AN INTRODUCTION TO INFORMATION-BASED COMPLEXITY*

J. F. Traub

Computer Science Department
Columbia University
November 1985

*This work was supported in part by the National Science Foundation under Grant MCS-82-14322 and by the Advanced Research Projects Agency under contract N00039-82-C-0427.

Abstract: Information-based complexity is based on three assumptions: information is partial, contaminated, and it costs. Its goal is to create a general theory about problems with such information, and to apply the results to solving specific problems in varied disciplines.

Some of the questions asked in information-based complexity are listed and some results are indicated.

Information-based complexity is contrasted with combinatorial complexity which studies problems such as the traveling salesman problem and linear programming. In combinatorial complexity information is assumed to be complete, exact, and free.

The relation between information theory and information-based complexity is briefly discussed.

Problems with only partial or limited information arise in many disciplines: in economics, computer science, physics, control theory, signal processing, prediction and estimation, scientific and engineering computation, medical imaging, geophysics, decision theory, and artificial intelligence. The goal of information-based complexity is to create a general theory about problems with partial or contaminated information, and to apply the results to solving specific problems in varied disciplines.

Traub and Woźniakowski (1980) and Traub, Wasilkowski, and Woźniakowski (1983) are research monographs on information-based complexity. Expositions and surveys may be found in Traub and Woźniakowski (1984), Traub (1985), Woźniakowski (1985), Wasilkowski (1985) and Traub and Woźniakowski (1986a).

I'll review what is meant by computational complexity.

By the computational complexity of a problem we mean its intrinsic difficulty as measured by the time, space, or other quantity *required* for its solution. For simplicity, I'll confine myself here to time complexity. Equivalently, the computational complexity of a problem is the cost of the optimal algorithm for its solution. Thus computational complexity defines optimal algorithm. For brevity I will usually refer to computational complexity simply as complexity.

Complexity is an invariant of a problem. It is independent of the algorithm, but may depend on the model of computation. See Traub (1985) for further discussion.

In general, determining the complexity of a problem is difficult. One establishes a lower bound by proving that a faster way of performing a task cannot exist; and an upper bound, which is the cost of a particular algorithm. The actual complexity is caught from above and below by these bounds. The deeper question is the lower bound, but good upper bounds are also very important.

There is sometimes confusion because people use the word complexity when they refer to the cost of an algorithm. When I say complexity I'll always mean complexity of the problem.

Basic concepts of information-based complexity include *information, partial information, contaminated information*, and *uncertainty*. I'll use an example drawn from modeling of the human visual system to illustrate these concepts.

How are humans able to see the world? The late David Marr of M.I.T. and his colleagues have developed a computational model of the human visual system, Marr and Poggio (1977), Marr (1981), Grimson (1981). I am going to give a simplified description of a small portion of this model. I want to point out, parenthetically, that similar issues arise when we design a vision system for a robot.

Imagine you're looking at an automobile. You can see what shape it has because, roughly speaking, your brain has performed a number of processes at various stages upon images it has received. For example, at one stage it has outlined the images of the various surfaces of the car by detecting the edges that separate them, such as the edge that separates the image of the windshield from the image of the hood. You can detect this edge because there's a sudden change in the slope of the surface; the window and hood do not join *smoothly*.

In the next stage, the human visual system identifies the three-dimensional shapes of the various surfaces. This stage will serve as our example.

How do you infer the shape of the hood? A depth value is the subjective distance to a point on the object as perceived by the the viewer. The model assumes that by binocularity or other means you obtain a *finite*, number of depth values. In general, between any pair of depth values the hood could have any shape. However, the visual system uses the assumption that the hood is smooth and therefore cannot change too much between depth values. (This notion of smoothness can be made mathematically precise.) Knowing the finite number of depth values and the smoothness of the surface, the visual system *approximately* determines the shape of the hood.

I'll use this example of determining the shape of the hood to introduce some fundamental concepts.

The first concept is *information*. I want to emphasize we do not mean information in the sense of Shannon, Shannon (1948). As you know, Shannon information is a certain number, the *information content* of a message, not the "information" contained in the message. Near the end of this paper I will briefly discuss connections between information theory and information-based complexity.

For present purposes, information is what is known about the problem to be solved. In determining the shape of the hood, the information is the finite number of depth values and the assumed smoothness of the surface. Because we may want to regard the surface smoothness as fixed, and study the effect of varying the depth values, we often regard the set of depth values as the information.

The number of depth values is finite. Many different surfaces may have the same depth values; there are not enough depth values to uniquely determine the surface. We say the information is limited, or *partial*.

Furthermore, the subjective distance perceived by the viewer is only an estimate of the true distance. Thus the information is *contaminated* by error.

Because the information is partial and contaminated, the problem of determining the shape of the hood can be solved only approximately. Alternatively, we can say there must be *certainty* in the answer and this uncertainty is inherently caused by the available information. It should be clear that partial or contaminated information always leads to inherent uncertainty.

As a second example, I'll use a mathematical problem. It is a simple problem, the computation of a definite integral. For most integrands we cannot compute the integral utilizing the fundamental theorem of the calculus since the antiderivative is not a "simple" function. We have to approximate the integral numerically. Usually, the integrand is evaluated at a finite number of points. The information is the values of the integrand at these points. Since an infinite number of integrands have the same values at these points, the information is partial. The integral is estimated by combining the integrand values. In addition, there will be round-off error in evaluating the integrand, and therefore the information is contaminated. Since with the information we're using we don't know the integrand, there is intrinsic uncertainty in the answer.

This example differs from the previous one in that we started with complete and exact information. The integrand was specified exactly as a function. But we couldn't use that information to solve our problem. We had to throw away our complete and exact information and replace it by partial and contaminated information.

Problems with partial and contaminated information arise elsewhere. In medical imaging, such as a CAT scan, we approximate the shape of internal organs from certain density measurements. In geophysical exploration, we estimate the location of mineral deposits or oil reservoirs by measurements taken at the earth's surface, Twomey(1977).

Two of the basic assumptions of information-based complexity are that information is partial and contaminated. There is one further assumption — information costs. For example, in mineral exploration, a seismologist might set off explosions whose effect is measured by sensors. That's an expensive process. THESE THREE ASSUMPTIONS ARE FUNDAMENTAL: INFORMATION IS PARTIAL, INFORMATION IS CONTAMINATED, AND INFORMATION COSTS.

I'll contrast that with the usual assumptions concerning the traveling salesman problem. The available information specifies just one task. That is, all the intercity distances are given; therefore the information is *complete*. Furthermore, the intercity distances are assumed to be given without error. Therefore the information is *exact*. Finally, the intercity distances are provided without charge. Thus the information is *free*. For the traveling salesman problem then, information is assumed to be complete, exact, and free. These assumptions are also made for other important problems. For example, for linear programming the information is the matrix and linear cost function.

Computational complexity is the study of the minimal cost for solving a problem. Computational complexity comes in two flavors, depending on which set of assumptions are made about information. Is information complete, exact, and free, as with the traveling salesman problem? We call this combinatorial complexity. Or is the information partial, contaminated, and priced, as with a host of other real-world problems?

What gives information-based complexity its particular flavor is based on the observation that the quality and amount of information determine the irreducible uncertainty in the solution. This irreducible uncertainty does not depend on any notion of algorithm. This observation enables us to decompose the problem of determining the computational complexity of solving a problem into an information phase and a combinatory phase. It's a very powerful principle.

Thus information-based complexity has both an *informational level* and a *combinatorial level.*

At the information level we answer questions such as:

- What is the intrinsic uncertainty in the solution of a problem due to the available information?
- How much information is needed to limit the uncertainty to a specified level? For example, in geophysical exploration, how many measurements must the seismologist make?
- What is the optimal information? Again, using our geophysical exploration example, where on the earth's surface should the seismologist place his instruments?
- Can the information be decomposed for parallel or distributed computation?

At the combinatorial level we answer questions such as:

- What is the minimal cost of combining the information to obtain an answer?

The central question of information-based complexity is the following:

- What is the computational complexity of solving a problem for a given level of uncertainty?

Answering this question requires both the information and combinatorial levels.

The answers to these questions depend on the setting. In both combinatorial complexity and information-based complexity we judge algorithms by their performance in various settings including worst case, average case, and probabilistic. I want to indicate some of the settings studied in information-based complexity.

In a *worst case* setting we want to minimize the cost while guaranteeing a solution for the most difficult problem instance. This is sometimes called a minimax criterion.

In an *average case* setting we want to minimize the expected cost while guaranteeing the average error.

In a *probabilistic* setting we want to minimize the expected cost while requiring that the probability of a large error is small. If a problem has a yes or no answer, then we require that the probability of a wrong answer is small.

Other settings are also of interest in information-based complexity. In a study of clock synchronization in distributed networks Wasilkowski (1985) uses a worst case setting in which he averages over the delays. In a study of whether adaptive information is more powerful than nonadaptive information for stochastic information, Kadane, Wasilkowski, and Woźniakowski (1984) minimize the worst case cost, guarantee the expected error, and average over the information error which is assumed to be stochastic.

In another setting of interest we want to minimize the expected cost for the most difficult problem instance. Such a setting is of interest for problems such as aircraft design and structural design for earthquakes.

A discussion of these and other settings may be found in Traub and Woźniakowski (1986a).

At the beginning of this paper I said that the goal of information-based complexity is to create a general theory about problems with partial or contaminated information, and to apply the results to solving specific problems in varied disciplines. I now will talk about several general results and then briefly indicate applications.

I'll discuss results concerning questions one and four at the information level.

Recall that the first question was:

– What is the intrinsic uncertainty in the solution of a problem due to the available information?

There exists a quantity which measures this intrinsic uncertainty. The radius of information is fundamental, not only to answering the first question but also to answering some of the other questions I posed. To emphasize this I will state the information principle: *There exists a quantity called the radius of information which measures the intrinsic uncertainty of solving a problem for specified information.*

This radius of information exists in every setting we study. It exists in the worst case, average case, and probabilistic settings.

In principle, one can always compute the radius of information for any problem and any specified information. For a particular instance, it may be technically very difficult to do this. Of course, we're used to that. For example, in principle one need only follow some basic prescriptions of quantum mechanics to solve certain problems of interest to physicists and chemists, but it's quite a different matter to actually follow the prescription, either symbolically or numerically, for a particular instance.

Next I'll discuss question four. The fourth question is:

– Can the information be decomposed for parallel or distributed computation?

One important way to solve a problem on a parallel or distributed system is to decompose the problem. That will be a subject for investigation for as far ahead as I can see. In particular, can the task of obtaining information be decomposed? That's the motivation for the above question.

I'll use a simple example to introduce the notions of sequential (adaptive) and parallel (nonadaptive) information. Consider estimating the value of an integral using as the information, values of the integrand at a finite number of points.

Imagine I've evaluated the integrand at a number of points and therefore have some idea of the shape of the curve. Can I take advantage of that knowledge to choose the next point for evaluation? For example, should I evaluate the integrand where its slope is large, that is, where its shape is changing rapidly?

The strategy I've just described uses sequential information. It is sometimes called *adaptive* information. *Parallel* information, on the contrary, means that I perform all the evaluations simultaneously without having any opportunity to take advantage of what we learn about the shape of the curve.

Those are informal descriptions of sequential and parallel information. Precise descriptions may be found in the literature.

The desirable results for distributed or parallel computation is that sequential information is not more powerful than parallel information; that adaption does not help. To see this for the integration example, assume I have p processors and that it takes a processor one unit of time to evaluate a function at a point. Then n functions evaluated sequentially cost n units of time, whereas n functions evaluated simultaneously, that is, in parallel, cost $\lceil n/p \rceil$ units of time. Thus the cost of collecting information is reduced by a factor of p, which is linear speed-up. Since the cost of evaluating the integrand dominates the cost of combining evaluations to produce an answer, the speed-up in computing an approximation to the integral is close to linear in the number of processors. See Traub and Woźniakowski (1984, 1984a, Section 4.5) for further discussion of speed-up.

The above discussion is predicated on the assumption that adaption does not help. However, intuition suggests that adaption does help.

The question of whether adaption helps was first analyzed for a worst case model. If a problem is linear, adaption does not help. Examples of linear problems are linear functionals such as intergration and differentiation, approximation, signal recovery, and linear partial differential equations. The vision example I described earlier is a special case of approximation. See Traub and Woźniakowski (1980), Traub, Wasilkowski, and Woźniakowski (1983).

We believed, along with others, that the result that adaption doesn't help, was due to the worst case model. To our surprise we were able to show that adaption does not help on the average. This was established in great generality. It is enough to assume that the problem is specified by a linear operator and that the distribution of problem instances is governed by a measure with a certain symmetry property, such as a Gaussian or Wiener measure. See Traub, Wasilkowski, and Woźniakowski (1984a), Wasilkowski and Woźniakowski (1984), Lee and Wasilkowski (1985), and Wasilkowski (1985).

What if the information is contaminated with stochastic error? Kadane, Wasilkowski, and Woźniakowski (1984) recently showed that again adaption does not help, provided only that the information is unbiased.

In all the work cited above the problem is specified by a linear operator. Perhaps adaption helps generally for nonlinear problems. However, Sikorski (1984) has shown that adaption does not help for the approximate solution of nonlinear equations satisfying a Lipschitz condition.

I must emphasize that there are, of course, many problems for which adaption helps. (See Traub, Wasilkowski, and Woźniakowski (1984b) for an especially constructed example. We are currently exploring linear problems, arising in practice, where adaption helps.) We *have*, however, identified many important problems for which we can prove that it does not and for which the answer to the question "Can the information be decomposed for parallel or distributed computation?" is yes.

I want to turn to applications of information-based complexity beginning with a general discussion of distributed systems.

There are two reasons for distributed systems. The first is that although a centralized system could be used, we select a distributed system for the sake of, say, efficiency. The second reason is that the problem is inherently distributed; examples include resource allocation in a decentralized economy, traffic networks, and reservation systems.

Consider now a large distributed system. One possibility is that the total system has complete information but the nodes have only local information, say, about themselves and their neighbors. Thus the information is distributed over the system. To give the nodes information about the total system would cost too much in time and/or space. Thus, decisions are made at nodes which have only partial information and that means a solution with uncertainty. In a dynamic system, even if complete information is initially available at the nodes, we cannot afford to update that information over time.

So far, I've assumed the total system has complete information. Of course, often even the total system has only partial or contaminated information and what I've said holds in spades.

An application of this discussion is that even the problems that are now exactly solved on a uniprocessor will be only approximately solved in the distributed environments of the future.

Let me return to the present. If a user wished to apply information-based complexity he provides just three items. They are;

1. A mathematical formulation of his problem and the setting (say, average case) in which he's interested.
2. The available information.
3. The model of computation; that is what information operations and combinatory operations are permitted and how much they cost.

That's all the user provides. The theory then provides everything else. For examples, the theory tells the user such things as how much information is needed to limit the uncertainty to a specified level, whether the information can de decomposed for parallel or distributed computation, the optimal algorithm, and the computational complexity of the problem.

At the beginning of this talk I indicated some of the disciplines for which only partial or limited information is available and to which information-based complexity can therefore be applied. We indicate some of the application areas which are currently being pursued. For each application we list one recent paper or book, often a survey. An extensive discussion of applications may be found in Traub and Woźniakowski (1986a).

Vision, Medical Imaging, Optimal Flow

Lee (1985)

Synchronization of Clocks in Distributed Systems

Wasilkowski (1985)

Nonlinear Constrained Optimization

Nemirovsky and Yudin (1983)

Prediction and Estimation

Milanese, Tempo and Vicino (1986)

Partial Differential and Integral Equations

Werschulz (1985a)

Ordinary Differential Equations

Kacewicz (1984)

Nonlinear Equations

Sikorski (1985)

Integration

Traub and Woźniakowski (1980, Section 6.4)

Topological Degree

Boult and Sikorski (1986)

Ill-Posed Problems

Werschulz (1985b)

Large Eigenpair Systems

Kuczynski (1986)

Large Linear Systems

Traub and Woźniakowski (1984b).

A natural question concerns connections between information-based complexity and information theory. The concepts of information and entropy as developed in thermodynamics by Boltzmann and Clausius and in information theory by Shannon are among the most significant and fruitful scientific ideas of the last hundred years. It seems to me, however, that the kind of questions were posed that posed earlier are not addressed by information theory. Indeed, as indicated above, information in the sense of Shannon is a certain number, the information content of a message, not the "information" contained in the message. It turns out, however, that there are connections between certain notions of entropy and information-based complexity. For example, Traub and Woźniakowski (1980, Section 7.4) have reported on the relation between Kolmogorov's notion of ε-entropy and our second question at the information level on how much information is needed to limit the uncertainty to a specified level.

More generally, uncertainty and information are basic concerns of a number of disciplines including information theory, statistics, economics, and control. Various measures of uncertainty and information have been proposed. For discussions see, for example, Brillorien (1956), Fisher (1950), Gallager (1968), Kullback (1961), Marschak and Radner (1972), Pinsker (1964), Shannon (1948), Theil (1967), and Wiener (1948).

In information-based complexity the radius of information measures the uncertainty. In Traub and Woźniakowski (1986b) we compare and contrast various measures of uncertainty and information including Shannon entropy, Kullback-Leibler entropy, Fisher information, variance and radius of information.

The aim of information-based complexity is to find unity across a great variety of what seem, at first, to be totally different disciplines. I believe we have made a beginning.

Bibliography

[1] Boult, T. and Sikorski, K. (1986), Complexity of Computing Topological Degree of Lipschitz Functions in n Dimensions, to appear in ***Journal of Complexity*, 2**.

[2] Brillorien, L. (1956), *Science and Information Theory*, Academic Press, New York.

[3] Fisher, R. A. (1950), *Contributions to Mathematical Statistics*, John Wiley and Sons, New York.

[4] Gallager, R. G. (1968), *Information Theory and Reliable Communication*, John Wiley and Sons, New York.

[5] Grimson, W. E. L. (1981), *From Images to Surfaces: A Computational Study of the Human Early Visual System*, MIT Press, Cambridge, Mass.

[6] Kacewicz, B. Z. (1984), How to Increase the Order to Get Minimal-Error Algorithms for Systems of ODE, *Numer. Math.*, **45**, 93–104.

[7] Kadane, J. B., G. W. Wasilkowski, and H. Woźniakowski (1984), Can Adaption Help on the Average for Stochastic Information?, *Report*, Columbia University.

[8] Kolmogorov, A. N. and V. M. Tihomirov, ε-Entropy and ε-Capacity of sets in Functional Spaces, *Usp. Mat. Nauk.* **14**, 3–80 (in Russian), (1959). Translation in *Amer. Math. Soc. Transl.* **17**, 277–364, (1961).

[9] Kuczynski, J. (1986), On the Optimal Solution of Large Eigenpair Problems, to appear in *Journal of Complexity*, **2**.

[10] Kullback, S. (1961), *Information Theory and Statistics*, John Wiley and Sons, New York.

[11] Lee, D. (1985), Optimal Algorithms for Image Understanding: Current Status and Future Plans, *Journal of Complexity*, **1**, 138–146.

[12] Lee, D. and G. W. Wasilkowski (1985), Approximation of Linear Functionals on a Banach Space with a Gaussian Measure, *Report*, Columbia University. To appear in *Journal of Complexity*, **2** (1986).

[12] Marr, D. (1981), VISION: *A Computational Investigation in the Human Representation and Processing of Visual Information*, W. H. Freeman, San Francisco.

[13] Marr, D. and T. Poggio (1977), From Understanding Computation to Understanding Neural Circuitry, *Neuroscience Research Program Bulletin* **15**, 470–488.

[14] Marschak, J. and R. Radner (1972), *Economic Theory of Teams*, Yale University Press, New Haven.

[15] Milanese, M., R. Tempo, and A. Vicino (1986), Strongly Optimal Algorithms and Optimal Information in Estimation Problems, to appear in *Journal of Complexity*, **2**.

[16] Nemirovsky, A. S. and D. B. Yudin (1983), *Problem Complexity and Method Efficiency in Optimization*, A Wiley-Interscience Publication, New York.

[17] Pinsker, M. S. (1964), *Information and Information Stability of Random Variables and Processes*, Holden-Day, San Francisco.

[18] Shannon, C. E. (1948), Mathematical Theory of Communication, *Bell Syst. Tech. J.* **27**, 379–423, 623–658.

[19] Sikorski, K. (1984), Optimal Solution of Nonlinear Equations Satisfying a Lipschitz Condition, *Numer. Math.* **43**, 225–240.

[20] Sikorski, K. (1985), Optimal Solutions of Nonlinear Equations, *Journal of Complexity*, **1**.

[21] Theil, H. (1967), *Economics and Information Theory*, North-Holland, Amsterdam.

[22] Traub, J. F. (1985), Information, Complexity, and the Sciences, University Lecture, Columbia University.

[23] Traub, J. F., G. W. Wasilkowski, and H. Woźniakowski (1983), *Information, Uncertainty, Complexity*, Addison-Wesley, Reading, Mass.

[24] Traub, J. F., G. W. Wasilkowski, and H. Woźniakowski (1984a), Average Case Optimality for Linear Problems, *Journal TCS*, **29**, 1–25.

[25] Traub, J. F., G. W. Wasilkowski, and H. Woźniakowski (1984b), When is Nonadaptive Information as Powerful as Adaptive Information, *Proceedings of the 23rd IEEE Conference on Decision and Control*, 1536–1540.

[26] Traub, J. F. and H. Woźniakowski (1980), *A General Theory of Optimal Algorithms*, Academic Press, New York, NY.

[27] Traub, J. F. and H. Woźniakowski (1984), Information and Computation, chapter in *Advances in Computers* **23**, M. C. Yovits, editor, Academic Press, New York 35–92.

[28] Traub, J. F. and H. Woźniakowski (1984a), On the Optimal Solution of Large Linear Systems, *Journal of the ACM*, **31**, 545–559.

[29] Traub, J. F. and H. Woźniakowski (1986a), Information-based Complexity, to appear in *Annual Review of Computer Science* **1**, Annual Reviews, Inc. Palo Alto.

[30] Traub, J. F. and H. Wozniakowski (1986b), Measures of Uncertainty and Information, in progress.

[31] Twomey, S. (1977), *Introduction to the Mathematics of Inversion in Remote Sensing and Indirect Measurement*, Developments in Geomathematics **3**, Elsevier Scientific Publ., Amsterdam.

[32] Wasilkowski, G. W. (1985), Average Case Optimality, *Journal of Complexity* **1**, 107–117.

[33] Wasilkowski, G. W. (1985), Clock Synchronization Problem with Random Delays, *Report*, Columbia University.

[34] Wasilkowski, G. W. and H. Woźniakowski (1984), Can Adaption Help on Average?, *Numer. Math.*, **44**, 169–190.

[35] Werschulz, A. G. (1985a), Complexity of Differential and Integral Equations, *Journal of Complexity*, **1**.

[36] Werschulz, A. G. (1985b), What is the Complexity of Ill-posed Problems?, *Report*, Columbia University.

[37] Wiener, N. (1948), *Cybernetics*, John Wiley and Sons, New York.