms is currently under intense investigation. Parts IV and V of this paper are devoted to an exposition of selected recent results.

B. *Basic formulation*. As before we want to approximate $S(f)$ for problem elements f chosen from a set F. The first issue faced in developing an average case model is that of assuming a probability measure μ on the set F. Using statistical language, we interpret μ as an *a priori* measure. It represents our belief about the distribution of problem elements f. Assuming that the solution operator $S: F \to G$ is measurable, the probability measure $\nu = \mu S^{-1}$ is an a priori measure of solution elements Sf. Both μ and ν represent distributions which are known before any information about problem elements f has been computed.

If F is a subset of k-dimensional Euclidean space $\mathbb{R}^k$, a natural choice for μ is a weighted Lebesgue measure, $\mu(A) = \int_A p(f) \, df$, where A is a Borel subset of $\mathbb{R}^k$. Here $p: \mathbb{R}^k \to \mathbb{R}_+$ is a density function.

Since many problems encountered in information-based complexity are defined on subsets of infinite-dimensional spaces, the choice of μ is generally not so obvious. Indeed, there exist no Lebesgue-type measures on infinite-dimensional spaces. In considering "infinite-dimensional" measures for such problems we propose that (at least theoretically) Wiener measures and, more generally, Gaussian measures serve as good candidates for the average case setting.

Assuming that a probability measure μ on F has been chosen, we proceed as follows. Let N be an information operator as defined in §A of Part II and let $\Phi: N(F) \to G$ be an algorithm using N. The *average error* of Φ is defined by

$$e^{\text{ave}}(\Phi, N) = \left\{ \int_F \|S(f) - \Phi(N(f))\|^p \mu(df) \right\}^{1/p},$$

where $p \in [1, \infty)$. Here we assume that Φ is chosen so that the above integrand is measurable. We also assume that

$$\int_F \|S(F)\|^p \mu(df) < \infty.$$

The *average cost* of Φ is given by

$$\text{cost}^{\text{ave}}(\Phi, N) = \left\{ \int_F [\text{cost}(N, f) + \text{cost}(\Phi, N(f))]^p \mu(df) \right\}^{1/p},$$

where, as defined earlier, $\text{cost}(N, f)$ is the information cost of computing $N(f)$ and cost $(\Phi, N(f))$ is the combinatory cost of computing $\Phi(N(f))$.

Given $\varepsilon \geqslant 0$, the *average ε-complexity* can now be defined as

$$\text{comp}^{\text{ave}}(\varepsilon) = \inf\{\text{cost}^{\text{ave}}(\Phi, N): \Phi, N \text{ such that } e^{\text{ave}}(\Phi, N) \leqslant \varepsilon\}.$$

Thus the average ε-complexity is the minimal average cost of computing approximations whose average error does not exceed ε. A survey of recent results for the average case setting can be found in Wasilkowski [1985].

C. *Average radius of information*. Our definition of the average radius of information will follow that of Wasilkowski [1983]. Given a problem element f and its computed information $y = N(f)$, we again concentrate on $SN^{-1}(y)$, the set of indistinguishable solution elements with respect to a given solution operator $S: F \to G$. In the worst case setting it was natural to define the

(worst case) radius of a subset A of the normed space G by $\text{rad}(A) = \inf_{x \in G} \sup_{a \in A} \|x - a\|$. For the average case setting we would like to replace the maximal distance of $\|x - a\|$ by its average value. Assuming for the moment that an appropriate probability measure ν_A is available on A, the average radius of A is defined by

$$\text{rad}^{\text{ave}}(A) = \inf_{x \in G} \left\{ \int_A \|x - a\|^p \nu_A(da) \right\}^{1/p}.$$

It is evident that $\text{rad}^{\text{ave}}(A) < \text{rad}(A)$.

With the set $SN^{-1}(y)$ playing the role of A, what measure should be chosen to represent the distribution of solution elements from $SN^{-1}(y)$? Clearly this measure must depend on the a priori measure μ over F, on the information operator N, and on the value $y = N(f)$. To get such a measure we assume that N is a measurable mapping. Then $\mu_1 = \mu N^{-1}$ is well defined and is a probability measure on the Borel sets of $N(F)$. For a measurable set $A \subseteq N(F)$, $\mu_1(A)$ gives the probability that information N takes values from A.

We now assume that there exists a unique (modulo sets of μ_1 measure zero) family of conditional probability measures $\mu_2(\blacksquare | y)$ defined on Borel sets of F such that

(i) $\mu_2(N^{-1}(y) | y) = 1$ for almost all $y \in N(F)$.

(ii) $\mu_2(B | \blacksquare)$ is μ_1-measurable for any Borel set B of F.

(iii) $\mu(B) = \int_{N(F)} \mu_2(B | y) \mu_1(dy)$ $\forall$ Borel sets B of F.

Such a family μ_2 will exist, for instance, if F is a measurable subset of a separable Banach space (see Parthasarathy [1967, p. 147]).

Property (i) requires that $\mu_2(\blacksquare | y)$ be concentrated on the set $N^{-1}(y)$ of problem elements yielding the same information. Property (ii) ensures that the integral in (iii) is well defined. The essence of (iii) is that for any measurable function $h : F \to \mathbb{R}_+$ we have

$$\int_F h(f) \mu(df) = \int_{N(F)} \left\{ \int_{N^{-1}(y)} h(f) \mu_2(df | y) \right\} \mu_1(dy).$$

Thus, we can first integrate over elements that share the same information and then over all information values y.

We are now ready to define the measure $\nu(\blacksquare | y)$ for the set $SN^{-1}(y)$. For each measurable set C of G, let

$$\nu(C | y) = \mu_2(S^{-1}(C) | y).$$

As a result of (i), $\nu(C | y) = \mu_2(S^{-1}(C) \cap N^{-1}(y) | y)$ and $\nu(SN^{-1}(y) | y) = 1$. Thus, $\nu(\blacksquare | y)$ is a probability measure on $SN^{-1}(y)$.

Using statistical language, we may say that the measure $\nu(\blacksquare | y)$ is an a posteriori measure that represents our belief about the distribution of solution elements which are indistinguishable after the information $y = N(f)$ has been computed.

The average radius of $SN^{-1}(y)$ can now be defined using ν by

$$\operatorname{rad}^{\text{ave}}\left(SN^{-1}(y)\right) = \inf_{x \in G} \left\{ \int_{SN^{-1}(y)} \|x - a\|^p \nu(da\,|\,y) \right\}^{1/p}.$$

The (global) *average radius of information* is then defined as the average value of $\operatorname{rad}^{\text{ave}}(SN^{-1})$ with respect to the measure μ_1. Thus,

$$r^{\text{ave}}(N) = \left\{ \int_{N(F)} \operatorname{rad}^{\text{ave}}\left(SN^{-1}(y)\right)^p \mu_1(dy) \right\}^{1/p}.$$

It follows immediately that $r^{\text{ave}}(N) \leqslant r(N)$.

We now show that the average radius of information is a sharp lower bound on the average error of any algorithm using N. Indeed, from (iii) the error of an algorithm Φ can be expressed as

$$\left[e^{\text{ave}}(\Phi, N) \right]^p = \int_{N(F)} \left\{ \int_{N^{-1}(y)} \|Sf - \Phi(y)\|^p \mu_2(df\,|\,y) \right\} \mu_1(dy).$$

Changing variables via $a = Sf$, we get

$$\left[e^{\text{ave}}(\Phi, N) \right]^p = \int_{N(F)} \left\{ \int_{SN^{-1}(y)} \|a - \Phi(y)\|^p \nu(da\,|\,y) \right\} \mu_1(dy)$$

$$\geqslant \int_{N(F)} \left\{ \inf_{x \in G} \int_{SN^{-1}(y)} \|a - x\|^p \nu(da\,|\,y) \right\} \mu_1(dy)$$

$$= \int_{N(F)} \operatorname{rad}^{\text{ave}}\left(SN^{-1}(y)\right)^p \mu_1(dy)$$

$$= \left[r^{\text{ave}}(N) \right]^p.$$

The single inequality above suggests that to construct an optimal error algorithm Φ we should define $\Phi(y)$ for each $y \in N(F)$ as follows. Choose $\Phi(y) = x$ where $x \in G$ enables the infimum to be attained:

$$(\text{I}) \quad \int_{SN^{-1}(y)} \|a - \Phi(y)\|^p \nu(da\,|\,y) = \inf_{x \in G} \int_{SN^{-1}(y)} \|a - x\|^p \nu(da\,|\,y).$$

If the infimum cannot be attained, $\Phi(y)$ should be chosen so that a value that is "close" to the infimum results. We have thus proved the following theorem, first presented in Wasilkowski [1983].

THEOREM 7. *The average radius of information is a sharp lower bound on the average error of any algorithm*:

$$r^{\text{ave}}(N) = \inf\left\{ e^{\text{ave}}(\Phi, N) \colon \Phi \text{ any algorithm using } N \right\}.$$

Thus we can find an average case ε-approximation based on N if and only if $r^{\text{ave}}(N) \leqslant \varepsilon$ (modulo a technical assumption that the infimum is attained).

D. *Optimal error algorithms.* In the previous section the conceptual construction of an optimal error algorithm was given. We now specialize this construction by assuming additional structure on the problem domain F and/or its codomain G. For simplicity we take $p = 2$ in the definition of the average error of an algorithm.

Assume that G is a Hilbert space over $\mathbb{R}$ with inner product denoted by $\langle \cdot , \cdot \rangle$. We define the *mean element* $m_y \in G$ of an a posteriori measure $\nu(\blacksquare \mid y)$ by noting that

$$\langle m_y, x \rangle = \int_{SN^{-1}(y)} \langle x, a \rangle \nu(da \mid y)$$

defines a bounded linear functional on G and hence a unique element $m_y \in G$. Recalling equation (I) above (with $p = 2$) and expanding $\|a - x\|^2 = \langle a - x, a - x \rangle$, we obtain

$$\int_{SN^{-1}(y)} \|a - x\|^2 \nu(da \mid y) = \int_{SN^{-1}(y)} \|a\|^2 \nu(da \mid y) - 2\langle m_y, x \rangle + \|x\|^2$$

$$= \int_{SN^{-1}(y)} \|a\|^2 \nu(da \mid y) - \|m_y\|^2 + \|x - m_y\|^2.$$

It follows by inspection that a minimum (with respect to x) results precisely when $x = m_y$, so $\Phi(y) = m_y$ minimizes the average error. Thus, choosing the mean of the a posteriori measure $\nu(\blacksquare \mid y)$ defines an optimal error algorithm. This approach was used by Wasilkowski [1983] to prove the following theorem.

THEOREM 8. *Let $S: F \to G$ with G a Hilbert space. Let μ be a probability measure on F with respect to which S and the information N are measurable. Let μ_2 be a family of conditional probability measures satisfying (i)–(iii) above. Then an optimal error algorithm Φ is provided by defining, for each $y \in N(F)$, $\Phi(y) = m_y$ where m_y is the mean of the conditional probability measure $\nu(\blacksquare \mid y)$ obtained from μ_2 by $\nu(C \mid y) = \mu_2(S^{-1}(C) \mid y)$:*

$$\langle m_y, x \rangle = \int_{SN^{-1}(y)} \langle x, a \rangle \nu(da \mid y).$$

In general it is hard to compute the mean m_y of an a posteriori measure. However, if F is a separable Banach space equipped with a Gaussian measure μ, the mean m_μ can be obtained in a relatively simple way. Since an essential role in the average case setting is played by Gaussian measures, we briefly recall their definition and some key properties (see Kuo [1975], Skorohod [1974], and Vakhania [1981]).

In the case where $F = \mathbb{R}^k$, k-dimensional Euclidean space, a measure μ is Gaussian if, for any Borel set A of $\mathbb{R}^k$,

$$\mu(A) = \frac{1}{(2\pi \det C_\mu)^{k/2}} \int_A \exp\left(-1/2 \langle C_\mu^{-1}(x - m_\mu), x - m_\mu \rangle\right) dx,$$

where $m_\mu \in \mathbb{R}^k$ and C_μ is a k by k positive definite matrix.

Now let F be a separable Banach space and μ a probability measure on F. Then μ is *Gaussian* iff μL^{-1} is a one-dimensional Gaussian measure for every continuous linear functional $L \in F^*$.

Gaussian measures can also be defined by using their characteristic functionals. The characteristic functional $\Psi_\mu : F^* \to \mathbb{C}$ of a probability measure μ on the Borel sets of F is defined by

$$\Psi_\mu(L) = \int_F e^{iL(f)} \mu(df), \qquad i = \sqrt{-1}.$$

A measure μ will be Gaussian if and only if its characteristic functional is of the form

$$\Psi_\mu(L) = \exp\left(iL(m_\mu) - \tfrac{1}{2}L(C_\mu L)\right)$$

for some linear operator $C_\mu: F^* \to F$ and some $m_\mu \in F$. The element m_μ is called the mean element of μ and is defined by the property

$$L(m_\mu) = \int_F L(F)\,\mu(df) \quad \forall L \in F^*.$$

The operator C_μ is called the *correlation operator* of μ and is defined by the property

$$L_1(C_\mu L_2) = \int_F L_1(f - m_\mu)L_2(f - m_\mu)\,\mu(df) \quad \forall L_1, L_2 \in F^*.$$

It can be checked that the correlation operator C_μ is symmetric ($L_1(C_\mu L_2) = L_2(C_\mu L_1)\ \forall L_1, L_2 \in F^*$) and nonnegative definite ($L(C_\mu L) \geqslant 0\ \forall L \in F^*$). If F is a separable Hilbert space then C_μ is fully characterized by being symmetric, nonnegative definite, and having a finite trace. For separable Banach spaces the complete characterization of correlation operators for Gaussian measures remains open.

E. *An integration example for the average case.* To illustrate the above ideas, we build upon the example introduced in Part II, §B. This time we take as our domain F the class $\mathbf{C}_0^r[0, 1]$ of real-valued functions on $[0, 1]$ that are r times continuously differentiable with $f^{(i)}(0) = 0$, $i = 1, 2, \ldots, r$. We give $\mathbf{C}_0^r[0, 1]$ the *sup* norm. The solution operator S is defined by $S(f) = \int_0^1 f(t)\,dt$.

For the measure underlying an average case analysis we let μ be the Wiener measure applied to rth derivatives,

$$\mu(A) = w\left(\{f^{(r)} : f \in A\}\right), \qquad A \text{ a Borel subset of } \mathbf{C}_0^r[0, 1],$$

where w is the standard Wiener measure (a special case of a Gaussian measure) on $\mathbf{C}[0, 1]$. We seek an ε-approximation using function evaluations (f and its first r derivatives) as our set Λ of allowable information operations. Based on results to be stated subsequently (Wasilkowski [1986a]), we may restrict ourselves to nonadaptive information

$$N(f) = \left[f(t_1), \ldots, f^{(k_1)}(t_1), \ldots, f(t_p), \ldots, f^{(k_p)}(t_p)\right],$$

where $0 \leqslant t_1 < t_2 < \cdots < t_p \leqslant 1$, $k_i \leqslant r$, and $n = \mathrm{card}(N) = k_1 + \cdots + k_p$ are fixed.

Results for this linear problem (see Lee and Wasilkowski [1986]) yield an estimate* $r^{\mathrm{ave}}(N) = \Omega(n^{-(r+1)})$ with this bound achievable from information consisting of equally spaced function evaluations:

$$N_n(f) = \left[f(1/(n+1)), f(2/(n+1)), \ldots, f(n/(n+1))\right].$$

---

*We say $g(n) = \Omega(h(n))$ if there exist positive constants c and C with $ch(n) \leqslant g(n)$ for $n > C$. By $g(n) = \Theta(h(n))$ we mean $g(n) = \Omega(h(n))$ and $h(n) = \Omega(g(n))$, i.e., there exist positive constants c_1, c_2, and C with $c_1 h(n) \leqslant g(n) \leqslant c_2 h(n)$ for $n > C$. An analogous definition applies when the parameter $\varepsilon \to 0$ replaces $n \to \infty$.

Furthermore, an optimal error algorithm for this information is the μ-spline algorithm Φ defined by $\Phi(N_n(f)) = S\sigma(N_n(f))$, where σ is the natural spline of degree $2r + 1$. Finally, an estimate for the average ε-complexity is $\text{comp}^{\text{ave}}(\varepsilon) = \Theta((1/\varepsilon)^{1/(r+1)})$. Thus we know the average complexity up to a constant.

V. Linear problems in an average case setting. Continuing our quest for an optimal error algorithm, we now assume that the solution operator $S : F \to G$ is linear and continuous with F a separable Banach space and G a separable Hilbert space. Let μ be a Gaussian measure defined on the Borel sets of F and let C_μ be its correlation operator. We assume, without loss of generality, that the mean element of μ is zero. Then the a priori measure $\nu = \mu S^{-1}$ on G is also Gaussian with mean zero and correlation operator $C_\nu : G^* \to G$ satisfying $C_\nu L = S(C_\mu(LS)) \, \forall L \in G^*$.

A. *Radius of information and an optimal error algorithm.* Given nonadaptive information $N(f) = [L_1(f), \ldots, L_n(f)]$ with $L_i \in F^*$, we may assume with no loss of generality that the L_i are orthonormalized so that $L_i(C_\mu L_j) = \delta_{ij}$. Then (Lee and Wasilkowski [1986]) $\mu_1 = \mu N^{-1}$ is a Gaussian measure defined on the Borel sets of $\mathbb{R}^n$ with mean zero and the identity as its correlation operator (matrix). Thus,

$$\mu_1(A) = \frac{1}{(2\pi)^{n/2}} \int_A e^{\sum_{j=1}^n t_j^2/2} \, dt_1 \, dt_2 \cdots dt_n.$$

The conditional measure $\mu_2(\blacksquare \,|\, y)$ on F whose existence was asserted earlier is also Gaussian with mean

$$m_{\mu_2, y} = \sum_{j=1}^n y_j C_\mu L_j, \qquad y = [y_1, y_2, \ldots, y_n].$$

Furthermore, its correlation operator is independent of y and is given by $C_{\mu, N} : F^* \to F$, where

$$C_{\mu, N}(L) = C_\mu(L) - \sum_{j=1}^n L\big(C_\mu(L_j)\big) C_\mu(L_j) \qquad \forall L \in F^*.$$

Note that $C_{\mu, N}(L_i) = 0 \, \forall i$ and that $L(C_\mu(L_i)) = 0 \, \forall i \Rightarrow C_{\mu, N}(L) = C_\mu(L)$. Thus the computed information $y = N(f)$ changes the measure μ by shifting the mean element from zero to $m_{\mu_2, y}$ and by modifying the correlation operator from C_μ into 0 in the linear subspace $\text{span}[L_1, \ldots, L_n]$.

The a posteriori measure $\nu(\blacksquare \,|\, y)$ defined by $\nu(C \,|\, y) = \mu_2(S^{-1}(C \,|\, y))$ is defined on the Borel sets of the Hilbert space G. It is Gaussian with mean $m_y = S m_{\mu_2, y}$ and has correlation operator $C_{\nu, N}(L) = S(C_{\mu, N}(LS)) \, \forall L \in G^*$. From Theorem 8 it follows that the algorithm

$$\Phi(y) = \sum_{j=1}^n y_j S C_\mu(L_j)$$

is an optimal error algorithm since it is the mean of the a posteriori distribution. The algorithm Φ can be expressed in the form $\Phi(y) = S\sigma(y)$, where $\sigma : N(F) \to F$ has two key properties:

(i) $N(\sigma(y)) = y$ (σ *interpolates* data).

(ii) Among interpolants satisfying (i), σ has minimal norm with respect to a certain norm depending on μ (σ is a *μ-spline*).

For this reason Φ is referred to as a *μ-spline algorithm* (see Wasilkowski and Woźniakowski [1982]).

More generally, Lee and Wasilkowski [1986] establish the optimality of Φ assuming only that G is a linear space (not necessarily normed), that S is linear (not necessarily continuous), and for a general error criterion (including the one we have discussed with an arbitrary $p \in [1, \infty)$).

Continuing the Hilbert space development, we summarize and push ahead by stating the following theorem.

THEOREM 9. *Let $S : F \to G$ be linear and continuous, with F a separable Banach space and G a separable Hilbert space. Let μ be a Gaussian measure on F with mean element zero. Let $N = [L_1, \ldots, L_n]$ be nonadaptive information, where the $L_i \in F^*$ are orthonormalized as indicated above. Then with correlation operators C_μ, C_ν, and $C_{\nu,N}$ as defined above,*

(a) *An optimal error algorithm is given by*

$$\Phi(y) = \sum_{j=1}^{n} y_j S C_\mu(L_j).$$

(b) *The radius of information is given by*

$$[r^{\text{ave}}(N)]^2 = \text{trace}(C_{\nu,N}) = \text{trace}(C_\nu) - \sum_{j=1}^{n} \left\| S(C_\mu L_j) \right\|^2.$$

PROOF. See Papageorgiou and Wasilkowski [1986].

The result of part (b) can be interpreted as follows. Before any computation has been done (via N) we only know the a priori measure ν and the best approximation to elements $g = Sf$ is zero (the mean of ν) with average error $[\text{trace}(C_\nu)]^{1/2}$. Formally this corresponds to the zero information operator whose average radius is given by $r^{\text{ave}}(0) = [\text{trace}(C_\nu)]^{1/2}$. Thus, $\text{trace}(C_\nu)$ measures the a priori uncertainty when only the formulation of the problem is known. After the computation of $y = N(f)$, the best approximation is the mean of the a posteriori measure $\nu(\blacksquare \mid y)$; and the trace of its correlation operator measures the a posteriori uncertainty $r^{\text{ave}}(N) = [\text{trace}(C_{\nu,N})]^{1/2}$. The quantity $\sum_{j=1}^{n} \|S(C_\mu L_j)\|^2$, the difference between these two trace values, depends on the solution operator S, the a priori measure μ (through its correlation operator), and the information operator N. The larger this difference is, the better the information provided by N. Under the assumptions of Theorem 9 the optimal error algorithm is linear. It would be nice to extend Theorem 9 to more general spaces and measures.

OPEN PROBLEM 5. Find general conditions on the spaces F and G and on the measure μ under which there exists a linear optimal error algorithm in the average case.

B. *Adaptive vs. nonadaptive information in the average case*. In surveying the question of whether adaption is more powerful than nonadaption for linear problems in an average case setting, we encounter a small surprise. While generally paralleling the results of the worst case, average case results are, in a sense, even stronger. We assume that the solution operator $S: F \to G$ is linear and that F is a separable Banach space equipped with a probability measure μ. We make no assumptions on G other than that it be a normed linear space.

Let N^a be adaptive information with *fixed cardinality* n ($n(f) = n \; \forall f \in F$). Thus,

$$N^a(f) = \left[L_1(f), L_2(f; y_1), \ldots, L_n(f; y_1, \ldots, y_{n-1}) \right],$$

where $y_i = L_i(f; y_1, \ldots, y_{i-1})$, $L_i(\blacksquare; y_1, \ldots, y_{i-1}) \in F^*$ and $L_i(f; \blacksquare)$ is measurable. For given adaptive information N^a we define nonadaptive information by fixing all values y_i via $y^* = [y_1^*, y_2^*, \ldots, y_{n-1}^*] \in \mathbb{R}^{n-1}$. We then define

$$N_{y^*}^{\mathrm{non}}(f) = \left[L_1(f), L_2(f; y_1^*), \ldots, L_n(f; y_1^*, y_2^*, \ldots, y_{n-1}^*) \right].$$

Such information is indeed nonadaptive, has the same cardinality as N^a, and is structurally simpler and easier to implement than N^a.

The following general result holds under a variety of assumptions on F, on its measure μ, and on G.

THEOREM 10. *Given adaptive information N^a with fixed cardinality there exists a vector y^* such that*

$$r^{\mathrm{ave}}\!\left(N_{y^*}^{\mathrm{non}} \right) \leqslant r^{\mathrm{ave}}(N^a).$$

Thus, in terms of intrinsic error as defined by the radius of information, adaption does not help in the average case.

PROOF. This result was proved by Traub, Wasilkowski, and Woźniakowski [1984] when F and G are finite-dimensional Hilbert spaces and μ is an elliptically contoured measure, and by Wasilkowski and Woźniakowski [1984a] when F and G are separable Hilbert spaces and μ is orthogonally invariant. With no conditions on G (other than that it be a linear space), the result was extended by Wasilkowski [1986b] to a separable Hilbert space F with a Gaussian measure; and by Lee and Wasilkowski [1986] to a separable Banach space F with a Gaussian measure.

With information of *varying cardinality*, Wasilkowski [1986a] shows that, while adaptive information *can* be more powerful, there are reasonable hypotheses under which adaption does not help. Furthermore, even in the general case, there exists equally powerful "nonadaptive" information provided we allow its cardinality $n(f)$ to have one of two values.

C. *Average case ε-complexity and optimal information for linear problems.* Throughout this section let $S: F \to G$ be linear and continuous, with F a separable Banach space and G a separable Hilbert space. Let $N = [L_1, L_2, \ldots, L_n]$ be nonadaptive information where the L_i can be chosen from the collection Λ of allowable continuous linear functionals. How should they be chosen to minimize the average radius of information?

We assume that μ is a Gaussian measure on F with mean zero and correlation operator C_μ. Without loss of generality, we further assume that $L_j(C_\mu(L_i)) = \delta_{ij}$ for any particular N that we are working with. Then the optimal error algorithm results from §A yield that the average radius of information is given by

$$[r^{\text{ave}}(N)]^2 = \int_F \left\| Sf - \sum_{j=1}^n L_j(f) SC_\mu(L_j) \right\|^2 \mu(df).$$

We wish to find continuous linear functionals L_j that minimize the average radius.

Following Papageorgiou and Wasilkowski [1986], let $\gamma_1, \gamma_2, \ldots$ be an orthonormal basis for G (if $\dim(G) < \infty$ set $\gamma_i = 0$ for $i > \dim(G)$) chosen so that

$$\text{span}\big[S(C_\mu L_1), \ldots, S(C_\mu L_n)\big] \subseteq \text{span}[\gamma_1, \ldots, \gamma_n].$$

Letting C_ν be the correlation operator of the measure $\nu = \mu S^{-1}$ and recalling that C_ν is nonnegative definite and has finite trace, we have

$$[r^{\text{ave}}(N)]^2 \geqslant \sum_{j=n+1}^\infty \int_F \langle Sf, \gamma_j \rangle^2 \mu(df) = \sum_{j=n+1}^\infty \langle C_\nu \gamma_j, \gamma_j \rangle$$

$$= \text{trace}(C_\nu) - \sum_{j=1}^n \langle C_\nu \gamma_j, \gamma_j \rangle.$$

Let γ_i^* be orthonormal eigenvectors for C_ν, $C_\nu \gamma_i^* = \lambda_i \gamma_i^*$, with $\lambda_i \geqslant \lambda_{i+1}$ ($\lambda_i = 0$ if $i > \dim(G)$). It is known (see, for example, Marcus and Minc [1964, Chapter 2]) that

$$\sum_{j=1}^n \langle C_\nu \gamma_j, \gamma_j \rangle \leqslant \sum_{j=1}^n \langle C_\nu \gamma_j^*, \gamma_j^* \rangle = \sum_{j=1}^n \lambda_j.$$

The above paragraph immediately gives $r^{\text{ave}}(N) \geqslant [\sum_{j=n+1}^\infty \lambda_j]^{1/2}$. This bound is sharp for $\Lambda = F^*$ (i.e., if all continuous linear functionals can be computed). Indeed, let $k = \sup\{i : \lambda_i > 0\}$ and let $m = \min\{k, n\}$. Then $\lambda_i > 0$ for $i \leqslant m$. Define information $N_m^* = [L_1^*, \ldots, L_m^*]$, where $L_i^*(f) = (1/\sqrt{\lambda_i})\langle Sf, \gamma_i^* \rangle$ for $i = 1, 2, \ldots, m$. Then

$$L_i^*(C_\mu L_j^*) = \frac{1}{(\lambda_i \lambda_j)^{1/2}} \int_F \langle Sf, \gamma_i^* \rangle \langle Sf, \gamma_j^* \rangle \mu(df) = \frac{1}{\lambda_i} \langle C_\nu \gamma_{i*}, \gamma_{j*} \rangle = \delta_{ij}.$$

From $L_i^*(C_\mu L_j^*) = \delta_{ij}$ and from the definition of L_i^*, we have

$$\frac{1}{\sqrt{\lambda_i}} \big(S(C_\mu L_j^*), \gamma_i^* \big) = \delta_{ij}.$$

Since the γ_i^* are orthonormal, $S(C_\mu L_j^*) = \sqrt{\lambda_j} \gamma_j^*$. Thus from part (b) of Theorem 9 we obtain $r^{\text{ave}}(N_m^*) = [\sum_{j=m+1}^\infty \lambda_j]^{1/2}$. We then have:

THEOREM 11. *In the above setting, N_m^* is the nth optimal information and the optimal error algorithm using N_m^* takes the form*

$$\Phi^{\text{opt}}\big(N_m^*(f)\big) = \sum_{j=1}^n \langle Sf, \gamma_j^* \rangle \gamma_j^*$$

with error given by $e(\Phi^{\text{opt}}, N_m^*) = r(N_m^*) = [\sum_{j=n+1}^\infty \lambda_j]^{1/2}$.

Note that Φ^{opt} is the truncated Fourier series of the solution element Sf with respect to the orthonormal basis consisting of eigenvectors of the correlation operator C_ν.

We illustrate Theorem 11 by the multivariate approximation problem studied by 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 partial derivatives of f with respect to x_j, $j = 1, 2, \ldots, d$.

Let F be the Banach space of functions for which $f^{(r, r, \ldots, r)}$ is continuous and $f^{(i_1, i_2, \ldots, i_d)}(t) = 0$ for all $i_j \leqslant r$ and any t with at least one zero component. The space F is equipped with the norm $\|f\| = \sup_{t \in D} |f^{(r, r, \ldots, r)}(t)|$ and with the classical Wiener measure placed on rth derivatives.

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

$$\lambda_1 = \left(\frac{(\ln i)^{d-1}}{\pi^d (d-1)! i} \right)^{2r+2} (1 + o(1)), \quad \text{as } i \to \infty.$$

The eigenvectors γ_i are the solutions of the differential equation

$$\lambda_i f^{(r+2, r+2, \ldots, r+2)}(x) - (-1)^{rd} f(x) = 0$$

with zero boundary conditions. The radius of information for N_n^* is given by

$$r(N_n^*) = \frac{1}{\sqrt{2r+1} \left[\pi^d (d-1)! \right]^{r+1}} \frac{(\ln n)^{(r+1)(d-1)}}{n^{r+1/2}} (1 + o(1)).$$

The problem of optimal information in the average case setting becomes technically much more difficult if Λ is a proper subset of F^*. In many practically important cases the only linear functionals that can be computed are function evaluations. For example, let F be a space of scalar or multivariate functions f defined on a domain $\Omega \subseteq \mathbb{R}^k$. Then $L \in \Lambda$ if $L(f) = f(t)$ for some $t \in \Omega$. Some interesting work on optimal choice of evaluation points (called the optimal design problem in the statistical literature) can be found in Ylvisaker [1975] and Wahba [1971]. We thus propose:

OPEN PROBLEM 6. For the general linear problem (or significant subclasses thereof) find nth optimal information consisting of function evaluations. The case of multivariate functions seems especially interesting.

Assuming now that $\Lambda = F^*$ and investigating ε-complexity we must consider ε-approximations using information operators N for which $r^{\mathrm{ave}}(N) \leqslant \varepsilon$. Using the result of Theorem 10 (adaption does not help), we can restrict ourselves to nonadaptive information N. Proceeding in parallel fashion to what we did in the worst case, let

$$m(\varepsilon) = \min\{\text{cardinality of } N : r^{\mathrm{ave}}(N) \leqslant \varepsilon\}.$$

From the results just developed above for $r^{\mathrm{ave}}(N)$, we then have

$$m(\varepsilon) = \min\left\{ n : \sum_{j=n+1}^{\infty} \lambda_j \leqslant \varepsilon^2 \right\},$$

where the λ_j are the eigenvalues of the correlation operator C_ν for the measure $\nu = \mu S^{-1}$. By applying the algorithm Φ^{opt} described above to the $m(\varepsilon)th$ optimal information $N^*_{m(\varepsilon)}$ we conclude that

$$cm(\varepsilon) \leqslant \mathrm{comp}^{\mathrm{ave}}(\varepsilon) \leqslant (c+2)m(\varepsilon) - 1.$$

Here c is the cost of a single continuous linear functional evaluation and we assume that each addition and scalar multiplication in G costs unity. If combinatory cost is dominated by information cost ($c \gg 1$), we obtain $\mathrm{comp}^{\mathrm{ave}}(\varepsilon) \approx cm(\varepsilon)$. We conclude that the information $N^*_{m(\varepsilon)}$ and the algorithm Φ^{opt} are almost optimal. As in the worst case, it can be shown that the average ε-complexity can go to infinity arbitrarily fast for any $\varepsilon > 0$. Thus linear problems can also be intractable in the average case setting.

As in the worst case setting, we believe that for many linear problems the ε-complexity is approximately proportional to the ε-cardinality number $m(\varepsilon)$.

OPEN PROBLEM 7. Obtain general conditions on the spaces F and G, the linear operator S, the measure μ, and the class of allowable linear functionals Λ such that $\mathrm{comp}(\varepsilon) \approx cm(\varepsilon)$.

VI. Concluding comments. We have concentrated on worst case and average case settings for treating error and cost, as embodied in the definitions of radius of information, ε-complexity, and algorithm optimality. A variety of other settings are possible—we briefly mention two of them here. In a *probabilistic setting* we relax the worst case requirement that the error be at most ε for all problem elements in the domain F. Instead we work with a probability measure on F and a parameter $\delta \in [0,1]$ and we require that the error be at most ε for a set of elements of measure at least $1 - \delta$. The cost in such a setting can be defined as in either our worst or average case treatments and either with or without concern about sets of measure $\leqslant \delta$. The reader is referred to Wasilkowski [1986a, 1986b], Lee and Wasilkowski [1986], and Woźniakowski [1986b] for some of the initial work in this new setting.

Another setting is motivated by the fact that it may be very difficult to determine the optimal information operator N^* of a given cardinality n and even if it is found we may have $r(S, N^*) = \infty$. In such cases one may be able to fix $f \in F$ and approximate $S(f)$ by a convergent sequence of algorithm values. For obvious reasons this approach is referred to as the *asymptotic setting*. See Trojan [1984] and Wasilkowski and Woźniakowski [1984b] for the first general treatment of the asymptotic setting. Earlier Traub [1961, 1964] studied the asymptotic setting for the solution of nonlinear equations. This was the first work which utilized information-centered arguments in complexity theory.

If, in conjunction with the various options for settings, we allow for different choices of error criteria such as absolute (the one we have used), relative, and normalized, we obtain a small combinatorial explosion of possible settings for information-based complexity (see Woźniakowski [1986c] for analysis of integration with respect to some of these possible settings). To complicate matters further, it may sometimes be appropriate to work in *mixed settings* where, for instance, error might be measured in the worst case and cost in the average case.

Our focus has been on theoretical aspects of information-based complexity. Much interesting work is also being done to relate the theory to more applied branches of mathematics. We cite a few examples in the paragraphs that follow.

For a survey of some results on approximate solution of linear partial differential and integral equations in the worst case setting see Werschulz [1985a]. To state one particular, Werschulz [1985b] considers the elliptic boundary value problem $Lu = f$ where L is a fixed linear elliptic operator and f belongs to the unit ball in an appropriately chosen Sobolev norm. Note that this is an example of a linear problem since solution values depend linearly on f. Werschulz [1982a, 1982b] gives necessary and sufficient conditions under which the finite element method is optimal.

We believe it is important to generalize this work by allowing the elliptic operator to vary, for instance, as a function of coefficients in the elliptic form. Then it is natural to assume partial information on L as well as on f. The problem then becomes nonlinear.

OPEN PROBLEM 8. Find or estimate the ε-complexity of the above nonlinear elliptic boundary value problem in the worst case or average case setting.

The following nonlinear constrained optimization problem is treated in Nemirovsky and Yudin [1983]: let B be a subset of a real Banach space, let E be a class of continuous real-valued functions on B, and let

$$F = \left\{ (f_0, f_1, \ldots, f_m) : f_j \in E \text{ for } j = 0, 1, \ldots, m \right\}.$$

The solution operator $S : F \to \mathbb{R}$ is defined for $f = (f_0, f_1, \ldots, f_m)$ by

$$S(f) = \inf\left\{ f_0(x) : x \in G \text{ and } f_j(x) \leqslant 0 \text{ for } j = 1, 2, \ldots, m \right\}.$$

Given $\varepsilon > 0$, sharp estimates are given on the number $m(\varepsilon)$ of function evaluations of f and f' required to find $S(f)$ within ε. Thus, if B is convex and E consists of convex functions, $m(\varepsilon) = \Theta(\ln(1/\varepsilon))$; while if B is compact of dimension d and E consists of r times continuously differentiable functions (not necessarily convex), $m(\varepsilon) = \Theta((\varepsilon^{-1})^{d/r})$. Nemirovsky and Yudin do not study combinatory cost and hence only obtain lower bounds on the ε-complexity of this problem.

OPEN PROBLEM 9. Find or estimate the ε-complexity for the above nonlinear optimization problems.

As our final applied example we mention the problem of solving the nonlinear equation $f(x) = 0$. This nonlinear problem is the subject of many papers in the worst case and asymptotic settings. A recent survey can be found in Sikorski [1985]. It would clearly be of interest to analyze this problem in an average case setting. Some significant results in this direction have been obtained by Smale [1985] and Shub and Smale [1985, 1986a, and 1986b] under the assumption that f is a polynomial of known degree. If, however, f is merely assumed to be a smooth function, the average case analysis remains to be done. The choice of an a priori measure is by no means obvious.

OPEN PROBLEM 10. Develop general results on the solution of nonlinear equations in the average case setting.

With regard to future directions for work in information-based complexity, we mention several general areas. As indicated above, many settings, both pure and mixed, have yet to be seriously explored. Likewise, little has been done with models of computation other than the real number model to which we have restricted ourselves. Finally, we mention the realistic possibility that the problem domain F may not be precisely specifiable (thus, for instance, we may not know exactly how many times functions in a given class are differentiable). The challenge of finding algorithms that work well for problems from a variety of possible domains is picturesquely called the "fat F" problem. See Motornyj [1973] and Werschulz [1982a, 1982b] for some specialized results on this problem.

We have tried to highlight some of the interesting mathematics that has come into play in information-based complexity. A number of subject and application areas are clearly providing fruitful connections with the general theory. We believe, and our short list of open problems is one testimony to this, that information-based complexity offers much to interest and challenge mathematicians working in a variety of fields.

BIBLIOGRAPHY

I. Adler, R. M. Karp, and R. Shamir (1983), *A simplex variant solving an $m \times d$ linear program in $O(\min(m^2, d^2))$ expected number of pivot steps*, Report UCB CSD 83/158, University of California, Berkeley.

I. Adler and N. Megiddo (1985), *A simplex algorithm whose average number of steps is bounded between two quadratic functions of the smaller dimension*, J. Assoc. Comput. Mach. **32**, 871–895.

M. Atteia (1965), *Fonctions-spline généralisées*, C. R. Acad. Sci. Paris **216**, 2149–2152.

N. S. Bakhvalov (1971), *On the optimality of linear methods for operator approximation in convex classes of functions*, Zh. Vychisl. Mat. Fiz. **11**, 244–249.

L. Blum and M. Shub (1986), *Evaluation of rational functions: Infinite precision is finite cose and tractable on average*, SIAM J. Comput. **15**, 384–398.

A. Borodin (1972), *Complexity classes of recursive functions and the existence of complexity gaps*, J. Assoc. Comput. Mach. **19**, 158–174, 576.

S. Gal and C. A. Micchelli (1980), *Optimal sequential and non-sequential procedures for evaluating a functional*, Applicable Anal. **10**, 105–120.

R. Holmes (1972), *R-splines in Banach spaces*, I. *Interpolation of linear manifolds*, J. Math. Anal. Appl. **40**, 547–593.

J. B. Kadane and G. W. Wasilkowski (1985), *Average case ε-complexity: A Bayesian view*, Bayesian Statistics **2**, Proc. Second Valencia Internat. Meeting (September 6/10, 1983) (J. M. Bernardo, M. H. Degroot, D. V. Lindley, A. F. M. Smith, eds)., 361–374.

R. M. Karp (1976), *The probabilistic analysis of some combinatorial search algorithms*, Algorithms and Complexity: New Directions and Recent Results (J. F. Traub, ed.) Academic Press, New York, 1–19.

_____ (1979), *Recent advances in probabilistic analysis of graph-theoretic algorithms*, Proc. 6th Colloq. Auto. Lang. and Program., Lecture Notes in Computer Sci. **71**, 338–339.

_____ (1980), *An algorithm to solve the $m \times n$ assignment problem in expected time $O(mn \log n)$*, Networks **10**, 143–152.

R. M. Karp and M. Luby (1983), *Monte-Carlo algorithms for enumeration and reliability problems*, Proc. 24th IEEE Found. of Comp. Sci. Sympos., 56–60.

_____ (1985), *A Monte-Carlo algorithm for the multiterminal reliability problem*, J. Complexity **1**, 45–64.

J. Kiefer (1953), *Sequential minimax search for a maximum*, Proc. Amer. Math. Soc. **4**, 502–505.

G. S. Kimeldorf and G. Wahba (1970a), *A correspondence between Bayesian estimation on stochastic processes and smoothing by splines*, Ann. Math. Statist. **41**, 2, 495–502.

_____ (1970b), *Spline functions and stochastic processes*, Sankhyā Ser. A **32**, 173–180.

H. H. Kuo (1975), *Gaussian measures in Banach spaces*, Lecture Notes in Math., vol. 463, Springer-Verlag, Berlin and New York.

F. M. Larkin (1972), *Gaussian measure in Hilbert space and application in numerical analysis*, Rocky Mountain J. Math. **2**, 378–421.

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

M. Marcus and H. Minc (1964), *A survey of matrix theory and matrix inequalities*, Allyn and Bacon, Inc., Boston.

C. A. Micchelli and T. J. Rivlin (1977), *A survey of optimal recovery*, Optimal Estimation in Approximation Theory (C. A. Micchelli and T. J. Rivlin, eds.), Plenum Press, New York, 1–54.

V. P. Motornyj (1973), *On the best quadrature formula of the form $\sum_{k=1}^{n} p_k f(x_k)$ for some classes of periodic differentiable functions*, Dokl. Akad. Nauk SSSR **211**, 1060–1062 [English transl.: Soviet Math. Dokl. **14**, 1180–1183].

A. S. Nemirovsky and D. B. Yudin (1983), *Problem complexity and method efficiency in optimization*, Wiley-Interscience, New York. Translated from *Slozhnost' zadach i effektivnost' metodov optimizatsii*.

S. M. Nikolskij (1950), *On the problem of approximation estimate by quadrature formulae* (in Russian), Uspekhi Mat. Nauk **5**, 165–177.

K. Yu. Osipenko (1976), *Best approximation of analytic functions from information about their values at a finite number of points* (in Russian), Mat. Zametki **19**, 29–40 [English transl.: Math. Notes **19**, 17–23].

E. W. Packel (1986a), *Linear problems (with extended range) have linear optimal algorithms*, Aequationes Mathematicae **31**, 18–25.

_____ (1986b), *Do linear problems have linear optimal algorithms?* SIAM Rev. (to appear).

A. Papageorgiou and G. W. Wasilkowski (1986), *Average case complexity for multivariate problems*, Report, Dept. of Computer Science, Columbia University.

K. R. Parthasarathy (1967), *Probability measures on metric spaces*, Academic Press, New York.

M. O. Rabin (1976), *Probabilistic algorithms*, Algorithms and Complexity: New Directions and Recent Results (J. F. Traub, ed.), Academic Press, New York, 21–39.

_____ (1983), *Randomized Byzantine generals*, Proc. 24th IEEE Sympos. Found. of Comp. Sci.

_____ (1986), *Removing singularities through randomization*, in progress.

J. Renegar (1984), *On the efficiency of Newton's method in approximating all zeros of a system of complex polynomials*, Report, Colorado State University.

J. Sacks and D. Ylvisaker (1970), *Statistical designs and integral approximation*, Proc. 12th Biennial Seminar Canadian Math. Congr., 115–136.

A. Sard (1949), *Best approximate integration formulas; best approximation formulas*, Amer. J. Math. **71**, 80–91.

M. Shub and S. Smale (1985), *Computational complexity: On the geometry of polynomials and a theory of cost*, Part I, Ann. Sci. École Norm. Sup. 4 serie, **18**.

_____ (1986a), *Computational complexity: On the geometry of polynomials and a theory of cost*, Part II, SIAM J. Comput. **15**, 145–161.

_____ (1986b), *On the existence of generally convergent algorithms*, J. Complexity **2**, 2–11.

K. Sikorski (1982), *Bisection is optimal*, Numer. Math. **40**, 111–117.

_____ (1985), *Optimal solution of nonlinear equations*, J. Complexity **1**, 197–209.

A. V. Skorohod (1974), *Integration in Hilbert space*, Springer-Verlag, New York.

S. Smale (1981), *The fundamental theorem of algebra and complexity theory*, Bull. Amer. Math. Soc. **4**, 1–36.

_____ (1983a), *The problem of the average speed of the simplex method*, Mathematical Programming: The State of the Art (A. Bachem et al., eds.), Springer-Verlag, New York, 530–539.

_____ (1983b), *On the average number of steps in the simplex method of linear programming*, Math. Programming **27**, 241–262.

_____ (1985), *On the efficiency of algorithms of analysis*, Bull. Amer. Math. Soc. **13**, 87–121.

S. A. Smolyak (1965), *On optimal restoration of functions and functionals of them*, Candidate's Dissertation, Moscow State Univ. (Russian).

A. V. Suldin, *Wiener measure and its application to approximation methods*, I, II, Izv. Vyssh. Uchebn. Zaved. Mat. **1959**, no. 6 (13), 145–158, **1960**, no. 5 (18), 165–179 (Math. Rev. **28** #722, **29** #1482).

J. F. Traub (1961), *On functional iteration and the calculation of roots*, Preprints of Papers 16th Nat. ACM Conf., Session 5A-1, pp. 1–4, Los Angeles.

_____ (1964), *Iterative methods for the solution of equations*, Prentice-Hall, Englewood Cliffs, New Jersey; reprinted by Chelsea Press, New York, 1982.

J. F. Traub, G. W. Wasilkowski and H. Woźniakowski (1983), *Information, uncertainty, complexity*, Addison Wesley, Reading, Mass.

_____ (1984), *Average case optimality for linear problems*, Journal TCS **29**, 1–25.

J. F. Traub and H. Woźniakowski (1980), *A general theory of optimal algorithms*, Academic Press, New York.

G. M. Trojan (1984), *Asymptotic setting for linear problems*, unpublished manuscript.

S. Twomey (1977), *Introduction to mathematics of inversion in remote sensing and indirect measurements*, Developments in Geomathematics 3, Elsevier Scientific Publ., Amsterdam.

N. N. Vakhania (1981), *Probability distributions on linear spaces*, North-Holland, New York.

G. Wahba (1971), *On the regression design problem of Sacks and Ylvisaker*, Ann. Math. Statist. **42**, 1035–1053.

_____ (1978), *Improper priors, spline smoothing and the problem of guarding against model errors in regression*, J. Royal Statist. Soc. B **40**, 3, 364–372.

G. W. Wasilkowski (1983), *Local average error*, Report, Dept. of Computer Science, Columbia University.

_____ (1985), *Average case optimality*, J. Complexity **1**, 107–117.

_____ (1986a), *Information of varying cardinality*, J. Complexity **2**, 204–228.

_____ (1986b), *Optimal algorithms for linear problems with Gaussian measures*, Rocky Mountain J. Math. **16**, 727–749.

G. H. Wasilkowski and H. Woźniakowski (1982), *Average case optimal algorithms in Hilbert spaces*, Report, Dept. of Computer Science, Columbia University; also in J. Approx. Theory **47**, 17–25.

_____ (1984a), *Can adaption help on the average?*, Numer. Math. **44**, 169–190.

_____ (1984b), *On optimal algorithms in an asymptotic model with Gaussian measure*, Report, Dept of Computer Science, Columbia University.

A. G. Werschulz (1982a), *Optimal error properties of finite element methods for second order Dirichlet problems*, Math. Comp. **38**, 401–413.

_____ (1982b), *Finite element methods are not always optimal*, Report, Univ. of Maryland, Baltimore County.

_____ (1985a), *Complexity of differential and integral equations*, J. Complexity **1**, 232–255.

_____ (1985b), *What is the complexity of elliptic systems?*, J. Approx. Theory **45**, 69–89.

A. G. Werschulz and H. Woźniakowski (1986), *Are linear algorithms always good for linear problems?*, Aequationes Math. **31**, 202–211.

H. Woźniakowski (1986a), *Information-based complexity*, Ann. Rev. Comput. Sci. Volume 1, 319–380.

_____ (1986b), *Probabilistic setting of information-based complexity*, J. Complexity **2**, 255–269.

_____ (1986c), *Complexity of integration in different settings*, Proc. Optimal Algorithms (Sofia, Bulgaria), 235–240.

D. Ylvisaker (1975), *Designs on random fields*, A Survey of Statistical Design and Linear Models (J. N. Srivastava, ed.), North-Holland, New York, 593–607.

DEPARTMENT OF MATHEMATICS, LAKE FOREST COLLEGE, LAKE FOREST, ILLINOIS 60045 (Affiliation of Edward W. Packel)

DEPARTMENT OF COMPUTER SCIENCE, COLUMBIA UNIVERSITY, NEW YORK, NEW YORK 10027

INSTITUTE OF INFORMATICS, UNIVERSITY OF WARSAW, WARSAW, POLAND