[3] P. H. Randolph, "Optimal stopping rules for multinomial observations," *Metrika*, vol. 14, pp. 48–61, 1968.
[4] T. Robbins, "A penalty-function approach to radiation—pattern synthesis," Ph.D. dissertation, Southern Methodist Univ., Dallas, Tex., 1972.
[5] A. Wald, *Sequential Analysis*. New York: Wiley, 1947.
[6] W. Zangwill, *Nonlinear Programming: A Unified Approach*. Englewood Cliffs, N.J.: Prentice-Hall, 1969.

A Comparative Evaluation of Two Global Search Algorithms

GEORGE A. BEKEY AND MAN T. UNG

Abstract—**Two heuristic methods for locating the global optimum of a multimodal performance index surface are described. One method is based on a modified random creep procedure which first locates a local minimum and then searches the parameter space with vector steps whose mean length gradually increases. The second is a modification of the Kiefer–Wolfowitz stochastic approximation procedure, in which a random perturbation is added to each measurement. Both algorithms are compared by applying them to finding the roots of a nonlinear algebraic equation and to a constrained dynamic optimization problem.**

Introduction

Computational methods of optimization have received a great deal of attention during the past decade. However, while most of the available methods perform well when locating the extremum of a unimodal surface, they are generally local in character when confronted by multiple extrema. In attempting to locate the global optimum in a nonlinear problem, the user must resort to one of two alternatives:

1) to perform a large number of local optimizations beginning from randomly selected initial conditions;
2) to divide the parameter space in advance into a number of subspaces and perform a search inside of each subspace.

Some combination techniques have been proposed, such as those due to McMurtry and Fu [1] or Hill [2] in which a grid procedure is used to divide the space into a number of hypervolumes, the particular hypervolume in which the global extremum is located is identified and finally a unimodal search is performed. In general, all such techniques are plagued by extremely long solution times as a result of the "curse of dimensionality." For example, in the case of six parameters, partitioning each parameter into as few as five equal partitions yields 5^6 (15 625) different initial conditions from which a local optimization may need to be performed.

An interesting heuristic global search procedure which is claimed to be free of dimensionality problems was proposed by Opacic [3]. The algorithm, called "SCAN," proceeds by setting up new starting points in the parameter space by reflection procedures around old starting points; from each of these a local minimum is found using the Fletcher–Powell–Davidon method.

Random search and random creep techniques have been applied to the global optimization problem for some time. In general, with such techniques, once a local minimum is located, the search range is expanded about this minimum in an attempt to locate a region where the performance index shows an improvement [4]–[6]. Such procedures offer the advantage of locating a global optimum from a single arbitrary initial condition within the constrained parameter space. The computer programs associated with the method are generally simple and constraints present no difficulty. However, convergence to the global minimum is obtained only in a probabilistic sense and the necessary convergence proofs are lacking at the present time.

As an alternative to the random search procedures, Sugiyama *et al.* [7], [8] have suggested a modification of the Kiefer–Wolfowitz stochastic approximation procedure which would enable it to locate all the local optima within a searching region. While heuristic in nature, the proposed algorithm was tested and found to be promising.

In this paper, the two algorithms previously cited are compared by evaluating their performance in two optimization problems. One of these problems is concerned with finding all the roots of a nonlinear algebraic equation, while the second is a dynamic optimization problem involving differential equations. The two algorithms are compared with respect to ease of implementation and performance.

Statement of the Problem

The parameter optimization problem may be stated as follows. Determine the values of the ordered sets of m parameters

$$\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_m)^T \tag{1}$$

which optimize (minimize or maximize) the criterion function $J(\alpha)$ subject to a set of equality and/or inequality constraints. The set of all α satisfying the constraints defines the feasible regions in which the solution must be located. (For some problems the constraints are not present or may be effectively eliminated by the use of penalty function techniques.)

The solution of the parameter optimization problem is denoted by

$$J(\alpha^*) = \min_{\alpha} J(\alpha) \tag{2}$$

where α^* is the optimum point. For convenience, (2) is stated as a minimization problem.

The optimization strategy is based on the selection of a starting value α^0 and computation of parameter increments $\Delta\alpha$ such that the ith parameter value is determined by

$$\alpha^i = \alpha^{i-1} + \Delta\alpha^{i-1}. \tag{3}$$

Description of the Algorithms

A. Random Creep Algorithm

The basic random creep algorithm may be stated as follows. Starting from an initial value α^0, the criterion function is measured at $\alpha^0 + \Delta\alpha$ where $\Delta\alpha$ is a vector with fixed or variable length and random direction. If

$$J(\alpha^0 + \Delta\alpha) < J(\alpha^0) \tag{4}$$

the step is termed a "success" and the base point is moved to $\alpha^1 = \alpha^0 + \Delta\alpha$; otherwise, the base point remains α^0 and another random step is attempted.

The basic strategy described here is quite old, having been described originally for analog computer application by Favreau and Franks [9] in 1958. A variable step-size random search was published by Mitchell [10] and its convergence properties were studied by Rastrigin [11]. There are many variations of the basic strategy described here, which are concerned with such matters as control of step-size, and adjustment of direction based on past successes or failures. The particular technique described in [5] has been termed absolute positive and negative biasing. In this case, if the last step produced a success it is used again for the next trial step, i.e.,

$$\Delta\alpha^i = \Delta\alpha^{i-1} \tag{5}$$

where the superscripts refer to the iteration or step number. If the last step resulted in a failure, one uses

$$\Delta\alpha^i = -\Delta\alpha^{i-1}. \tag{6}$$

The algorithm used in this study is based on [5] but does not include absolute biasing. It may be described as follows.

A Gaussian number generator subroutine supplies the next trial increment $\Delta\alpha = N(\sigma)$ in the parameter space. The vector $\Delta\alpha$ will have zero mean and standard deviation $\{\sigma_1', \sigma_2', \cdots, \sigma_m'\}$. The trial value

$$\alpha' = \alpha + \Delta\alpha \tag{7}$$

Manuscript received March 30, 1973; revised June 24, 1973. This work was supported in part by the U.S. Air Force Office of Scientific Research under Grant AF/AFOSR 71-2008.

The authors are with the Department of Electrical Engineering–Systems, University of Southern California, University Park, Los Angeles, Calif. 90007.

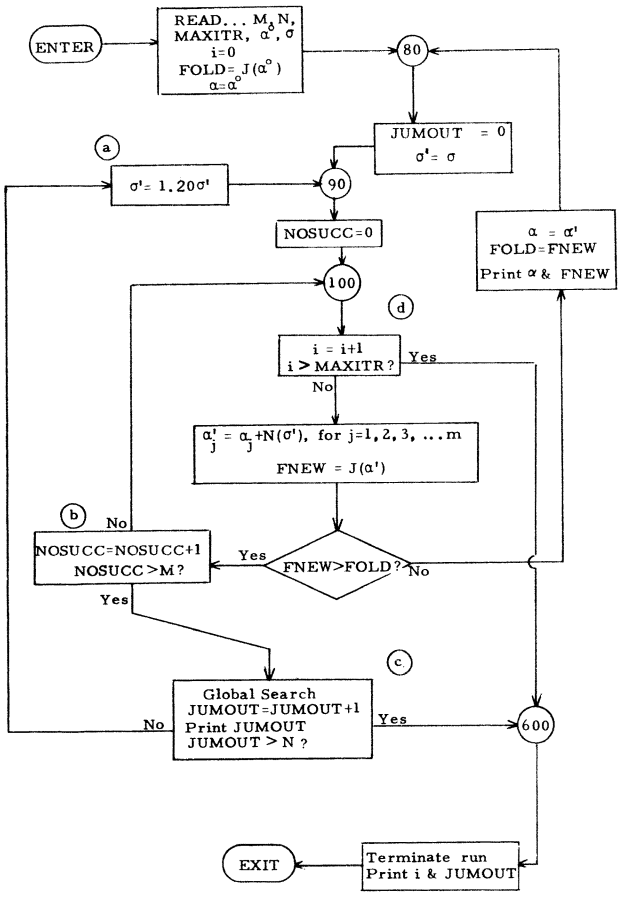

Fig. 1. Random creep algorithm.

will be abandoned if it does not result in an improvement over the present performance index $J(\alpha)$. The standard deviations chosen are sufficiently small so that whatever steps taken remain in the neighborhood of the current vector α, thus taking advantage of the local continuity properties of $J(\alpha)$. Starting from an initial parameter vector α^0, this algorithm will find a zigzag path downward to a relative minimum. Following a convergence test for a local minimum (based on a number of successive failures), a global search subroutine is entered in which the standard deviation of the Gaussian random number generator is increased by 10 percent. Continued failures will result in a gradual increase in the standard deviation of the trial steps, which in this algorithm are random in both length and direction. While the algorithm does not guarantee the discovery of all the local minima, experience has shown it to be quite successful. Fig. 1 depicts the flowchart for this method (program listings can be obtained by writing to the authors). The significance of the symbols used in the flowchart is as follows.

α — $\{\alpha_1, \alpha_2, \cdots, \alpha_m\}^T$ corresponds to the optimum $J(\alpha)$ obtained so far, unless α' can produce $J(\alpha') \leq J(\alpha)$, α' will be rejected and another trial value is attempted.

σ — Original standard deviation vector.

σ' — Current value of σ, which might be many times the size of σ, every time the immediate neighborhood of α fails to yield an improvement over $J(\alpha)$ after an intensive search, σ' is increased by 20 percent (box a) to allow an ever larger region to be searched.

FOLD, FNEW — The current optimum value of the criterion function and the new trial value, respectively.

NOSUCC — Number of consecutive trials rejected for lack of improvement over FOLD. After M successive tries (box b), the search is widened in quest of the global optimum.

JUMOUT — Number of times that σ was expanded to reach the current value σ'. If no improvement in $J(\sigma)$ is found at the end of N expansions in σ (box c), the program is terminated.

MAXITR — Whenever the total number of iterations i reaches MAXITR, the program is terminated regardless of the outcome in $J(\alpha)$. This integer MAXITR places an absolute upper limit on i.

$N(\sigma_j')$ — A number obtained from the Gaussian noise generating subroutine having a standard deviation σ_j' and zero mean.

B. Modified Kiefer–Wolfowitz Algorithm

Stochastic approximation is an iterative procedure for finding the root of an unknown monotone regression function originally proposed by Robbins and Monro, which was extended by Kiefer and Wolfowitz to the problem of locating the maximum or minimum of an unknown unimodal regression function [12]. The technique can be shown to converge with probability one under fairly general conditions. The distribution of the disturbances in the noisy measurement data need not be known, provided that they are finite and have zero mean. While originally proposed for a one dimensional situation, the method was subsequently extended to the multidimensional unimodal case.

The method is visualized most easily in the one dimensional case. Assume that we desire to find the minimum of an unknown function $J(\alpha)$ on the basis of observations $Y(\alpha)$, where

$$Y(\alpha) = J(\alpha) + Z(\alpha) \tag{8}$$

where $Z(\alpha)$ is zero mean finite variance observation noise. Kiefer and Wolfowitz have demonstrated that a convergent sequence of parameter values α may be obtained from the algorithm

$$\alpha^{n+1} = \alpha^n - \frac{a^n}{c_n}[Y(\alpha^n + c_n) - Y(\alpha^n - c_n)] \tag{9}$$

where a_n and c_n are sequences of numbers obeying particular restrictions. Sequences which satisfy the necessary requirements are

$$a_n = \frac{G}{n}, \qquad G = \text{constant}$$
$$c_n = \frac{1}{n^{1/3}}. \tag{10}$$

The procedure is basically designed for locating the maximum or minimum of unimodal functions in the presence of observation noise. It was recently suggested by Sugiyama *et al.* [7], [8] that the basic Kiefer–Wolfowitz procedure could be forced to follow a variety of trajectories in the parameter space which nevertheless lead to the proper optimum by the simple device of adding to each observation a zero mean artificial noise with a nonzero variance which decreases with each iteration. The procedure is heuristic and no formal proof of its ability to locate other minima within the searching region is available at this time. The modification raises the intriguing possibility that a judicious choice of the artificial perturbation may cause the process to converge to a certain number of optimum points with probability one.

If we denote the artificial perturbation $N(\sigma_n)$, then the algorithm of (9) is modified to become

$$\alpha^{n+1} = \alpha^n - \left(\frac{a_n}{c_n}\right)[Y^*(\alpha^n + c_n) - Y^*(\alpha^n - c_n)] \tag{11}$$

where

$$Y^*(\alpha^n) = Y(\alpha^n) + N(\sigma_n). \tag{12}$$

As indicated in the preceding

$$E[N(\sigma_n)] = 0 \tag{13}$$

and

$$\lim_{n \to \infty} E[N^2(\sigma_n)] = 0. \tag{14}$$

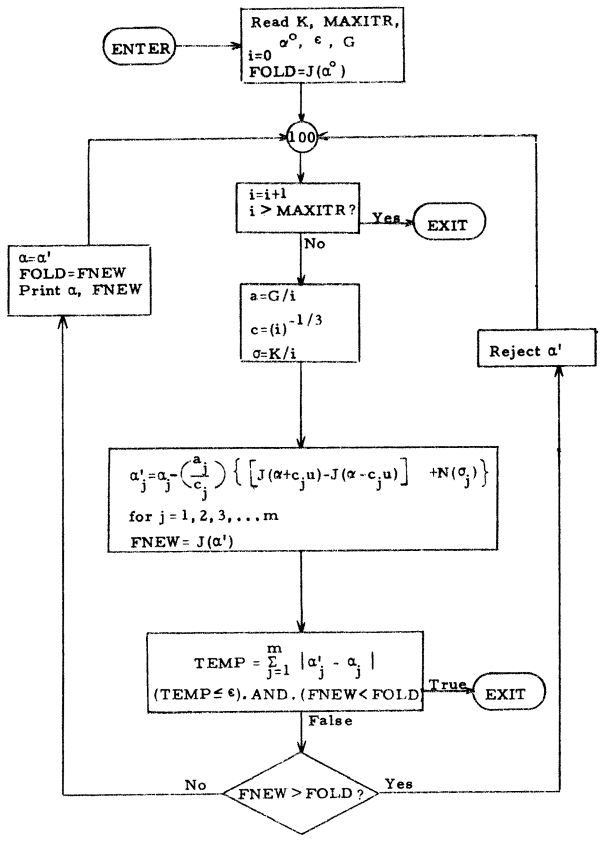

Fig. 2. Modified Kiefer–Wolfowitz method.

Specifically, the perturbation used in this study was Gaussian and characterized by

$$N(\sigma_n) \in N\left(0, \frac{K}{n^2}\right). \tag{15}$$

For the case of m parameters, the basic algorithm is modified to ensure that only one of the elements of the parameter vector is perturbed at one time. If we let U be an m-dimensional vector with all zero elements except $U_n = 1$, then the $(n+1)$st search point for the jth component of the parameter vector is given by

$$\alpha_j^{n+1} = \alpha_j^n - \frac{a_n}{c_n}[Y^*(\alpha_j^n + c_nU) - Y^*(\alpha_j^n - c_nU)]. \tag{16}$$

The constants K in (15) and G in (10) depend on the size of the region to be searched. In the current study they were chosen by trial and error, the aim being to ensure that the artificial perturbation signal is sufficiently large to force the searching trajectory to traverse a large portion of the parameter space from a single initial condition.

A flow chart of the modified Kiefer–Wolfowitz method is shown in Fig. 2. Except for the notation indicated previously, this figure resembles Fig. 1 in many ways. Two conditions will cause the program to exit, namely, $i >$ MAXITR and an improvement in $Y(\alpha)$ is made, but the norm of the step falls below a preselected quantity ε.

As with many other discrete parameter adjustment procedures, it is possible for this algorithm to "hang-up" on a narrow ridge. If such ridges or valleys are anticipated, a ridge-following subroutine can be included in the flow chart of Fig. 2.

Applications of the Method

The methods described in this paper were successfully tested on two systems. One is algebraic and the other is made up of differential equations. Consider first the nonlinear algebraic equation described by Hosaki [8]. We desire to find the roots of the expression

$$J(\alpha) = (+1 - 8\alpha_1 + \tfrac{7}{3}\alpha_1^2 - \tfrac{7}{3}\alpha_1^3 + \tfrac{1}{4}\alpha_1^4)\alpha_2^2 e^{-\alpha_2} \tag{17}$$

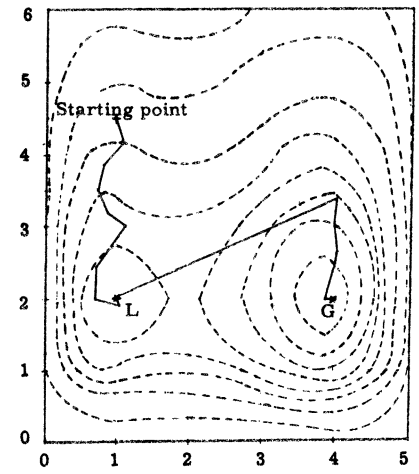

Fig. 3. Trajectory of random creep algorithm.

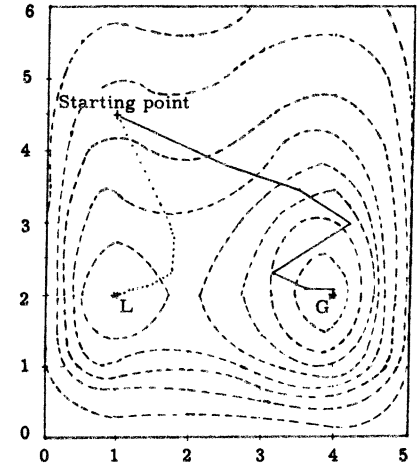

Fig. 4. Trajectories of modified Kiefer–Wolfowitz algorithm.

where

$$0 \leq \alpha_1 \leq 5$$

$$0 \leq \alpha_2 \leq 6.$$

The $J(\alpha)$ surface possesses two minima in the region of interest at

$$J(1,2) = -1.127, \quad \text{(relative optimum)} \tag{18}$$

$$J(4,2) = -2.345, \quad \text{(global optimum)}.$$

Starting at (1,4.5) the random creep method found the relative minimum after 108 steps. The search would have stalled there had it not been for the widening of σ. This led to a region where $J(\alpha) < -1.127$; one point of that region was (4.08,3.53). From there on, the search is again confined to a small neighborhood which eventually yielded the global optimum at (4,2). Fig. 3 shows the trajectory of this run. It is not possible to show all the steps taken, large and small, so only the compounded trajectory of many tiny steps and single large steps is shown. The computer run was made with the following input parameters: $M = 20$; $N = 30$; MAXITR = 500; $\alpha_0 = \{1,4.5\}$; and $\sigma = \{0.2,0.2\}$. The accompanying computer printout can be found in Appendix A. Notice that, after iteration #108, the local minimum (point L) was found. It took 11 consecutive widenings of the search before an improvement was made at iteration #249. At that time, $\sigma = \{1.486,1.486\}$. Subsequently, the global optimum (point G) was within reach. A total of 27 runs were made all starting at $\{1.0,4.5\}$. Each run was cut off after 500 iterations. All but one of the runs were successful in locating the global minimum as reflected by the means of

TABLE I
SUMMARY OF RESULTS
ALGEBRAIC SYSTEM

Algorithm	α^0	Average Running Time of Successful Runs
Random creep	(1,4.5)	4 s
Modified K–W	(1,4.5)	less than 0.5 s
Differential Equation System		
Random creep	(0.2,0.9)	6 min
Modified K–W	(0.2,0.9)	3.5 min

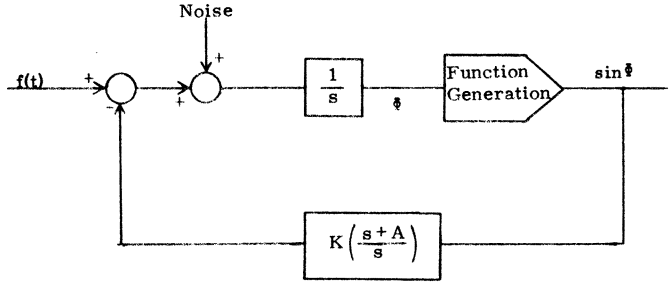

Fig. 5. System under study.

the final answers: $\bar{\alpha}_1 = 3.9968$ and $\bar{\alpha}_2 = 2.0819$ as opposed to the exact answers of {4,2}.

When the modified Kiefer–Wolfowitz method was applied to the same problem, the situation depicted by Fig. 4 arises. From the starting point {1,4,5}, the solution converged to either the relative minimum or the global minimum depending on the random numbers (σ_n) as shown in (12). Two typical trajectories are drawn in Fig. 4. Running times for all cases are summarized in tabular form in Table I.

Let us next apply the two methods to a nonlinear dynamic system, taken from communication theory. The example system is a simplified model of a baseband loop which is supposed to track the input frequency $f(t)$ and in which synchronization effects take place, as shown in block diagram form in Fig. 5. For those interested in the working details of this example, consult [13]. Translating the block diagram of Fig. 5 into a system of ordinary differential equations, assuming that the function generator is a sine function

$$\frac{d\Phi}{dt} = f(t) - K(\psi + \sin\Phi) + n(\sigma^*), \qquad 0 \le K \le 2.8$$

$$\frac{dJ}{dt} = \left(\frac{d\Phi}{dt}\right)^2$$

$$\frac{d\psi}{dt} = A \sin(\Phi), \qquad 0 \le A \le 1.04 \qquad (19)$$

where ψ is an intermediate state variable introduced in the process of transforming the feedback block of Fig. 5 into a differential equation. Here, as well as in other parts of this paper, J stands for the criterion function proportional to the error between incoming frequency and the tracked frequency. The term $n(\sigma^*)$ represents the Gaussian noise with standard deviation σ^* and zero mean.

In order to obtain some insight into the behavior of the method, the criterion function surface for this problem was plotted by making a large number of computer runs using the IBM Continuous System Modeling Program (CSMP). The resulting contour lines are shown in Fig. 6. In this program the parameters $\{\alpha_1,\alpha_2\}$ were called $\{A,K\}$, where A is the filter time constant and K stands for the feedback loop gain. Again there exist two minima, both situated on the boundary

$$J(1.04,1.2) = 23.3, \quad \text{(relative minimum)}$$
$$J(1.04,2) = 22.3, \quad \text{(global minimum)}. \qquad (20)$$

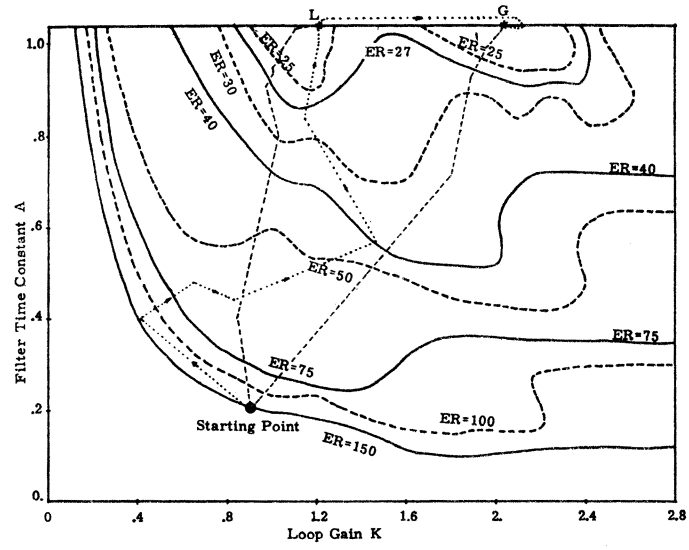

Fig. 6. Criterion function for phase-locked loop; note: $ER(A,K) = J(\alpha)$ by definition.

The locations of the optima serve to further test the convergence properties of the two methods. Compared to other systems of differential equations, this is admittedly a small problem. In terms of computer time it still requires a vast amount of time because of the multiple evaluations of the performance index $J(\alpha)$ during each iteration, each requiring the solution of differential equations. A Runge–Kutta fourth-order algorithm was used in this program. Also in this program the observed noise $n(\sigma^*)$ was kept relatively low (compared to the signals), having a standard deviation of $\sigma^* = 0.01$.

From the starting point $\alpha^0 = \{0.2,0.9\}$ the random creep algorithm sometimes found the global optimum without stopping at the local minimum. By contrast, the dotted path on Fig. 6 found the local minimum first, the global minimum afterwards. This run consumed 6 min on the IBM 370/155. When the modified Kiefer–Wolfowitz method was applied to the same problem it produced two families of trajectories, one set converging to {1.4,1.2} while the other coming together at {1.04,2.0}. Fig. 6 shows one typical trajectory of each (dashed curves). Because the Kiefer–Wolfowitz method relies on gradient calculations, it becomes sluggish in the neighborhood of the optima. Hence the search was stopped upon satisfying the relationship $\|\alpha' - \alpha\| \le \varepsilon$, for $\varepsilon = 0.01$. The two dashed trajectories on Fig. 6 used up to 10 min of computer time.

Table I puts the execution times of both methods in perspective to facilitate direct comparison. The average running times given are those of successful runs only. In these runs the global minima were reached (with or without stopping at the relative minima). Unsuccessful runs were not accounted for in this table.

The preceding problem specifically included only parameter constraints. If functional constraints appear in the problem, it will be necessary to verify the feasibility of the computed random perturbation and creep before proceeding. Alternatively, it may be possible to transform the problem to one of unconstrained optimization by means of penalty functions.

CONCLUSION

Two heuristic procedures for finding all the maxima (or minima) of a surface within a constrained searching region have been demonstrated. Both techniques were applied to two sample problems, one algebraic and one dynamic. The results illustrated the ability of both methods to locate the global optimum from a single arbitrary initial condition.

No general convergence proofs for either method are available at this time in the multimodal case. In addition to the theoretical work needed in this connection, there are probably acceleration procedures which could be tried and which could be very important in response surfaces of many dimensions.

APPENDIX A

Iteration	α_1	α_2	J(α)	σ_1 or σ_2
0	1.00000	4.50000	-0.46866	0.20000
3	1.07864	4.14765	-0.56397	0.20000
5	0.88992	3.85862	-0.64812	0.20000
6	0.73725	3.63153	-0.68240	0.20000
7	0.78464	3.58483	-0.71298	0.20000
17	0.79484	3.27404	-0.81487	0.20000
19	0.89123	3.22341	-0.85388	0.20000
23	1.12829	3.05294	-0.90724	0.20000
25	0.93272	2.91856	-0.95516	0.20000
27	0.83677	2.75890	-0.98259	0.20000
28	0.76319	2.58384	-0.99818	0.20000
30	0.70626	2.42434	-0.99820	0.20000
35	0.68263	2.06214	-1.02059	0.20000
42	0.75268	2.10778	-1.06371	0.20000
44	0.80029	2.03451	-1.08912	0.20000
46	0.83687	1.84182	-1.09569	0.20000
47	1.10311	1.84612	-1.11297	0.20000
56	0.97345	1.83422	-1.11904	0.20000
71	0.93538	1.96726	-1.11954	0.20000
85	0.92422	1.90217	-1.12004	0.20000
87	0.97299	2.04229	-1.12669	0.20000
103	1.02506	1.98295	-1.12721	0.20000
108	1.00528	1.98600	-1.12771	0.20000
JUMOUT = 1				
JUMOUT = 2				
JUMOUT = 3				
JUMOUT = 4				
JUMOUT = 5				
JUMOUT = 6				
JUMOUT = 7				
JUMOUT = 8				
JUMOUT = 9				
JUMOUT = 10				
JUMOUT = 11				
349	4.08324	3.53003	-1.57444	1.48601
350	3.99468	3.28736	-1.74914	0.20000
352	3.99070	3.07293	-1.89385	0.20000
354	3.93506	2.99573	-1.93899	0.20000
355	4.16349	2.93650	-1.94227	0.20000
358	4.04506	2.84550	-2.03576	0.20000
361	4.06679	2.54815	-2.19399	0.20000
364	4.13033	2.15666	-2.30275	0.20000
367	3.88618	2.03345	-2.32544	0.20000
370	3.90530	2.06060	-2.32991	0.20000
373	4.05143	1.93819	-2.33911	0.20000
385	4.00596	1.95753	-2.34470	0.20000
396	4.00477	1.97394	-2.34540	0.20000
JUMOUT = 1				
JUMOUT = 2				
451	3.99093	2.01405	-2.34555	0.28800
JUMOUT = 1				
JUMOUT = 2				
MAX. NO. ITERATION EXCEEDED				
ITER=501 JUMOUT= 2				

Acknowledgment

The authors wish to express their appreciation for the suggestions and advice that Dr. H. Sugiyama has given during the preparation of this paper.

References

[1] G. J. McMurtry and K. S. Fu, "A variable structure automaton used as a multimodal searching technique," *IEEE Trans. Automatic Control*, vol. AC-11, pp. 379–387, May 1966.

[2] J. D. Hill, "A search technique for multimodal surfaces," *IEEE Trans. Syst. Sci. Cybern.*, vol. SSC-5, pp. 2–8, Jan. 1969.

[3] J. Opacic, "A heuristic approach to multimodal function minimization —the global aspect," in *Proc. 9th Allerton Conf. on Circuit and System Theory*, 1971, pp. 369–376.

[4] I. N. Bocharov and A. A. Feldbaum, "An automatic optimizer for the search for the smallest of several minima," *Automat. Telemekh.*, vol. 23, pp. 289–301, 1962.

[5] G. A. Bekey, M. H. Gran, A. E. Sabroff, and A. Wong, "Parameter optimization by random search using hybrid computer techniques," in *Proc. Fall Joint Computer Conf.*, 1966, pp. 191–200.

[6] E. C. Stewart, W. P. Kavanaugh, and D. H. Brocker, "A study of a global search algorithm for Optimal control," in *Proc. 5th Int. Analog Computer Conf.*, 1958, pp. 437–443.

[7] S. Kusano, M. Tasaka, and H. Sugiyama, "Some additional aspects of the Robbins–Monro stochastic approximation process for multiple roots searching," *Technol. Rep. of Osaka Univ.*, vol. 18, p. 1013, 1968.

[8] K. Hosaki, H. Imamura, M. Tasaka, and H. Sugiyama, "A heuristic method for maxima searching in case of multimodal surfaces," *Technol. Rep. of Osaka Univ.*, vol. 20, pp. 337–344, 1970.

[9] R. R. Favreau and R. G. Franks, "Statistical optimization," in *Proc. 2nd Int. Analog Computer Conf.*, 1958, pp. 437–443.

[10] B. A. Mitchell, "A hybrid analog-digital parameter optimizer for ASTRAC II," in *Proc. AFIPS Spring Joint Computer Conf.*, vol. 25, 1964, pp. 271–285.

[11] L. A. Rastrigin, "The convergence of random search method in the presence of noise," *Automat. Remote Contr.*, vol. 24, pp. 1337–1342, 1963.

[12] J. Wolfowitz, "On the stochastic approximation method of Robbins and Monro," *Ann. Math. Statist.*, vol. 23, pp. 457–461, 1952.

[13] W. C. Lindsey, *Synchronization Systems in Communication and Control*. Englewood Cliffs, N.J.: Prentice Hall, 1972.

The Best Two Independent Measurements Are Not the Two Best

THOMAS M. COVER

***Abstract*—Consider an item that belongs to one of two classes, $\theta = 0$ or $\theta = 1$, with equal probability. Suppose also that there are two measurement experiments E_1 and E_2 that can be performed, and suppose that the outcomes are independent (given θ). Let E_i' denote an independent performance of experiment E_i. Let $P_e(E)$ denote the probability of error resulting from the performance of experiment E. Elashoff [1] gives an example of three experiments E_1,E_2,E_3 such that $P_e(E_1) < P_e(E_2) < P_e(E_3)$, but $P_e(E_1,E_3) < P_e(E_1,E_2)$. Toussaint [2] exhibits binary valued experiments satisfying $P_e(E_1) < P_e(E_2) < P_e(E_3)$, such that $P_e(E_2,E_3) < P_e(E_1,E_3) < P_e(E_1,E_2)$. We shall give an example of binary valued experiments E_1 and E_2 such that $P_e(E_1) < P_e(E_2)$, but $P_e(E_2,E_2') < P_e(E_1,E_2) < P_e(E_1,E_1')$. Thus if one observation is allowed, E_1 is the best experiment. If two observations are allowed, then two independent copies of the "worst" experiment E_2 are preferred. This is true despite the conditional independence of the observations.**

I. Introduction

In the classification or hypothesis testing problem, it is well known that the most informative k element subset of n measurements is not necessarily the union of the k individually most informative measurements. An easy example can be generated by allowing statistical dependence among the measurements, thereby making it redundant to use the two measurements with individually lowest probabilities of error.

If one goes no further than this example, one might be led to believe that the difficulties in finding the best k measurements arise solely from dependence among the measurements. Elashoff [1] and Toussaint [2] have shown that the essence of the problem is retained even when all the measurements are (conditionally) independent. This correspondence provides another such example.

II. Family of Examples

Consider the following example:

$$\begin{array}{lll} & E_1 & E_2 \\ \theta = 0 & X = \begin{cases} 1, & p_0 \\ 0, & 1 - p_0 \end{cases} & Y = \begin{cases} 1, & r_0 \\ 0, & 1 - r_0 \end{cases} \\ \theta = 1 & X = \begin{cases} 1, & p_1 \\ 0, & 1 - p_1 \end{cases} & Y = \begin{cases} 1, & r_1 \\ 0, & 1 - r_1 \end{cases} \end{array}$$

where $\Pr\{\theta = 0\} = \Pr\{\theta = 1\} = \frac{1}{2}$. Also, let (E_i,E_i') denote two independent repetitions of E_i. Thus, for example, for (E_1,E_1'), $\Pr\{(X,X') = (1,0) \mid \theta\} = p_\theta(1 - p_\theta)$.

The Bayes probability of error is given for a discrete random variable X by

$$P_e(E) = \sum_x \min\{\Pr\{\theta = 0\}P_0(x), \Pr\{\theta = 1\}P_1(x)\}.$$

Manuscript received June 5, 1973; revised July 30, 1973. This work was supported by the Air Force Office of Scientific Research under Contract F44620-69-C-0101.

The author is with the Departments of Electrical Engineering and Statistics, Stanford University, Stanford, Calif. 94305.