A. Grothendieck and J. L. Verdier, *Topos*, Artin et al. (eds.) Lecture Notes in Math., vol. 269, Springer-Verlag, Berlin and New York, 1972, 229–515.

L. A. Henkin, *Completeness in the theory of types*, J. Symbolic Logic **15** (1950), 81–91.

P. T. Johnstone, *Topos theory*, London Mathematical Society Monographs, vol. 10, Academic Press, London 1977.

S. A. Kripke, *Semantic analysis of intuitionistic logic*, I, J. N. Crossley et al. (eds.), Formal Systems and Recursive Functions, North-Holland Publ. Co., Amsterdam, 1965.

J. Lambek, *From types to sets*, Advances in Math. **36** (1980), 113–164.

____, *On the unity of algebra and logic*, F. Borceux (ed.), Categorical Algebra and its Applications, Lecture Notes in Math., vol. 1348, Springer-Verlag, Berlin and New York, 1988, pp. 221–229.

J. Lambek and P. J. Scott, *Intuitionistic type theory and the free topos*, J. Pure Appl. Algebra **19** (1980), 215–257.

____, *New proofs of some intuitionistic principles*, Z. Math. Logik Grundlag. Math. **29** (1983), 493–504.

____, *Introduction to higher order categorical logic*, Cambridge, Univ. Press, 1986.

F. W. Lawvere, *An elementary theory of the category of sets*, Proc. Nat. Acad. Sci. U.S.A. **52** (1964), 1506–1511.

____, *Introduction to toposes, algebraic geometry and logic*, Lecture Notes in Math., vol. 274, Springer-Verlag, Berlin and New York, 1972, pp. 1–12.

____, *Variable quantities and variable structures in topoi*, A. Heller et al. (eds.), Algebra, Topology and Category Theory, Academic Press, 1976, pp. 101–131.

F. W. Lawvere et al. (eds.), *Model theory and topoi*, Lecture Notes in Math., vol. 445, Springer-Verlag, Berlin and New York, 1975.

M. Makkai and G. E. Reyes, *First order categorical logic*, Lecture Notes in Math., vol. 661, Springer-Verlag, Berlin and New York, 1977.

W. Mitchell, *Boolean topoi and the theory of sets*, J. Pure Appl. Algebra **2** (1972), 261–274.

B. Russell and A. N. Whitehead, *Principia Mathematica* I–III, Cambridge Univ. Press, pp. 1910–1913.

M. Tierney, *Sheaf theory and the continuum hypothesis*, Toposes, Algebraic Geometry and Logic, F. W. Lawvere (ed.), Lecture Notes in Math., vol. 274, Springer-Verlag, Berlin and New York, 1972, pp. 13–42.

H. Volger, *Logical categories, semantical categories and topoi*, Model Theory and Topoi, F. W. Lawvere et al. (eds.), Springer-Verlag, Berlin and New York, 1975, pp. 87–100.

J. LAMBEK

McGILL UNIVERSITY

Information-based complexity, by J. F. Traub, G. W. Wasilkowski and H. Woźniakowski. Academic Press, Boston, San Diego, and New York, 1988, xiii + 523 pp., $64.50. ISBN 0-12-697545-0

In the past ten years, there have been a good number of developments in information-based complexity theory; in addition, the field and related issues have gained more attention in the mathematical community. This book fills a need for information on recent developments, and it comprehensively describes older and better-known results.

Beside its many applications in computational sciences and numerical analysis, information-based complexity theory has a strong functional analytic leaning which might profitably be exemplified at this point. Given two infinite dimensional normed linear spaces F and G and a map $S: F \to G$, consider the problem of how such a map might be best expressed in terms of finite dimensional objects (the relevance of this question to real-world computations may be clear, but we will discuss this further later).

For concreteness, let us given an example. Let $F = H^r(\Omega)$ be the Sobolev space of functions on the bounded, simply connected domain $\Omega \subset \mathbf{R}^n$ with r derivatives which are square integrable. Let A be a uniformly elliptic differential operator of order $2m$ on Ω (we will assume here that $r \geq -m$). Given a function $f \in F$ and the problem $Au = f$, the solution is given by $u = Sf$, where $S = A^{-1}$. One might imagine that Sf could be computed by finding the integral kernel of S and integrating it against f. If this map were implemented computationally, it is clear the computer would have to be working with finite dimensional objects. The procedure followed in the computation would first involve a truncation of f to something finite dimensional which can be represented in a computer. For example, f may be represented through its values at a finite set of points or through a finite number of coefficients in an eigenfunction expansion. Let us denote by $N: F \to \mathbf{R}^n$ the operator which "truncates" f. The truncated object Nf must then be operated on in order to calculate (or approximate) the effect of the resolvent operator S on f. Let us denote this second operation by $\phi: \mathbf{R}^n \to G = H_0^m(\Omega)$. Here $H_0^m(\Omega)$ denotes those functions in $H^m(\Omega)$ whose first $m - 1$ normal derivatives vanish on the boundary $\partial\Omega$. The essential problem here is (for fixed dimension n) to find optimal N and ϕ so that the "computed" function $\phi(N(f))$ is a good approximation to the solution Sf, or equivalently, so that the diagram in Figure 1 commutes to the greatest extent possible.
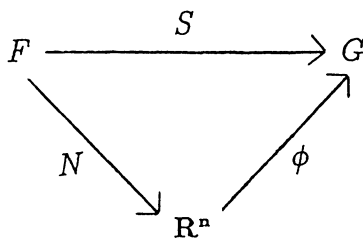

FIGURE 1

It is interesting to note that in this case the so-called finite element method used in numerical analysis has been analyzed by Werschulz [W] in the context of this formulation. With properly chosen parameters the method yields an almost optimal pair (N^*, ϕ^*), denoted as "information" and "algorithm," respectively, for solving this problem. That is, the maximal error (for f in the unit ball of $H^r(\Omega)$) of approximating Sf by $N^*(\phi^*(f))$ is no more than a constant multiple of that for the best possible choice of (N, ϕ) (using a linear information operator N). In this case, the

information operator of rank n is $N^*f = (\langle f, s_1\rangle, \langle f, s_2\rangle, \ldots, \langle f, s_n\rangle)$. Here, $\langle \cdot, \cdot \rangle$ denotes inner product, and $\{s_n\}$ are basis functions for the family of splines, i.e., functions in $C^{m-1}(\Omega)$ which are piecewise polynomials (on a pre-defined triangulation of Ω) of a fixed degree larger than $2m - r + 1$. The algorithm is a linear map $\phi^*: \mathbf{R}^n \to H^m$ which when composed with N^* has the form $\phi^*(N^*(f)) = \sum_{j=1}^n \langle f, s_j\rangle g_j$, where the functions g_j can be pre-computed. Thus, an entirely different method from our initial suggestion (which was finding a kernel and integrating against it) proves to be almost optimal. Indeed, the finite element method described above has long been used to solve partial differential equations.

In general, the process of applying N to f abstractly represents the extraction of usable information from f; hence N is called an information operator. The operator ϕ represents how the computer uses the information Nf in order to approximate the solution Sf, and hence is called an algorithm. To summarize the above discussion, the most basic paradigm of information-based complexity is mathematically a study of the problem of filtering infinite dimensional maps through finite dimensional spaces.

It should be mentioned that above, a numerical "cost" (or complexity) can be assigned to the algorithm ϕ, based on the types and number of operations (e.g., addition, multiplication) the computation of ϕ involves. It can be asked what the smallest cost is of a ϕ for which

$$\sup_{\|f\| \leq 1} \|Sf - \phi(N(f))\| < \varepsilon$$

for given ε, where $\| \cdot \|$ denotes norm in the space F or G. In the recent work on complexity of root finding for real and complex functions, the focus is on the minimization of this cost, based on the assumption of full information about the function whose roots are to be computed. Study of complexity of root finding was originally investigated extensively by Traub [T], and has more recently been initiated by Hirsch, Smale, and Shub (see [HS, S]).

Information-based complexity belongs to the field of analytic (or continuous) complexity theory, as opposed to combinatorial complexity. The latter field is epitomized by the solution of problems which ask essentially the number of permissible operations required to solve problems such as matrix inversion or multiplication. These combinatorial complexity problems are also obviously important. Any nontrivial reduction in the enormity of the number of steps in solving a problem is generally significant, certainly asymptotically as the problem size becomes very large.

Analytic complexity has crystallized as a distinct field more recently. Here numerical bit operations do not play a central role (indeed, such calculations are generally assumed to have infinite precision), and basic arithmetic operations (e.g., additions or multiplications of real numbers) are treated as primitive operations. Into this latter category falls the recent work in complexity of root finding for real and complex functions.

As indicated above, the approach in information-based complexity is analytic rather than algebraic or combinatorial. It is motivated by a desire to formalize approaches to the solution of analytic problems involving

partial or contaminated information. To the analyst it may initially not
seem that many problems fall into this category, but it is indeed a fact that
they do, for the following reason. While the work of the typical analyst
involves relations of analytic objects such as functions and operators, the
relation of these objects to what one can get one's hands on can be made in
only two ways. Sometimes analytic problems have exact analytic solutions,
and that is the end of it. However, existence of exact analytic solutions in
such situations is exceedingly rare. It may happen that one wants to get
one's hands on the solution of a partial differential equation, and thus will
need to compute the resolvent of an operator applied to a given function.
It is unlikely that for a given function, the resolvent applied to it will be
exactly computable. The alternative, as in the above example, is to repre-
sent the function in a computer, and evaluate the resolvent applied to the
function by integrating against the resolvent kernel. Aside from the details
involved in such a procedure, it is clear that if one is confident of the accu-
racy of the computer's ability to add and multiply with sufficient precision,
the essence of interest here should be in how the computer represents the
function internally, and what it does with that information once it has it.

Just as in combinatorial complexity theory, a central problem in inform-
ation-based complexity theory is how much work it takes to compute an
approximation such as the one mentioned in the second paragraph (com-
putational complexity). One can define work however one wishes, and this
is done typically by assigning numerical cost (or complexity) to the oper-
ations involved in computing the approximation $\phi \circ N$, and asking what
choice of N and ϕ minimizes cost while producing an error smaller than
a given tolerance ε. The ε-complexity of a problem S is the minimal cost
of computing an approximation $\phi \circ N$ of S.

Just as in thermodynamics, where there are precise upper bounds on the
output of work from a thermodynamical system, there are similarly lower
bounds on the amount of work required to solve a problem; this work is
the complexity of the problem. An important aspect of this field is related
to the study of those lower complexity bounds.

Upper and lower bounds are of interest, and in some cases exact com-
plexities of problems can be obtained. For example, the complexity of
approximating a function in the unit ball of the Sobolev space of func-
tions in the unit cube of $\mathbf{R}^n$ which have r derivatives is (under some mi-
nor assumptions) $C(n, r) = O(1/\varepsilon^{n/r})$. This is, in order to identify such a
function f within ε units in L^2 norm, it is required that at least $C(n, r)$
linear functionals of f be calculated, and $C(n, r)$ linear functionals also
suffice. (It turns out that in this case the costs of the operations involved
in computing the algorithm ϕ do not change the answer.)

One difference between the two approaches is that a desired accuracy
of approximation needs to be specified beforehand in analytic complex-
ity theory, while combinatorial complexity problems are generally solved
exactly.

A paradigm in information-based complexity is the identification of a
vector f in large or infinite dimensional space based on partial informa-
tion about that vector. If f is in the unit ball of a space H^r of r-times

differentiable functions on $\mathbf{R}$, for example, information about f may be available in the form of the vector $y = N(f) = (f(x_1), f(x_2), \ldots, f(x_k))$, $x_i \in \mathbf{R}$, which is a set of k continuous linear functionals evaluated at f. Given just the information above, there are of course many functions f which that information may correspond to. One important strategy for choosing among these corresponds to the constraint that one choose the vector $f' \in N^{-1}(y)$ which satisfies some other condition, such as minimization of norm in the original Sobolev space H^r. This requirement may be invoked as one of simplicity if nothing else, e.g., the requirement that we choose our approximation to f to be as smooth as possible within the constraints of the information y.

To see this better, one might imagine how any system which infers shape from visual cues would extrapolate the shape of a surface from the partial distance information it were provided with. It is known that one appropriate method to do this is to infer the smoothest possible surface consistent with the data; this should be compared with the process of finding a function of smallest possible Sobolev norm still consistent with the data, as described above. This example not only encompasses the problem of finding the shape of a surface with partial information, but also with contaminated information, i.e., information with error in it. Such problems come up in areas like signal processing, prediction and estimation, remote sensing, and ill posed problems.

Providing a general format for modeling such problems is the central theme in information-based complexity. The problem of evaluating $A^{-1}f$ (discussed above) even as an integration against the kernel $K(x, y)$ of $S \equiv A^{-1}$ also involves partial information. That is, even though K can be known with arbitrary accuracy in principle, it is in practice known only by its values at a finite number of points. This limits the calculation of the integral $(Sf)(x) = \int K(x, y)f(y)\, dy$, unless K and f happen to have analytic forms which make the integral calculable analytically. It may seem here that the former situation does not come up in practice; however, if this is so it is an artifact of the reader's experience. In practice, for example, in the solution of a differential equation, such an integral is calculated either with the full knowledge of the analytic form of $K(x, y)$, or with partial knowledge of K, say, at a finite number of points (in case an exact analytic form of K is not available).

What is the criterion used to decide whether an approximation algorithm is good or not? One such criterion is called the worst case criterion, wherein one picks the problem element f of norm less than or equal to 1 for which the norm error $e(\phi, N, f) \equiv \|Sf - \phi(N(f))\|$ is largest, and chooses N and ϕ so that this worst case error is smallest.

A second criterion which is used has to do with so-called average case complexity. Here there is assumed to be a probability measure μ on the space F which somehow corresponds to the frequency with which elements of the F will be selected in the process of approximating $Sf = A^{-1}f$. In applications so far, this measure has been assumed to be an infinite dimensional Gaussian or a combination of Gaussians, partly because there is a

limited collection of interesting and usable measures on infinite dimensional spaces. The quantity which one wants to minimize in this setting is the so-called average case error, given by $e^{\mathrm{avg}}(\phi, N) \equiv E(\|Sf - \phi(N(f))\|)$, where the expectation is taken with respect to the probability measure μ. This situation is important because in some approximation schemes worst case error may be quite large, but the elements $f \in F$ for which the worst possible error is actually attained may have small likelihood of being selected. In these cases the averaged error over all possible problem elements f is a more accurate measurement of the usefulness of the information operator N and the algorithm operator ϕ.

It has become well known that for the solution of some problems it is useful to use random information rather than deterministic information. To illustrate this, consider the Monte Carlo method of evaluating an integral of a continuous function f over a Borel subset B of a topological space M, with Borel measure μ. Points $x_i \in B$ are chosen at random according to a scheme wherein essentially the density of points in a subset is proportional to the measure of the subset, and the sum $(\mu(B)/n) \sum_{i=1}^{n} f(x_i)$ is used to approximate the integral $Sf \equiv \int_B f(x) \, d\mu(x)$. Thus, whereas the above deterministic methods of operator approximation would dictate that Sf be evaluated, for example, using the information $Nf \equiv (f(x_1), \ldots, f(x_n))$ where $x_1, \ldots, x_n$ are predetermined via some approximation scheme (say "uniformly" distributed over the set B in the case $\mu =$ Lebesgue measure), here we are choosing our information at random in that the x_i are random. Clearly notions of error in this setting must be probabilistic; information-based complexity is the right formalism for studying Monte Carlo procedures such as this. The setting has added relevance given recent highly effective "quantum Monte Carlo" methods for solving problems in quantum mechanics involving finding eigenfunctions of operators using methods involving analogues of simulated annealing [CA].

A second related issue is that of contaminated information. This can be modeled as above, with error smaller than or equal to ε added to the information $N(f)$ above. In this case, one often tries to minimize the worst case error, given by $\sup_{\|f\| \leq 1, \|\varepsilon_1\| \leq \varepsilon}(\|Sf - \phi(N(f) + \varepsilon_1)\|)$. This is studied in [MR].

The book under review is a comprehensive treatment of the work done in this area to date. It begins with a simple example which elucidates the issues involved in this type of complexity theory (namely those of problem formulation, information, and model of computation). This example is the binary search (twenty questions game), wherein the player has the goal of identifying an unknown element x of the interval $(0, 1)$, with the possibility of asking and getting the answer to a fixed finite number of questions with only a yes or a no. Elementary notions are identified in this setting.

In the worst case setting, one tries to use an algorithm which minimizes worst case error. In the average case setting, one is provided with a probability measure on the set of problem elements (in this case on $(0,1)$), and one tries to adopt a strategy which minimizes the average error, i.e., error

averaged over the set of all possible solutions $x \in (0, 1)$. In the probabilis-
tic setting, one is again equipped with a probability measure on $(0, 1)$, and
tries to adopt an algorithm so that the probability that the final guess is in
error by more that ε is smaller than δ, for given ε and δ. These are natural
settings, in view of the fact that sometimes we need strategies which are
guaranteed to bring us within ε units of x, and sometimes we are satisfied
with strategies which do this on the average. In addition, strategies which
seem to fail utterly in the worst case setting are sometimes quite useful
ones when viewed from the standpoint of the average-case setting.

Other settings are also considered in the context of the binary search,
including one in which information is contaminated (i.e., answers may be
erroneous sometimes).

The next chapter considers the general setting, as described earlier, in
which a map between normed linear spaces is to be approximated by a
composition of an operator of finite rank with another operator (in the
above example of binary search, this map S is the identity).

The following five chapters deal with theory and applications of the
worst case setting, average case setting, and the probabilistic setting. Also
covered in the following chapter is a comparison of these settings.

The next chapter considers the asymptotic setting, in which errors are
studied as functions of the rank of the information operator (cardinality of
information), and one seeks algorithms for which this error has the greatest
asymptotic decrease as rank becomes infinite. The final two chapters deal
with random information (as exemplified in Monte Carlo methods) and
noisy (contaminated) information.

For further concise information on the topics discussed here, the reader
is directed to two recent reviews of this area by Packel and Woźniakowski
[PW], and Packel and Traub [PT].

This book being meant for nonanalysts as well as analysts, there is at the
end an expository utilitarian section on the functional analysis necessary,
including elements of Banach and Hilbert space theory, theory of linear
operators, elementary spectral theory and measure theory, and Gaussian
measures on Banach spaces.

The book is a compilation and synthesis of a large body of work, done
over several decades (although a large part of it is more recent). To its
credit, it has simple proofs, and gives a clear and lucid exposition of the
span of the field. There are problems at the ends of sections, which vary
in difficulty. Referencing in the book is thorough, and it has an exten-
sive bibliography (about 500 references) and subject and author indices.
One useful feature which the reviewer looks for in a book is an index of
symbols; this book does not have one. However, its organized exposition
should outweigh this difficulty.

Overall, the book is an excellent introduction to an area making ad-
mirable use of functional analytic techniques to frame and analyze basic
issues in the computational sciences and theoretical numerical analysis.
It is recommended to anyone interested in more mathematical aspects of
these subjects.

REFERENCES

[CA] D. Creperley and B. Adler, *Quantum Monte Carlo*, Science **231** (1986), 555–560.

[HS] M. Hirsch and S. Smale, *On algorithms for solving* $f(x) = 0$, Comm. Pure Appl. Math. **2** (1979), 281–312.

[MR] C. Micchelli and T. Rivlin, *A survey of optimal recovery*, Optimal Estimation in Approximation Theory (C. Micchelli and T. Rivlin, eds.), Plenum, New York, 1977.

[PW] E. Packel and H. Woźniakowski, *Recent developments in information-based complexity*, Bull. Amer. Math. Soc. (N.S.) **17** (1987), 9–36.

[PT] E. Packel and J. Traub, *Information-based complexity*, Nature **327** (1987), 29–33.

[S] S. Smale, *On the efficiency of algorithms of analysis*, Bull. Amer. Math. Soc. (N.S.) **13** (1985), 87–121.

[T] J. Traub, *Iterative methods for the solution of equations*, Prentice-Hall, Englewood Cliffs, New Jersey, 1964.

[W] A. Werschulz, *Optimal error properties of finite element methods for second order elliptic Dirichlet problems*, Math. Comput. **38** (1982), 401–413.

MARK A. KON
BOSTON UNIVERSITY

The homology of Hopf spaces, by R. M. Kane, North-Holland, Amsterdam, New York, Oxford and Tokyo, 1988, xv + 479 pp., $105.25. ISBN 0-444-70464-7

In 1988 Richard Kane published his book, *The homology of Hopf spaces*. I remember at the 1986 Arcata Topology Conference (before the International Congress of Mathematicians) when Alex Zabrodsky, John Harper, Clarence Wilkerson and I received mimeographed preprints of Richard's book, we were all pleasantly surprised that someone had taken the time to amass many of the details of this growing field into a coherent book. More recently, Frank Williams commented that Kane's book will probably be the only book on finite H-spaces published in the 1980s. For myself and others who have Ph.D. students working in the area, Kane's book is an excellent first reference for many of the ideas currently used by the experts.

An H-space (or Hopf space) is simply a pointed space X, $*$ together with a binary pairing $X \times X \to X$ such that the two inclusions

$$X \times * \to X \times X \to X$$
$$* \times X \to X \times X \to X$$

are homotopic to the identity.

Mathematicians are interested in these spaces because all topological groups are H-spaces, and further, if one takes an arbitrary topological space, its loop space is an H-space. The interplay between space and loop space has been an important area of study. For example, the homotopy and homology of space and loop space are intimately related by suspension