Development of direct-search strategies in hill-climbing problems

KENT L. NORMAN
University of Maryland, College Park, Maryland

Hill-climbing problems were presented to undergraduate students to investigate search strategies for finding a maximum point on the surface. Subjects guessed the values of two variables, X and Y, that could result in the maximum value of a criterion variable Z. The function relating X and Y to Z was an inverted paraboloid. Subjects tended to use a systematic strategy in which they adjusted only one variable at a time in order to increase Z. When a failure to increment Z occurred with one variable, they switched to the other variable and adjusted it. Problems in which the paraboloid was highly elliptical were more difficult than ones with spherical paraboloids, especially when the axes of ellipse were not lined up with the X,Y axes. Direct-search strategies are discussed in terms of their efficiency with respect to the number of moves to maximization and cognitive effort.

A wide variety of problems involve a search through a problem space for a state whose evaluation is at the maximum point. In formal problem theory, these are known as hill-climbing problems. Given the initial state in the problem space, one attempts to select a move that increments the evaluation of the state. A trivial example of this type of problem is the children's game of "hot-cold." The child is blindfolded, and an object is placed in the room. The child's search for the object is aided by the other children's crying "warmer" as he or she approaches the object and "colder" as he or she withdraws from it. Hill climbing as a strategy in human problem solving has been studied by Newell and Simon (1972) in subject protocols. Others have suggested that this is a useful strategy in common problem-solving tasks. For example, Wickelgren (1974) discusses the strategy in his book on how to solve problems. Although hill climbing has been suggested as a strategy for a number of problems, the psychological processes used to solve the problem of hill climbing itself have escaped consideration.

Simple nonlinear optimization problems provide a direct task for studying hill climbing. Points on a surface are specified by X,Y coordinates, and the height of the surface at some point is Z = f(X,Y). Figure 1 shows a nonlinear hill-climbing problem in three-dimensional space. The only information that the problem solver has is the height of the surface at each point that he tests. From the information
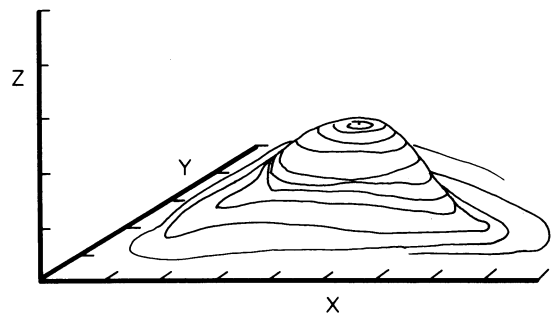
Figure 1. A hill-climbing problem represented in three-dimensional space.

that he gleans from these points, he is to search for the maximum.

A number of computer algorithms have been devised to locate the maximum (Dixon, 1972). In general, these procedures test nearby points to see if their height is greater than the current height. If so, they move to that point and repeat the search. This process produces an ascending path to the maximum point. Of course, strategies differ in terms of their efficiency and speed. For example, one might want to use a strategy that requires a minimum number of moves in order to find that maximum. One such strategy is called STEPIT, devised by Chandler (1969) from the Hook and Jeeves (1961) pattern-search algorithm. This procedure starts by making exploratory moves in order to find an increasing gradient. It then makes pattern moves in that direction until it reaches an unsuccessful move. Having encountered a failure, it returns to making exploratory moves and repeats the process until it can no longer find successful exploratory moves. It then concludes that it has found the maximum point.

It is hypothesized that human subjects working on optimization problems also develop systematic strategies. Although these strategies result in solution, it is predicted that they will not be efficient in the sense of minimizing the number of moves. Rather, they represent a balance of conserving cognitive processing at the expense of efficiency. The experiment reported here varied the difficulty of the problem by varying the shape of the surface as well as the amount of practice.

METHOD

Subjects

Ten (six male and four female) undergraduate students were recruited from introductory computer science and introductory statistics classes at the University of Maryland. The subjects had had no prior exposure to strategies of optimization.

Task

The subjects were told that they were to search for the X,Y location that resulted in a Z value of 10,000 in as few moves as possible. On each trial, the subjects supplied values for X and Y. The allowable values for X and Y were integers from 0 to 100. A value of Z was then computed from X and Y and shown to the subject. The functions to be optimized were inverted elliptical paraboloids of the form $(X/a)^2 + (Y/b)^2 = Z/c$, where a and b are constants that determine the lengths of the major and minor axes and c is a constant that is set by the constraint that Z takes on values between 0 and 10,000.

Half of the subjects completed 8 problems and half completed 16 problems. The optimal point, the ratio of the axes, and the rotation of the major axis were varied across problems. The optimal point was constrained to occur within the window $20 < X,Y < 80$ and assumed a uniform random value across problems. The ratio of the minor axis to the major axis was set at 1:1, 1:2, 1:3, or 1:4. These were assigned in a randomized-blocks order. The angle of rotation of the major axis from the horizontal X-axis was varied in a uniformly random manner across problems. Using contour curves, Figure 2 shows the graphic representation of the problem for an inverted elliptical paraboloid with a ratio of 1:2.

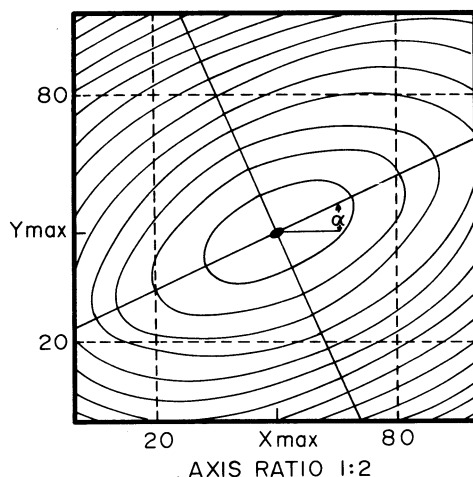Problems were presented on interactive DecWriter computer terminals by an executive program. The subjects were instructed either individually or in small groups and were then allowed to work on the problems at their leisure. At the beginning of each session on the terminal the instructions were printed along with a diagram showing the X,Y coordinates from 0 to 100.

RESULTS

Figure 3 shows the effect of practice on the number of moves to find the maximum averaged over the subjects. A decrease in the number of moves was found as a function of the number of problems worked. This decrease was significant $[F(3,27) = 3.49, p < .05]$ over the first four blocks of problems and suggests that the subjects were developing more efficient strategies with practice. The ratio of the minor axis to the major axis had a significant effect on the number of moves to maximization $[F(3,27) = 3.40, p < .05]$. The means were 19.4, 27.0, 25.0, and 31.4 for the ratios of 1:1, 1:2, 1:3, and 1:4, respectively. The more elliptical the paraboloid, the more difficult optimization tended to be. Furthermore, as the axes of the paraboloid were rotated away from the X,Y axes, there was an increase in the number of moves to optimization for axis ratios of 1:2, 1:3, and 1:4. The correlations between the number of moves and the angle of rotation in radians were .39, .44, and .54, respectively, for the three axis ratios $(p < .05$ in each case). Finally, the distance from the initial guess to the optimal point had an effect: the greater the distance, the greater the number of moves to optimization $[t(9) = 2.20, p < .05$ one-tail test].

The path defined by the set of moves taken by a subject on any problem constitutes a subject-item protocol. Such protocols can be analyzed to determine the overall strategy of search taken by the subject. Figure 4 shows a typical path taken by one subject. Although a number of strategies might be devised to find the optimum point, the following four general cases of strategies, in which X and Y are interchangeable, can be identified:

(1) Alternating univariate optimization—Y is held constant and X is adjusted in a series of moves to find a partial maximum of Z with respect to X; then X is held constant and Y is adjusted in a series of moves to find a partial maximum of Z with respect to Y. The procedure is repeated until the overall maximum is found.

(2) Alternating step optimization—Z is incremented by one successful X move; then Z is incremented by one successful Y move. The procedure is repeated until the maximum is reached.

(3) Simultaneous multivariate optimization—Both X and Y are changed on each move in a direction so as to increment Z. When a failure to increment Z occurs, the direction is changed.

(4) Pattern search—Exploratory moves are made in which X and Y are adjusted individually in order to determine their partial gradients.

Figure 2. Contour graph of one of the inverted elliptical paraboloids used in the experiment (α is the angle of rotation).
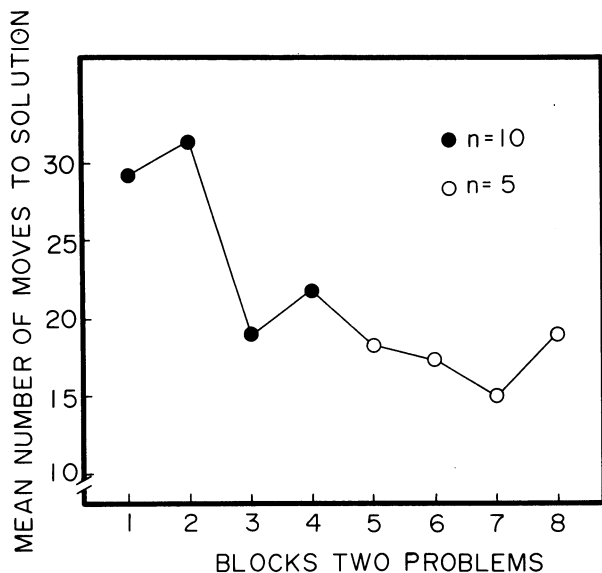
Figure 3. Mean number of moves to find the maximum point as a function of problems.
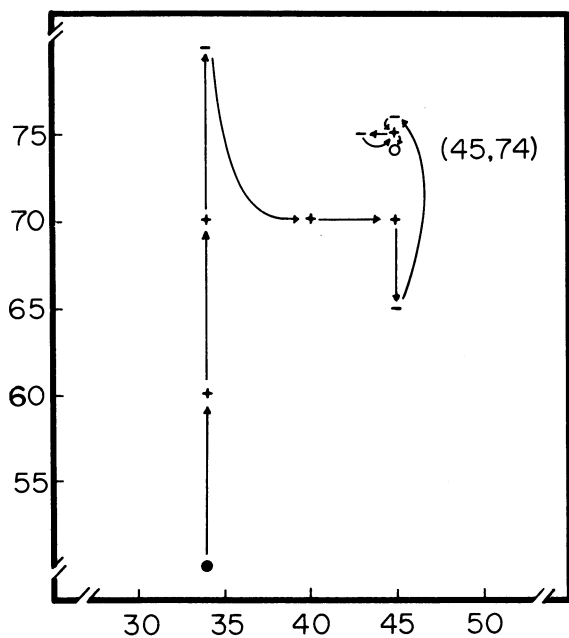

Figure 4. The path of a typical subject in the experiment (arrows indicate direction of move; the solid circle is the first point tested; pluses and minuses indicate either an increase or a decrease in Z from the last point; and the open circle is the maximum point).

Then a pattern move is made in which X and Y are changed simultaneously in the direction of a vector determined by the exploratory moves. The pattern move is repeated until a failure occurs, and the process is repeated until the maximum is found.

The majority of subjects tended to use the alternating univariate optimization strategy, as typified in the path shown in Figure 4. These subjects changed one variable at a time. If it resulted in an increase in Z, they continued to change it in that direction. When a failure to increment Z occurred, they returned to the last successful value of that variable and started to change the other variable. This strategy is particularly effective when the major and minor axes are parallel to the X,Y axes. However, when the axes are not aligned, this strategy necessitates a number of alterations, which results in a zigzag path and leads to an increase in the number of moves to maximization, as reported above. A few subjects used the simultaneous multivariate strategy during an initial phase of locating the maximum and then switched to the alternating univariate strategy as they approached the maximum. These subjects generally incremented X and Y by the same absolute amounts and therefore moved only along the diagonal. Consequently, this strategy turned out to be formally equivalent to the alternating univariate strategy rotated 45 deg.

A number of changes occurred in the subjects' moves as they approached the maximum point. It has been mentioned already that some subjects switched from the simultaneous strategy to the univariate strategy. The greatest change, however, was in terms of step size. Step size is defined here as the Euclidean distance from one X,Y point to the next. Figure 5 shows a Vincentized curve of step size as a function of 10 blocks of trials. Initially, the subjects made rather large steps, which decreased significantly in size as the subjects approached the optimum [$F(9,81) = 19.81$, $p < .01$]. Most subjects using direct-search strategies, on the other hand, start with small exploratory steps, then increase step size as they move in the direction of the optimum point, and finally
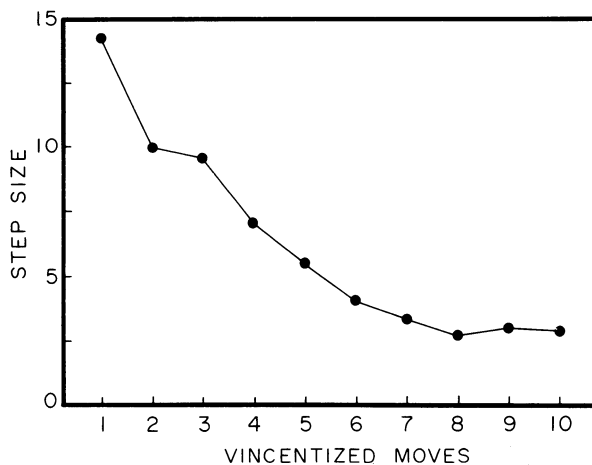

Figure 5. Mean step size as a function of Vincentized trials.

reduce step size as they converge on the maximum and test to see if they have indeed found the maximum. In the present task, however, the subjects knew that the maximum was at 10,000. Consequently, large exploratory moves would not be unreasonable if Z was appreciably less than the maximum value. As subjects homed in on the maximum, step size was reduced to a minimum so as not to overstep it.

DISCUSSION

It is well known that problem-solving strategies differ in terms of cognitive requirements and efficiency. Subjects seek to meet task requirements while at the same time minimizing cognitive effort. It is a tradeoff between maximizing performance and minimizing effort. The hill-climbing problem, as presented to the subjects in this experiment, incorporated two variables X and Y. Because these may be thought of as separate dimensions, the subjects tended to minimize complexity of the task by dealing with only a single dimension at a time. This strategy was efficient only when the mixed partial derivatives of the surface with respect to X and Y were zero. This was the case for the ratio 1:1 and whenever the rotation of the axes was zero. When the surface was elliptical and rotated, the simultaneous multivariate optimization and pattern-search strategies were superior, because they used information about the direction of the steepest gradient in order to point to the greatest increment in Z. It is not yet clear whether subjects can learn to use this information for such surfaces. Additionally, it needs to be seen how instructions that vary the goal of minimizing the number of moves or time or distance from the initial point to the maximum affect the choice of strategy.

Finally, it should be mentioned that nonlinear hill-climbing problems have several advantages over other tasks used in the study of problem-solving strategies. The hill-climbing task ensures the externalization of the problem-solving protocol by employing an unknown evaluation function. The subject-item protocol path is left as a trail, because the subject must sample each point. This allows for a fairly complete analysis of the search strategy, as well as a measure of the parameters of the search process. In the present case, it was clear that search was generally accomplished by finding a partial maximum on one variable and then a partial maximum on the other, until the overall maximum was obtained.

Direct-search strategies are important in many typical everyday problems. There are numerous instances in which people must find the right combination of variables in order to maximize some function. In medicine, the physician is interested in the right combination of dosage levels; in athletics, one may be interested in the right combination of exercise repetitions; and in seeking an overall quality of life, one is interested in environmental, economic, and societal variables. Such problems can be thought of as hill-climbing or optimization problems.

REFERENCES

CHANDLER, J. P. Subroutine STEPIT—Finds local minima of a smooth function of several parameters. *Behavioral Science*, 1969, **14**, 81-82.

DIXON, L. C. W. *Nonlinear optimization*. London: English Universities Press, 1972.

HOOK, R., & JEEVES, T. Direct search solution of numerical and statistical problems. *Journal of the Association for Computing Machinery*, 1961, **8**, 212-228.

NEWELL, A., & SIMON, H. A. *Human problem solving*. Englewood Cliffs, N.J: Prentice-Hall, 1972.

WICKELGREN, W. A. *How to solve problems: Elements of a theory of problems and problem solving*. San Francisco: Freeman, 1974.