Dynamic Programming

Richard Bellman

Man is a decision-making animal and, above all, conscious of this fact. This self-scrutiny has led to continuing efforts to find efficient ways of behaving in the face of complexity and uncertainty. Aristotelian logic and probability theory, psychology and psychiatry—these are some of the by-products of the efforts to understand conscious and subconscious decision making. Over the last 25 years there has been a vast intensification of research in this intriguing area, as well as a more widespread recognition of its explicit and implicit role in all segments of society. The many varieties of planning and programming required to overcome, circumvent, or neutralize the myriad intricacies of engineering and social systems have generated a number of new mathematical theories. Dynamic programming is one of these, a child of its century. It is dedicated to the study of multistage decision processes, processes where a sequence of decisions over space and time is required. Making full use of the unique "as if" attitude of mathematics, the theory can be equally well applied to situations which can be profitably construed as multistage decision processes.

One of the most important types of problems requiring a sequence of decisions is that of the control of a system. It may be a matter of harnessing the tides, of allocating hydroelectric power, of preventing a chain reaction of power failure, of cutting and planting trees, of breeding disease-resistant cattle, or of destroying a pest which menaces the trees. Alternatively, it may be the management of a university endowment program that is involved, or the maintenance and replacement of a fleet of taxicabs. What is common to all these operations, and to many other control and decision processes, is the fact that they are not one-shot affairs. On the contrary, many interconnected observations and actions are required. The procedure is as follows. One examines the system with an eye to its needs and obligations. To fulfill its needs and satisfy its obligations a number of courses of action are available, one of which must be chosen. At a subsequent time, a further examination is made to determine the effect of this decision and to obtain the information required for the next decision. This combination of observation, interpretation, and decision is then repeated, in some cases indefinitely.

Dynamic programming supplies a systematic technique for determining what information is required and how it should be used effectively. The applications of this mathematical theory have grown constantly in number as the capabilities of the digital computer to store, process, and retrieve information have grown in power.

Feedback Control

To illustrate some of the foregoing ideas, let us consider the guidance of a space vehicle from its blast-off to its soft landing on the moon. Through utilization of available astronomical data in Newton's equations of motion, an appropriate path is plotted (Fig. 1).

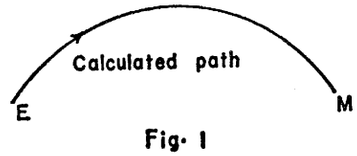

Fig. 1

Let us suppose that this path has been determined by the requirement that it be a path requiring the least flight time. The determination can be made by any of a number of different methods, although it is by no means a trivial matter. Despite this theoretical success, there are difficulties as far as actual flight is concerned. After all, a spacecraft is not the point particle of Newtonian mechanics. It is a huge, unwieldy system possessing different components with different propensities for malfunctioning or for behaving in a fashion different from a prediction based upon a simpler, idealized system. For a variety of reasons the rocket may begin to wobble and may ultimately deviate considerably from the plotted path, as indicated in Fig. 2.

The space probers are, of course, aware of this possibility. From the time of launching, the position and velocity of the vehicle are carefully monitored, the actual and calculated paths being constantly compared. At a suitable time, a control mechanism is actuated to force the vehicle back to the desired trajectory. This is an example of feedback control. The amount of effort required to carry out the correction depends on the extent of the deviation from the planned flight. The monitoring and control continue until the voyage has been successfully concluded. The final part of the control process involves adjusting the velocity so that the giant ship lands on the surface of the moon with minimum impact.

Difficulties

Guidance of the spacecraft is effected through a sequence of observations and decisions. Should a correcting force be exerted, and, if so, how? The theory of feedback control is quite elegant, and we are surrounded by examples of its successful application. Nevertheless, severe difficulties can arise in practice. One source of these is discussed in the following extension of the spacecraft illustration. The mathematical problem of neutralizing midcourse meandering is a difficult one. In order to render it more tractable, the mathematician considers the case where only a small amount of straying is allowed. Most of the conventional theory of feedback control is based upon this reasonable initial hypothesis, which allows the use of linear equations and thus the use of the full power of classical analysis. As a consequence, we have available a number of techniques for control which are quite effective provided nothing untoward occurs. Suppose, however, that the well-known perversity of inanimate objects begins to operate and a number of small malfunctions accumulate to produce a serious deviation from

The author is professor of mathematics, electrical engineering, and medicine, University of Southern California, Los Angeles.

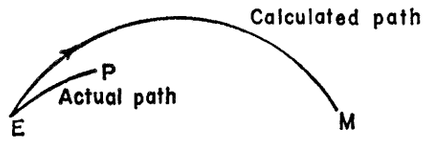

Fig. 2

the calculated trajectory. Suppose, further, that corrections based upon small-deviation theory are made. Then something like the zig-zag course shown in Fig. 3 can occur. If the oscillations around the calculated path are considerable, the time required for the spacecraft to traverse the corrected path may be much greater than the time originally contemplated, and the vehicle may run out of fuel. Moreover, the situation may become similar to that of a beginner attempting to ride a bicycle. The correcting influences become greater and greater and occur at shorter and shorter intervals until the trajectory becomes so erratic that no correction is possible. This is the ultimate in instability.

The source of difficulty has been a much too narrow use of the concept of feedback control. Instead of concentrating on the original goal, that of traveling from the earth to the moon in minimum time, we foolishly focused on the subgoal of adhering rigidly to the calculated trajectory. At *P* of Fig. 2 we should have recognized that we were quite far from the original path and should then have plotted a new path starting from *P* and ending at the moon (see Fig. 4).

If it turns out that the vehicle is forced off this new path by further malfunctions, we should repeat the foregoing procedure. We should locate the current position in space and then chart the course which minimizes the time required to get from this position to the moon. This is a simple commonsense approach to the control of a system, and to multistage decision making in general. We do the best we can starting from where we are.

Policy Concept

This is the essence of dynamic programming. Basic to this procedure is the concept of a policy, a rule for telling what decision to make in terms of the current position of the system. The major advantage of this new control concept over the classical ideas of control lies in its flexibility. We are prepared for all eventualities. No matter what the current position, a policy informs us what control to exert. No longer are we bound by preconceived notions of the nature of the most desirable path. As we see below, implicit in the idea of a policy is the basic notion of learning from experience.

Multistage decision making is regarded as the repeated application of a policy. A policy which is most efficient in the sense of minimizing time, or fuel, or cost or of maximizing profit is called an optimal policy.

These optimal policies can be quite simply characterized by means of an intuitively derived "principle of optimality": an optimal policy has the property that, whatever the initial state and initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision.

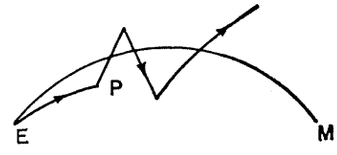

Fig. 3

This is a slightly more abstract formulation of the commonsense approach discussed above. The translation of this formulation into mathematical terms provides us with equations which enable us to determine the optimal policies.

Application

The very flexibility of the dynamic programming approach generates difficulties in its application. As indicated above, a policy is a rule which determines the optimum decision ("decision" being equivalent to "control") in every conceivable state of the system. Since complicated systems can exist in a very large number of conceivable states, it is necessary to store a considerable set of possible decisions for future use. Consequently, until very recently the limited storage and retrieval capacities of digital computers severely restricted the application of dynamic programming.

Thus, for example, the largest computers commercially available a few years ago could, at most, store and retrieve quickly 64,000 ten-digit numbers. This seems to be an enormous capacity until we make a brief count of the possible positions of a space vehicle. When the vehicle is idealized as a point, there are three position coordinates and three velocity coordinates, the usual six-dimensional phase space. An allowance of ten possible values for each of these coordinates yields 10^6 possible positions of the system.

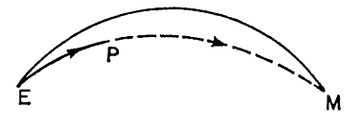

Fig. 4

There are, of course, more efficient ways of storing a policy. Nevertheless, I hope I have indicated some of the difficulty encountered in describing a policy for a complex system. Fortunately, with the current generation of computers the situation has improved enormously. The number 64,000 has been increased to 10^6, and even to 2×10^6, while operating times have decreased. We can expect 10^6 to be replaced, within 10 years, by 10^8, with the time of operation cut again by a factor of at least 100. All this means that dynamic programming methods will be readily applicable to many of the vital engineering, economic, and industrial systems of our society.

What is extremely important is the fact that these techniques in many cases do not require as much advanced mathematical training as the classical methods do. Many of the classical theories were constructed to avoid the then impossible task of making large-scale arithmetic calculations. As the computer is developed, and as new mathematical ideas based upon the ability of the computer to carry out billions of simple instructions become widely known, we will see an elimination of the mathematical middleman throughout many economic, engineering, and scientific domains. The mathematician will thus be a typical victim of automation and sophistication. Fortunately, of course, he has more than enough wilderness to which to retreat. For example, major research efforts are required to reduce combinatorial problems and questions in statistical mechanics to a formulation where a computer can be used.

Multistage Decision Making

Let us now consider some more-complicated types of decision making. With all the difficulties we conjured up, we nevertheless made certain simplifying assumptions.

1) We can accurately determine the state of the system at any time; or, equivalently, knowing where we are, we can determine the time.

2) We know when the control is exerted and can accurately predict its effect.

For a variety of reasons, some theoretical, some operational, all these assumptions are idealizations. This does not mean that they are not useful. It merely means that we should be prepared to modify them when the important simplification they provide intellectually begins to limit us scientifically in our ability to understand and predict.

Let us consider here only the case where cause-and-effect predictions have to be modified. If we cannot predict exactly what the effect of a decision is going to be, or predict it accurately enough, clearly the problem of effective decision making becomes one of a higher order of difficulty. To begin with, it isn't even clear what one means by "effective" or "optimal."

Fortunately, there already exists a mathematical theory devoted to the study of unpredictable effects, the theory of probability. Naturally, this theory does not cover all kinds of chance events, only carefully chosen types of violations of certainty. And, as long as we are forcing ourselves to study decision making under conditions of uncertainty, let us return to a problem area vastly more entertaining than that of space travel—the area of gambling systems. We can, if we desire, protect ourselves against the charge of frivolity with the observation that the mathematics involved is abstractly identical to that involved in the actuarial activities of insurance companies, in the investment plans of Wall Street, and in divers questions arising in reliability theory, inventory theory, and so forth.

Let us then leave the space vehicle in midcourse and turn to the dilemma of the casual visitor to Las Vegas who would like to cover his expenses by means of some judicious wagers. Intuitively, we feel that the amount he bets on each turn of the wheel or roll of the dice should depend upon how much money he has and what his objective is. Once again, it is clear that what is required in a gambling system is a policy, a rule that tells the gambler what bets to make, and, more generally, what decision to make, in every possible situation. Indeed, dynamic programming has been used with success in the game of blackjack, and in other gambling as well.

A flexible policy is thus ideally suited to the exigencies of a multistage decision process involving chance events. What is esthetically satisfying is the fact that the principle of optimality provides a mathematical apparatus for treating deterministic and probabilistic decison processes in a uniform fashion. The term *stochastic* is often used to describe a process involving chance events, precisely because it is a word with no daily-life connotations.

Adaptive Control

The examples we have considered so far may be called conventional decision processes, inasmuch as they have involved the following further tacit assumption.

1) We know all the basic variables required to describe the system.

2) We are familiar with all possible decisions.

3) We know the general structure of cause and effect in either the deterministic or the stochastic sense.

4) The overall objective of the decision or control process is clearly and precisely defined.

In a number of important situations (in fact, in all decision-making situations), we face the problem of making decisions without a full knowledge of the basic workings of the underlying system. This is the state of affairs in running a major industry, in making policy in the economic and military spheres, in designing experiments, and in doing research in general. Consider, for example, the matter of constructing a gambling system without knowing whether the dice are loaded. We would prefer to take time out to obtain the missing data, but the rules of the game do not allow it. It is necessary to learn and act at the same time. We start with certain preconceptions of the nature of an optimal policy and then systematically modify this policy on the basis of experience. This is called adaptive control. There is obviously an intimate connection between these concepts and what the psychologist calls adaptation.

In dealing with large complex systems we encounter the formidable problems of deciding what information is to be used in decision making, how the maximum amount of information can be extracted from small samples of data, what possibly can be learned from further experimentation and observation, and how the available information can be used to modify initial policies. Not only do we have to decide upon the allocation of time and other resources to the control activity; we also must decide how much time and effort to devote to studying the intrinsic nature of the system.

The concept of a policy extends to this wider class of decision and control processes, with the proviso that a policy now tells us what to do in terms of where we are—and what we know. When "information" is taken to include a set of additional state variables, the principle of optimality can be used to obtain a precise mathematical formulation of adaptive control processes. The mathematical theory is on a much higher conceptual and analytic level than that for the deterministic and stochastic processes described above. Despite all the power of current mathematics and the power of computers now in the design stage, we cannot expect mastery in these areas until sometime far in the future. This is an excellent example of a scientific area requiring a great deal of sophisticated conceptualization and formulation before any arithmetic can be done.

As stated above, in an adaptive control process a decision affects not only the position of the system but also our information about the system. The study of these processes thus forces us to analyze in detail what we mean by "information" and by "learning." This is nowhere clearer than in the attempt to write a computer program for an adaptive process. Much to his distress, the mathematician is confronted with the problem of analyzing "thinking."

Hierarchy of Decision Making

As might be expected, there is no simple, or even unique, explanation of this phenomenon of the human mind. Let us consider ways in which a mathematician can approach this thorny subject. Do machines think? It is not surprising that a great deal of controversy, much of it emotional and visceral, has arisen over this issue. What is surprising is that many people who should know better are not aware of the fact that the question is devoid of meaning. Until we have defined what we mean by a "machine," what is meant by "think," and, particularly, what is meant by "can," all we can agree on is

that a question—some question—is implied. In fact, as we shall see, many questions are implied.

To begin with, by "machine" we shall mean a digital computer of the type commercially available. Others may mean a different type of computer or device, or even a hypothetical computer. This is their prerogative. In order to define "think," we shall employ the device used above in connection with decision making. Rather than attempt to define thinking in abstract terms, we consider it only in connection with certain carefuly delineated types of decision processes. We then equate levels of thinking processes with levels of decision-making processes. Once this has been accomplished, it is meaningful to ask if we are capable of writing a digital computer program that can carry out a specific decision-making process in a stated time. Thus, "can" has different meanings, depending upon whether we allow a time of 2 minutes, 2 hours, or 2 years or merely require that the time be finite, although not predictable.

For example, in connection with the recognition of handwriting in a banking firm, or the analysis of x-rays or tissue cultures, a time of more than 2 minutes may make the ability of a computer merely a mathematical curiosity. That a computer can be programmed to play legal chess or checkers is hardly remarkable; that it can learn from experience to play master checkers is interesting and represents a feat of programming. It cannot, at present, learn from experience to play master chess, and an ability of this type would represent an astounding breakthrough in the theory of adaptive processes. That a computer can produce music which we recognize as a poor imitation of Beethoven, Mozart, or Bach is hardly remarkable; that it can produce an accidental tune among countless cacophonies is again to be expected. If it could systematically produce beautiful music —that is, "create"—this would represent a far more astounding breakthrough in the intellectual domain. Even if these breakthroughs should be achieved, we would not presume to say that we understood what goes on in the human brain. We are asking for duplication only of the result, not of the process. Again, it is anyone's prerogative to attempt to perform these feats on the basis of an analysis of what the human mind does, but in the opinion of most people, in view of what little is known, such an attempt represents a foolhardy effort.

Perhaps the essential point I am trying to make is that there are levels of decision-making processes, and that it is therefore important to find systematic ways of cataloging these processes. The idea of introducing categories of processes is borrowed from the method used by Russell in mathematical logic, the theory of types. This theory enables one to construct hierarchies of statements.

To give an example of how a hierarchy can be introduced, let us begin by introducing processes on the first level. These are processes of the deterministic or stochastic type described above. On the second level, we consider processes involving learning about the structure of the system. A local policy is required for making decisions at each stage, and a global policy is required for modifying the local policy on the basis of experience. This is decision making about decision making. Choosing a global policy is decision making about decision making about decision making. We can now continue in this fashion. Needless to say, a certain amount of effort is required to formulate this hierarchy precisely.

This is one particular way of introducing a hierarchy. There are other ways, and there are always problems outside any particular formulation. Where, for example, do we place the problem of ascertaining the level of a specific decision-making process?

Conclusion

Little has been done in the study of these intriguing questions, and I do not wish to give the impression that any extensive set of ideas exists that could be called a "theory." What is quite surprising, as far as the histories of science and philosophy are concerned, is that the major impetus for the fantastic growth of interest in brain processes, both psychological and physiological, has come from a device, a machine, the digital computer. In dealing with a human being and a human society, we enjoy the luxury of being irrational, illogical, inconsistent, and incomplete, and yet of coping. In operating a computer, we must meet the rigorous requirements for detailed instructions and absolute precision. If we understood the ability of the human mind to make effective decisions when confronted by complexity, uncertainty, and irrationality, then we could use computers a million times more effectively than we do. Recognition of this fact has been a motivation for the spurt of research in the field of neurophysiology.

The more we study the information-processing aspects of the mind, the more perplexed and impressed we become. It will be a very long time before we understand these processes sufficiently to reproduce them.

In any case, the mathematician sees hundreds and thousands of formidable new problems in dozens of blossoming areas, puzzles galore, and challenges to his heart's content. He may never resolve some of these, but he will never be bored. What more can he ask?