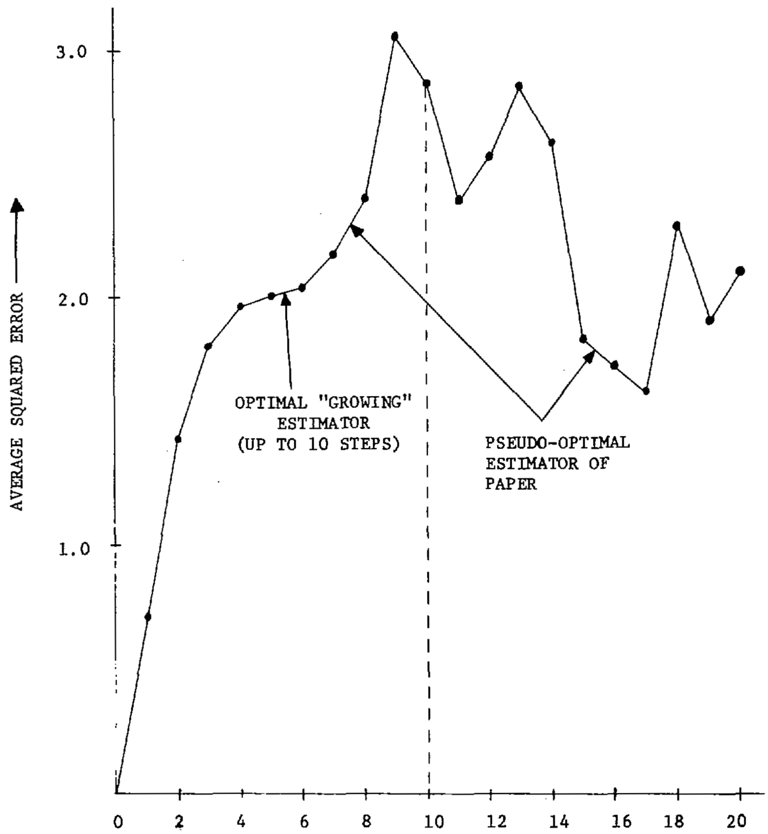

Fig. 3. Average squared error (Markov γ_k).

References

[1] D. Middleton and R. Esposito, "Simultaneous optimum detection and estimation of signals in noise," *IEEE Trans. Inform. Theory*, vol. IT-14, pp. 434–444, May 1968.
[2] N. E. Nahi, "Optimal recursive estimation with uncertain observation," *IEEE Trans. Inform. Theory*, vol. IT-15, pp. 457–462, July 1969.
[3] R. E. Kalman, "A new approach to linear filtering and prediction problems," *Trans. ASME, J. Basic Eng.*, ser. D, vol. 82, pp. 35–45, Mar. 1960.
[4] G. A. Ackerson and K. S. Fu, "On state estimation in switching environments," *IEEE Trans. Automat. Contr.*, vol. AC-15, pp. 10–17, Feb. 1970.
[5] Y. C. Ho and R. C. K. Lee, "A Bayesian approach to problems in stochastic estimation and control," *IEEE Trans. Automat. Contr.*, vol. AC-9, pp. 333–339, Oct. 1964.
[6] A. G. Jaffer and S. C. Gupta, "Recursive Bayesian estimation with uncertain observation," *IEEE Trans. Inform. Theory* (Corresp.), vol. IT-17, pp. 614–616, Sept. 1971.
[7] A. P. Sage, *Optimum Systems Control*. Englewood Cliffs, N. J.: Prentice-Hall, 1968.

An Evaluation of Adaptive Step-Size Random Search

LEE J. WHITE, MEMBER, IEEE, AND ROBERT G. DAY, MEMBER, IEEE

***Abstract*—An evaluation is made of a proposed adaptive step-size random search method to minimize a quadratic function, and the computational results are replicated and improved. Comparing this method to the Fletcher–Powell technique, it is shown that the required number of function evaluations grow linearly with dimension for both algorithms, with the Fletcher–Powell method superior to the proposed random search technique. In addition, several interesting characteristics of the Fletcher–Powell method are experimentally noted.**

Manuscript received August 17, 1970; revised May 7, 1971. Paper recommended by J. M. Mendel, Chairman of the IEEE S-CS Adaptive and Learning Systems, Pattern Recognition Committee. This work was supported in part by NSF Grant GN 534.1.

L. J. White is with the Department of Computer and Information Science, Ohio State University, Columbus, Ohio, 43210.
R. G. Day is with the Digital Equipment Corporation, Maynard, Mass. 01754.

Introduction

Recently Schumer and Steiglitz [4] proposed a method of adaptive step-size random search (ASSRS) for minimizing a given function. Their criterion for comparison was the number of function evaluations required to locate the minimum of the function to within some prespecified accuracy. This criterion is one that is generally accepted; e.g., see Box [1].

Schumer and Steiglitz used their ASSRS algorithm to minimize a number of test functions, where the number of function evaluations required were compared to those of the Newton–Raphson method and the simplex technique of Nelder and Mead [3]. The functions they used were the hypersphere, $F(x) = \sum_{i=1}^{n} x_i^2$, the hyperellipsoid, $F(x) = \sum_{i=1}^{n} a_i(x_i)^2$, and the function $F(x) = \sum_{i=1}^{n} a_i(x_i)^4$. For the hypersphere, it was found that when the dimension n exceeded 78, ASSRS required fewer function evaluations than the Newton–Raphson method to reduce the function to within 10^{-8}. Similar results were obtained for the other functions, and for comparisons to the simplex method. In addition, a theoretical analysis was made for the hypersphere function, which proved the asymptotic superiority of ASSRS to the Newton–Raphson method in terms of function evaluations.

Intuitively, these results seem surprising. The implication is that when the dimension of the function to be minimized is sufficiently large, this minimum can be located more effectively by disregarding all information that can be obtained from the function values and gradients at previous iterations, and proceeding at random, rather than using this information in some deterministic manner.

It is the purpose of this short paper to show that the Fletcher–Powell method [2] is superior to ASSRS even for large dimensional problems. The Fletcher–Powell method is well known for its property of converging to the minimum of a quadratic function in n iterations, if one neglects computer round-off error. The number of function evaluations per iteration are $3(n+1)$. However, we wish to show that for the class of functions considered by Schumer and Steiglitz, the number of iterations required to converge to within some prespecified value of the minimum of a quadratic function are nearly independent of n.

Adaptive Step-Size Random Search

Schumer and Steiglitz proved in [4] that, for the hypersphere function $F(x) = \sum_{i=1}^{n} x_i^2$, the number of function evaluations required to obtain $F(x) < \epsilon$, for some prespecified ϵ (such as 10^{-8}), are asymptotically a linear function of n as n becomes large. The algorithm they consider is as follows.

The search for the minimum is started at some initial point v_0, $i = 0$, and $F_0 = F(v_0)$. A random step u_i of unit length is found from the current position v_i. This random step is obtained by selecting a vector uniformly from the surface of a unit hypersphere. A scalar s_i is determined so as to maximize the expected reduction of $F(v_i)$ and is defined as the *optimum step size*. If the scaled random step reduces F, v_{i+1} is set to $(v_i + s_i u_i)$; if not, v_{i+1} is set to v_i. The iteration is repeated until $F(v_i) < \epsilon$.

An expression was found for the maximum expected improvement

$$I_{\max} = k/n, \qquad k = 0.406.$$

Using this expression, they find the number of function evaluations m for which $F_m = F(v_m) < \epsilon$ to be asymptotically given by

$$m = [-(1/k)\log(F_m/F_0)]n.$$

Thus if $F_m < \epsilon$, and F_0 is a constant, then the number of function evaluations are asymptotically a linear function of n.

However, the necessity of finding an optimum step size at every iteration requires excessive computation. Thus Schumer and Steiglitz developed a practical version of this algorithm called adaptive step-size random search. This version is next discussed.

An initial value of s for the step size is chosen before each iteration. Two random steps of size s and $s(1+a)$ are taken, where a is a uniform random number $0 < a < 1$. The step size that produces the most improvement in the function is chosen as the step size for the next iteration. If neither step results in improvement, the step size remains unchanged and the iteration is repeated. If no improvement is obtained for p successive iterations, the step size is multiplied by a factor $\delta < 1$ and the sequence is repeated. Thus the algorithm tends to adjust the step size toward its optimum value. In addition, each time q iterations have passed, a step with the current step size is compared to a step K times as large. This test guards against the possibility that the step size has become excessively small.

The present authors programmed the ASSRS algorithm in PL/1 and tested this code on an IBM 360/75 computer in order to replicate the experiments of Schumer and Steiglitz. For each dimension tested, five independent trials were made. The starting point for each trial was determined by choosing a point in a random direction one unit away from the origin so the initial value F_0 was the same for each trial. The procedure was terminated when $F < \epsilon = 10^{-8}$. Since the parameters p, δ, q, and K were not specified by Schumer and Steiglitz, through extensive experimentation the authors found that the following values yielded the best results:

$$p = 5, \qquad \delta = 1/2, \qquad q = 50, \qquad K = 50.$$

For the case of a hypersphere function, Fig. 1 shows that not only were the experimental results replicated but also improved upon. The authors obtained a slope of 62, somewhat less than the slope of 80 in the original experiments. This improvement is probably partly due to the use of different algorithm parameters, and although the same stopping criterion was used, also partly due to the selection of the initial vector. Schumer and Steiglitz used an initial vector of ones, whereas the authors used a unit vector of random direction, since a vector of ones corresponds to an increasing distance from the origin with increase in dimension. Recalling the argument requiring F_m/F_0 to be constant, the effect of the increasing distance of the initial position may be a very slight convexity observed in their data.

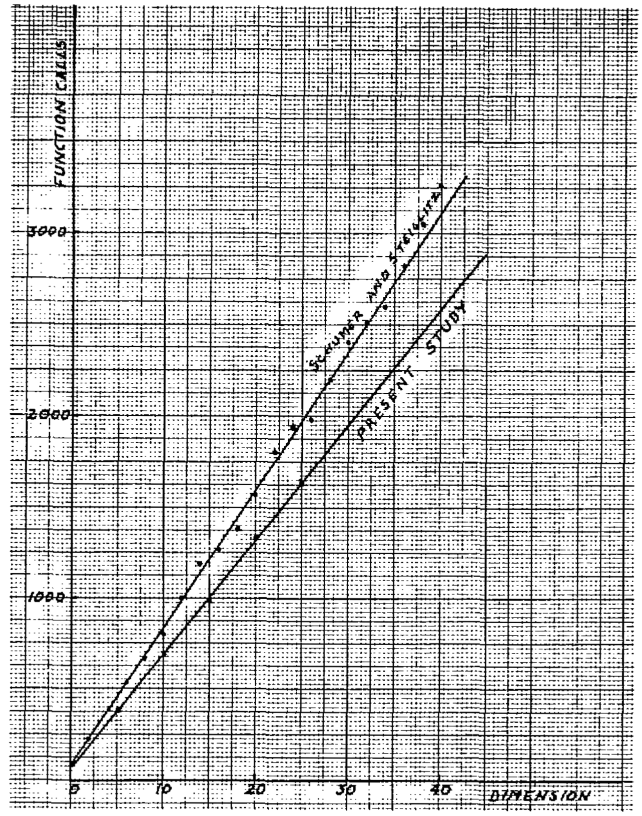

Fig. 1. Comparison of the results of Schumer and Steiglitz to the present study for ASSRS applied to a hypersphere, convergence criterion $F < 10^{-8}$.

The other experimental results of Schumer and Steiglitz were also replicated with slight improvement by the authors.

Experimental Results—A Comparison

In order to compare the ASSRS algorithm to that of Fletcher–Powell [2], a program called FLEPOW was written in PL/1 and implemented on the IBM 360/75 computer. The termination condition was decreased to $\epsilon = 10^{-12}$, and in order to compare these algorithms for large dimensions, each function was minimized by FLEPOW for $n = 20$ to 240, in steps of 20 dimensions. The slopes of the lines for ASSRS had to be adjusted for the different termination condition and were found by means of the expression for slope

$$-(1/k)\log(\epsilon/F_0).$$

Fig. 2 shows a comparison of ASSRS and FLEPOW for the hypersphere function. Note that clearly the Fletcher–Powell algorithm generates a number of function calls that are linear with dimension in order to satisfy the terminating condition $F_m < \epsilon = 10^{-12}$. The data points determining this linear relation are labeled by the number of iterations required for FLEPOW to terminate. The required number of iterations grow so slowly with n that one

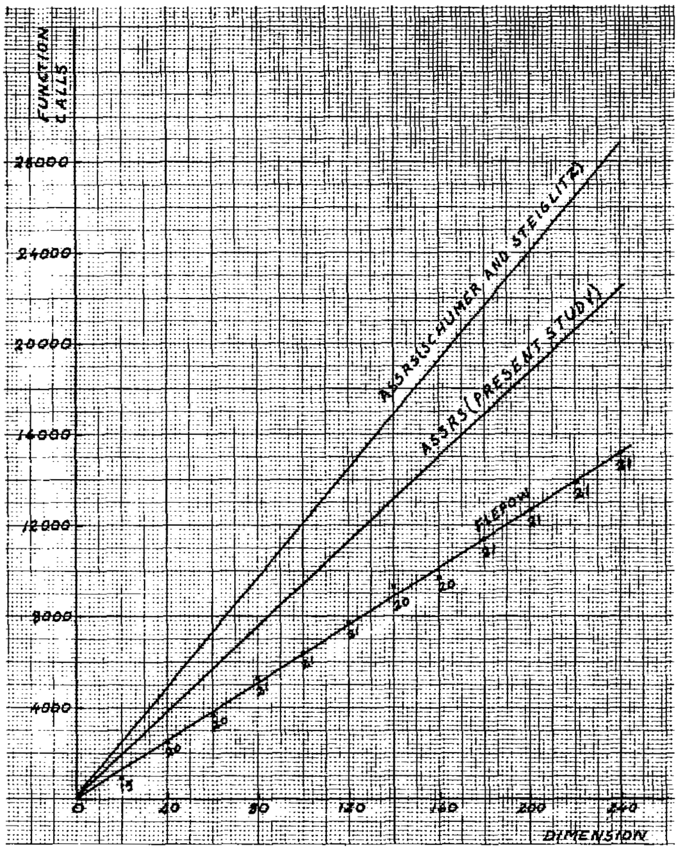

Fig. 2. Comparison of ASSRS and FLEPOW for a hypersphere function; $F < 10^{-12}$.

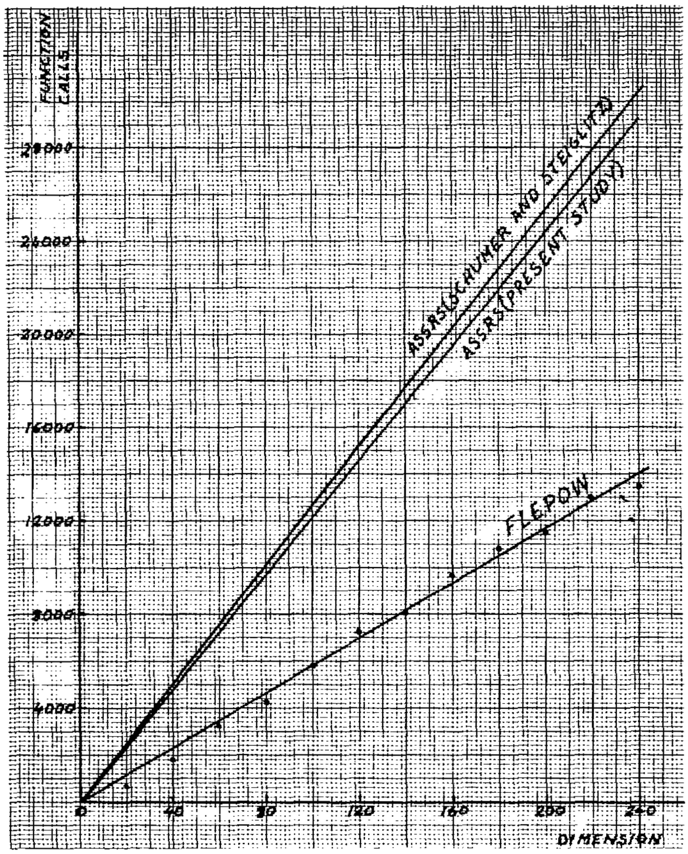

Fig. 3. Comparison of ASSRS and FLEPOW for a hyperellipsoid function; $F < 10^{-12}$.

might easily consider a bound on the number of iterations to be dependent upon ϵ, but independent of n.

Fig. 3 illustrates a similar comparison for the hyperellipsoid function $F(x) = \sum_{i=1}^{n} a_i(x_i)^2$, where the coefficients a_i were randomly chosen from a uniform distribution on [0.1, 1], as done by Schumer and Steiglitz. Clearly, in both studies the Fletcher–Powell algorithm is superior, with the slopes shown in Table I.

It might be argued that if the parameters in the ASSRS algorithms were optimized, a line of lower slope than of the line for FLEPOW could be obtained. At least for the hypersphere, this argument is not valid. Schumer and Steiglitz found that for the optimum step size, asymptotically that slope is

$$-(1/k) \log (\epsilon/F_0).$$

Substituting $k = 0.406$, which they obtained, the smallest slope for any choice of ASSRS parameters is

$$-(1/0.406) \log (10^{-12}/1) = 68.$$

Consequently, for any choice of parameters for the ASSRS algorithm, FLEPOW will require fewer function evaluations than this realization of ASSRS. Although no similar analysis has yet been made for hyperellipsoid functions, Fig. 3 shows such a drastic difference between FLEPOW and ASSRS in terms of the number of function evaluations that it does not appear that the parameters of ASSRS could be optimized so as to reverse this superiority.

Further Experiments with FLEPOW

In an elaboration on the observation that the number of function evaluations are linear with dimension n for the Fletcher–Powell algorithm, the necessity becomes evident for the number of required iterations to be independent of n, since $3(n + 1)$ function evaluations are made for each iteration. The number of iterations are constant if the objective function is reduced by a constant percentage with each iteration, i.e.,

TABLE I

Method	Slope for Hypersphere	Slope for Hyperellipsoid
ASSRS (Schumer and Steiglitz)	120	126
ASSRS (present study)	94	122
Optimal step-size random search	68	not known
Fletcher-Powell	63	58

$$F_{i+1}/F_i = A$$

where A is constant. Then if F_0 and F_m are the initial and final values, respectively,

$$F_m = A^m F_0$$

where m is the required number of iterations.

To determine experimentally whether this hypothesis is true, graphs of the objective function versus iterations required were plotted on semilog paper for $n = 220$ and 240. Fig. 4 shows the result of this experiment for the hypersphere function; similar results were obtained for the hyperellipsoid and other functions.

It can be seen that after the first two or three iterations, log F is quite linear with respect to required iterations. Since we are dealing with quadratic functions, their minima will be found to be within the limits imposed by round-off error in n iterations by FLEPOW. It has been observed that after about $n/2$ iterations, F begins to decrease extremely rapidly, and above that point, the linear relationship no longer holds.

Conclusions

In this paper the authors have replicated the computational results of the ASSRS algorithm of Schumer and Steiglitz. These results showed that, at least for a very narrow class of functions to be

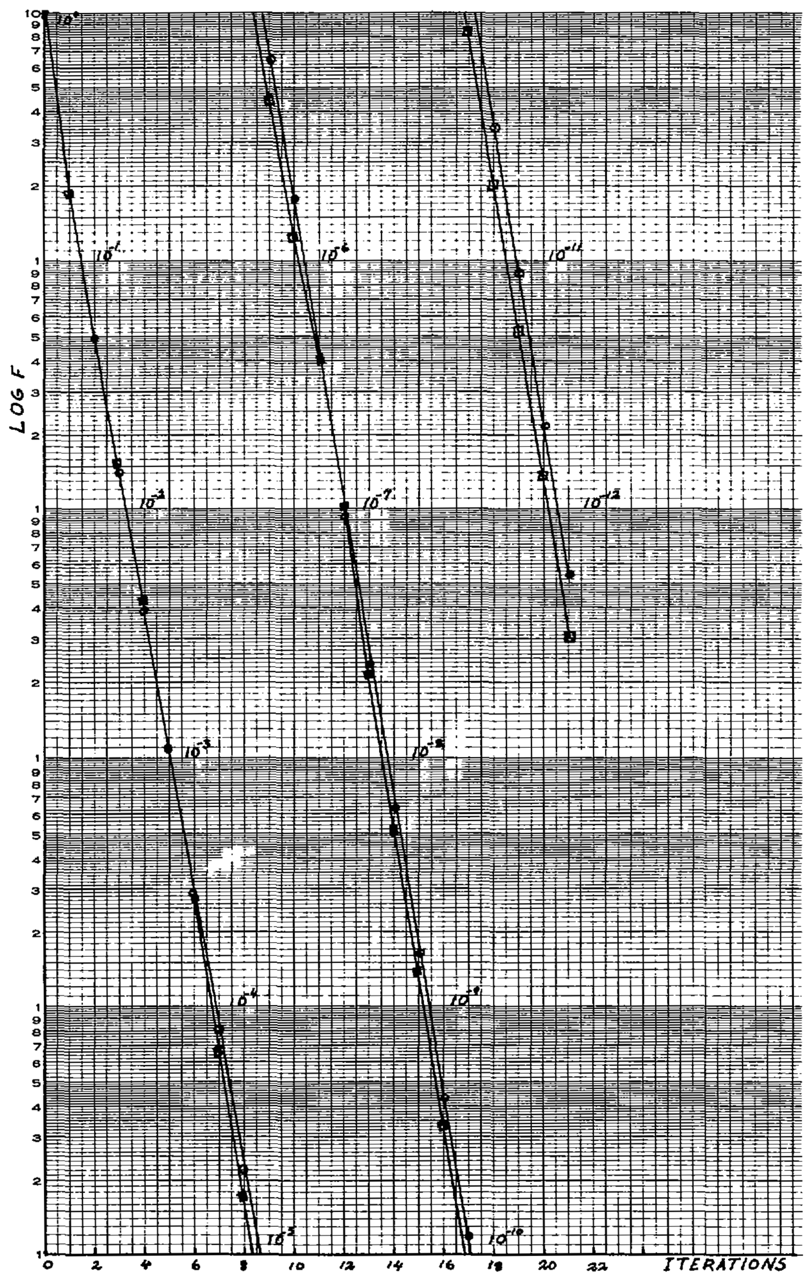

Fig. 4. Log F versus iterations for a hypersphere function.

minimized, the number of function evaluations required to reduce the function to within a prespecified value of the minimum grows only linearly with dimension. It has been shown that the variation of the parameters of ASSRS can improve the performance of the algorithm.

FLEPOW, a computer implementation of the Fletcher–Powell algorithm, has been applied to the same class of functions considered by Schumer and Steiglitz. It would appear that for large dimensions FLEPOW is superior to ASSRS for any adjustment of ASSRS parameters. However, this conclusion may not hold if the distribution of test points on the unit sphere is made adaptive as well as the step size. Although no indication was given in [4], the random optimization method may be applicable to a wider class of search problems than the Fletcher–Powell method, e.g., noisy or multimodal surfaces.

The number of function evaluations for FLEPOW were experimentally shown to be linear with dimension for a fixed terminating function value. This introduced the interesting observation that the number of iterations required to reduce the function below the specified value ϵ are asymptotically independent of n, at least for wide ranges for ϵ. More research is needed in order to characterize mathematically why this should occur.

References

[1] M. J. Box, "A comparison of several current optimization methods and the use of transformations in constrained problems," *Comput. J.*, vol. 9, pp. 67–77, May 1966.
[2] R. Fletcher and M. J. D. Powell, "A rapidly convergent descent method for minimization," *Comput. J.*, vol. 6, pp. 163–168, 1963.
[3] J. A. Nelder and R. Mead, "A simplex method for function minimization," *Comput. J.*, vol. 7, pp. 308–312, Jan. 1965.
[4] M. A. Schumer and K. Steiglitz, "Adaptive step size random search," *IEEE Trans. Automat. Contr.*, vol. AC-13, pp. 270–276, June 1968.