Papers, Volume 1 1962-1985

Thoft-Christensen, Palle

Publication date:
2006

Document Version
Publisher's PDF, also known as Version of record

CHAPTER 4

APPLICATION OF STOCHASTIC METHODS IN OPTIMIZATION[1]

P. Thoft-Christensen, Aalborg University, Denmark

1. INTRODUCTION

It is well known that a great number of engineering problems are intimately connected with problems of optin1ization although most engineering problems probably are solved by a good intuition rather than by a good knowledge of optimization techniques. The reason for this situation may be found in a lack of experience in the use of digital computers. But to-day access to fast digital computers have stimulated a great deal of research in the field of optimization and one can expect that this field will become more important for engineers in the future.

It is also well known that the theory of non-linear optimization still contains many unsolved problems. One of the most difficult problems by all optimization methods is to obtain global rather than local optimal estimates. This problem is detailed discussed in this paper and a rather new strategy by Thoft-Christensen & Hartmann [1] is presented.

The purpose of this paper is furthermore to review some of the most common methods of parameter optimization with emphasis on the random search techniques. The advantages and disadvantages by using random methods compared with non-random methods are stressed and the different types of random methods are compared with each other. Finally some suggestions for future research in this field are given.

2. METHODS OF NON-LINEAR OPTIMIZATION

The discussions in this paper will for the sake of simplicity be restricted to the area of parameter optimization, that is the problem of finding minimum (or maximum) for a so-called *criterion function*

[1] VII International congress on the Application of Mathematics in Engineering, Weimar, DDR, June 1975.

$$F: \quad M \subset R^n \rightarrow R.$$

The ordered set of n parameters $\bar{x}_0 \in M$ is called a *global minimum* for the function *F*, if

$$F(\bar{x}) \geq F(\bar{x}_0) \quad for \quad all \quad \bar{x} \in M,$$

where *M* is called the *feasible region* for the function *F*.

If there exists a neighbourhood $\omega(\bar{x}_0)$ of the point $\bar{x}_0$ so that

$$F(\bar{x}) \geq F(\bar{x}_0) \quad for \quad all \quad \bar{x} \in M \cap \omega(\bar{x}_0)$$

then $\bar{x}_0$ is called a *local minimum* for the function *F*.

In the classical theory of minima (and maxima) a necessary condition for a point $\bar{x}_0 \in M$ to be a local minimum, (or maximum) for the function *F* of class C^2 is that

$$\frac{\partial F(\bar{x}_0)}{\partial \bar{x}_i} = 0 \quad for \quad i = 1, 2, ..., n \tag{1}$$

A point $\bar{x}_0 = (x_1, x_2, \ldots, x_n)$ that satisfies (1) is called a stationary point and its nature is determined by the following determinants

$$\Delta_k(\bar{x}_0) = \begin{vmatrix} \frac{\partial^2 F}{\partial x_1^2} & \cdots & \frac{\partial^2 F}{\partial x_1 \partial x_k} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 F}{\partial x_k \partial x_1} & \cdots & \frac{\partial^2 F}{\partial x_k^2} \end{vmatrix}, \; k = 1, 2, ..., n.$$

It can be shown that a stationary point is a local minimum if

$$\Delta_k(\bar{x}_0) > 0 \quad \text{for} \quad k = 1, \; 2, ..., \; n.$$

It is clear that determination of local or global minimum based on this analytic method can be very difficult if the number of parameters is not very small. Further complications arise if equality or inequality constraints are introduced or if the criterion function *F* is not of class C^2.

In the literature a great number of different methods of optimization are discussed. One way of making a classification is the following division:

a. Random methods,

b. Non-random methods.

But it must be noticed that a systematic classification usually is very difficult as the methods often overlap both with regard to technique and with regard to applicability.

It is outside the intention with this paper to give a comprehensive review of all methods of optimization. For this purpose it can be referred to the books by Pierre [2], Lavi & Vogl [3], Lootsma [4], Kuester & Mize [5], Wilde [6], Wilde & Beightler [7], Kowalik & Osborne [8] and many others.

In section 3 some non-random methods are briefly treated and in section 4 - 6 random methods are more detailed discussed.

3. NON-RANDOM METHODS

An excellent presentation and discussion of the theoretical foundation for the steepest descent methods is given by Sargent & Sebastian [9] on the basis of a work by Davidson [10].

The intuitive ideas behind the classical steepest descent methods (or the hill climbing methods) are the following. An initial starting point $\bar{x}_0$ for the search for minimum is given. The next point $\bar{x}_1$ in the search is the sought by proceeding in the direction which gives the greatest decrease in $F(\bar{x})$, that is in the direction of the steepest slope. The gradient

$$\bar{g}(\bar{x}_0) = \left(\frac{\partial F(\bar{x}_0)}{\partial x_1}, \ldots, \frac{\partial F(\bar{x}_0)}{\partial x_n}\right)$$

can be shown to be the sought direction. However, this direction is not always the "best". In the so-called *DFP-method* presented below the gradient direction is not the search direction.

Following the presentation by Sargent & Sebastian [9] the "new" point $\bar{x}_{k+1}$ is determined from the "old" point $\bar{x}_k$ by the equation

$$\bar{x}_{k+1} = \bar{x}_k - \alpha_k \bar{\bar{S}}_k \bar{g}_k = \bar{x}_k + \bar{p}_{k+1}\ , \tag{2}$$

where the n×n matrix $\bar{\bar{S}}_k$ is defined by the recursion formula

$$\bar{\bar{S}}_k = \bar{\bar{S}}_{k+1} - \frac{\bar{\bar{S}}_{k-1}\bar{q}_k \bar{q}_k{}^T \bar{\bar{S}}_{k-1}}{\bar{q}_k{}^T \bar{\bar{S}}_{k-1} \bar{q}_k} + \frac{\bar{p}_k \bar{p}_k{}^T}{\bar{p}_k{}^T \bar{q}_k}\ , \tag{3}$$

where $\bar{q}_k = \bar{g}_k - \bar{g}_{k-1}$ and α_k is chosen so the function Φ, where

$$\Phi(\alpha) = F(\bar{x}_k - \alpha \bar{\bar{S}}_k \bar{g}_k) \tag{4}$$

is minimum.

The steps in this algorithm are as follows:

(1) A starting point $\bar{x}_0$ is selected.

(2) A search direction $\bar{\bar{s}}_0 = -\bar{g}_0$, where $\bar{\bar{S}}_0$ can be chosen arbitrarily (e.g. the unit matrix $\bar{\bar{I}}$) and $\bar{g}$ is the gradient vector $\bar{g}(\bar{x}_0)$, is computed.

(3) A one-variable minimization in the direction chosen by the previous step is conducted on the function

$$\Phi(\alpha) = F(\bar{x}_0 - \alpha \bar{\bar{S}}_0 \bar{g}_0)$$

to determine α_0.

(4) A "new" point $\bar{x}_1$ is computed by the formula

$$\bar{x}_1 = \bar{x}_0 - \alpha_0 \bar{\bar{S}}_0 \bar{g}_0.$$

This procedure is then continued on the basis of the formulae (2), (3) and (4). Fletcher & Powell [11] proved that $\bar{\bar{S}}_k$ is positive definite provided $\bar{s}_0 = -\bar{g}_0$ is positive definite and that the denominators in (3) are non-zero. Further, they proved that a function decrease is obtained at every step. The last assertion is only valid if the function $\Phi(\alpha)$ is uni-modal.

If the objective function is quadratic it can be shown that by the method just described the minimum will be obtained in at most n steps. A detailed treatment of this method can be found in e.g. Jacoby, Kowalik & Pizzo [12].

It is generally believed that the DFP-method is a very highly efficient search technique with fewer disadvantages than the usual "steepest descent" techniques. Before leaving the non-random methods a brief presentation of the conjugate direction method by Fletcher & Reeves [13] will be given.

By their method the search direction is determined by the formula

$$\bar{s}_{k+1} = -\bar{g} + \frac{\bar{g}_k^{\,T}\bar{g}_k}{\bar{g}_{k-1}^{T}\bar{g}_{k-1}}\bar{s}_k$$

with $\bar{s}_0 = -\bar{g}_0$. As before $\bar{g}_k = \bar{g}(\bar{x}_k)$. This method can be shown to have some of the advantages of the DFP-method and further it is simpler to work with and its storage requirement is lower. With a quadratic objective function *F* the directions $\bar{s}_k$, *k* = 0, 1, . . . , n-1 are conjugate, so the minimum will also by this method be located in n steps. For a non-quadratic function the algorithm must be restarted for every (n+ 1) steps with $\bar{s}_0$ equal to the steepest descent direction.

It is commonly asserted that the DFP-algorithm is more successful than the method by Fletcher and Reeves. There exists a great number of experiments with the two methods and most of them confirm this statement although it always is possible to find a specific function for which a given method seems to be better in some sense than all other methods.

In case the derivatives cannot be evaluated from analytical formulae finite difference approximations must be used and truncation errors and rounding errors will occur. The performance of the gradient methods may then be critical.

4. PURE RANDOM METHODS

When gradient searches can be used, that is when the objective function *F* has continuous derivatives on the feasible region $M \subset R^n$. These methods are generally more efficient as compared with random methods. But in some situations such as the case when the objective function *F* is not even continuous random methods can often be used with success, but an enormous number of function evaluations are usually needful. Further a very slow rate or convergence is often seen by these methods, although it is theoretically possible to show that they converge to the global solution with probability one.

In spite of these reservations we shall see in the following that the random methods also have their advantages.

The most primitive types of random search methods are the so-called pure random methods. Such methods are apparently first proposed by Brooks [14]. The method is very simple to program even in the case of an objective function of many variables but a great number of function evaluations are needed.

The idea behind this method is to select a number of points from a uniform probability distribution over the feasible region. The function value is calculated in each of these points and the smallest value is chosen as the minimum. Unfortunately the number of points to select is going up exponentially with the dimension n (see White [15] and Spang [16]).

The only practical application of the pure random method is perhaps to use it in choosing a starting point for the creeping methods.

5. CREEPING RANDOM METHODS

Some fundamental investigations concerning the convergence of creeping random methods made ten to fifteen years ago by Rastrigin [17] and Gurin & Rastrigin [18] gave rise to an extensive research in random methods. The main results of these investigations seem to indicate that some creeping random methods are superior to the steepest descent method for problems with many parameters. On the other hand later experiments with the non-random methods presented in section 3 have shown the opposite result. The only tenable conclusion to day regarding this problem is that the random methods in same situations may by superior to non-random methods with regard to convergence. But, as shown later, there may exist other more serious reasons for choosing a random method.

The basic principle in the creeping random search can be illustrated by the following very simple method. A starting point $\bar{x}_0$ in the feasible region *M* for the objective function $F : M \to R$ and a fixed step size (step length) is chosen. The first step $\Delta\bar{x}_0$ is supposed to have a random direction and the fixed step size. If $F(\bar{x}_0 + \Delta\bar{x}_0) < F(\bar{x}_0)$ the trial step $\Delta\bar{x}_0$ is called a *success*, and the point $\bar{x}_1 = \bar{x}_0 + \Delta\bar{x}_0$ is used as the starting point for the next step in the search. If $F(\bar{x}_0 + \Delta\bar{x}_0) \geq F(\bar{x}_0)$ the trial step $\Delta\bar{x}_0$ is called a *failure* and the old point $\bar{x}_0$ is again used as starting point for a new random trial step until a success is found. The method continues in this way until the search is terminated by some stopping rule. This simple algorithm can be described by the formula

$$\bar{x}_{i+1} = \bar{x}_i + \varepsilon_i \Delta\bar{x}_i$$

where

$$\varepsilon_i = \begin{cases} 1 & \text{if} \quad F(\bar{x}_i + \Delta\bar{x}_i) < F(\bar{x}_0) \\ 0 & \text{if} \quad F(\bar{x}_i + \Delta\bar{x}_i) \geq F(\bar{x}_0) \end{cases}$$

If the random trial step $\Delta\bar{x}_i$ is a failure, then it is easily seen that the step $-\Delta\bar{x}_i$ with opposite direction is a success if the objective function can be approximated by a linear function in a neighbourhood of $\bar{x}_i$. The *reverse search* can be presented by

$$\bar{x}_{i+1} = \bar{x}_i + \varepsilon_i \Delta\bar{x}_i$$

where

$$\varepsilon_i = \begin{cases} 1 & \text{if} \quad F(\bar{x}_i + \Delta\bar{x}_i) < F(\bar{x}_i) \\ -1 & \text{if} \quad F(\bar{x}_i + \Delta\bar{x}_i) > F(\bar{x}_i) \end{cases}$$

It is a characteristic of the methods presented above that the step size is fixed and that the step direction is chosen random. It is clear that information about the behaviour of the function in the neighbourhood of the point $\bar{x}_i$ is used only very little for the future steps. The first algorithm at all uses no local information and by the second only the reverse property is used. It is therefore likely to expect better performance if the step direction is not chosen random and if the step size is not fixed. It is as an example reasonable that a successful step should affect the next step with regard to its direction. Further intuition indicates that the step size should be decreased when the search is close to the minimum to improve convergence. It is important to notice that the considerations here only are valid if the nearest minimum is considered that is if the serious problem of finding global rather than local minimum is neglected. In section 6 problems concerning global minimum is closely discussed.

Directional adaptation can be introduced by some kind of correlation between past successful steps and future trial steps. In this way some local property for the function *F* is taken into account as it is done in the non-random methods. A simple way of doing it is used by de Graa [19] who defined the next step $\Delta\bar{x}_i$ as a linear combination of e.g. the last successful step $\Delta\bar{x}_{i-1}$ and some kind of random vector $\bar{r}_i$ that is by the formula

$$\Delta\bar{x}_i = a\Delta\bar{x}_{i-1} + b\bar{r}_i \tag{5}$$

where *a* and *b* are constants. He records a substantial improvement in the number of function evaluations.

A more general algorithm than (5) was earlier suggested by Matyas [20] based on the same idea namely a kind of combination of a vector depending of old successful steps and a random vector.

There exists of course a lot of different ways of making directional adaptation (see e. q. White [15]). Thoft-Christensen & Hartmann [1] suggests the following directional adaptation. Let $\bar{x}_i$ be the last determined point in the search and let the step length $|\Delta\bar{x}_i|$ be given. On the hyper sphere

$$|\bar{x} - \bar{x}_i| = |\Delta\bar{x}_i| \tag{6}$$

a number of points are then chosen until a successful step $\Delta\bar{x}_i$ is found by some probability distribution defined on the hyper sphere. In experiments with two-parameter test functions the directional adaptation was carried out by letting the directional angle of the new trial step follow the normal probability distribution N[α, *v*], that is with a mean equal to the directional angle α for the last successful step and a standard deviation *v*. For three-parameter tests the directional adaptation was accomplished by use of a po1ar description of the steps. The directions of the next trial steps are now given by two angles θ and φ, which are assumed to follow normal probability distributions with means equal to the corresponding angles of the last successful step and with a given standard deviation. In the general n-parameter situation the following form for directional adaptation was used. Let $\Delta\bar{x}_{i-1}$ be the last successful step and let the coordinates of its normalized directional vector be $(\alpha_1,...,\alpha_n)$. The direction of the next trial step is then determined by a normalized vector with coordinates given by n normal probability distributions with means equal to α_i i=1,...,n and standard deviations depending of the desired degree of directional adaptation. This way of

determining the step directions has been used with success on test functions with up to 15 parameters.

It is more difficult to choose an effective algorithm with regard to the step size. The reason for this will appear from the following considerations. To-improve convergence the step size should be decreased when the search is close to the minimum. But it is very difficult to judge during the search how close the search is to the minimum, because only discrete values of the objective function are at hand. And a small change here does not guarantee that the random search is close to the minimum, as the objective function may have a flat part. In the random methods presented in the literature several procedures for step-size adjustment have been used. Some authors increase the step size if a success takes place and decrease it if none occurs within some number of trials. Other authors reduce the step- size following a number of failures. There seems to be no universal answer to this problem (see Matyas [20] and Schumer & Steiglitz [21].

The strategy proposed by Thoft-Christensen & Hartmann [1] is based on a step-size adjustment, which is not very critical. After arriving in a point $\bar{x}_i$ a prescribed (small) number of points on the hyper-sphere (6) with the radius $|\Delta\bar{x}_i|$ is checked. If no successful step is found the length of the trial steps (and therefore the radius of the hyper-sphere) is bisected and doubled. In this way two sets of hyper spheres with radius

$$|\Delta\bar{x}_i|/2,\ |\Delta\bar{x}_i|/4,\ |\Delta\bar{x}_i|/8,\ldots$$

and

$$2|\Delta\bar{x}_i|,\ 4|\Delta\bar{x}_i|,\ 8|\Delta\bar{x}_i|,\ldots$$

are defined and the tracking of a successful trial step is continued on these spheres one after each other. When a successful step is located the same procedure continues in the new point $\bar{x}_{i+1} = \bar{x}_i + \Delta\bar{x}$, but with a step length modified by some kind of gain-dependency.

It is obvious that many different methods of optimization can be constructed on basis of the considerations above. By combining different kinds of directional adaptation with different ways of adjusting the step-size random methods with unpredictable properties appears. The only way of comparing them is to test them on some test functions (see e. g. G. Schack & Borowski [22] and Brooks [23]), but it is very difficult to generalize on basis of such experiments.

6. GLOBAL OPTIMIZATION

All the methods mentioned previous have one very serious drawback, namely that they if no special precautions are taken only work satisfactory if the objective function is unimodal. If the objective function is multimodal there exists a risk that methods based on gradient techniques or on a rapid convergent random strategy only will disclose a local optimum. This will happen when the starting point is outside of the field of influence of the global optimum. So in such problems where the field of influence of the global optimum only is a small part of the feasible region the change of obtaining a local rather than a global optimum is corresponding high.

It is easy to see that these problems are extremely difficult to solve, as the main objective by most methods is to arrive in the nearest minimum as fast as possible and

only the local behaviour of the objective function is used. Therefore it is often stated that the method should be initiated from a number of different starting points (see e.g. Weisma &Wood [24]). Bocharov & is Feldbaum [25] have discussed a number of different methods for locating starting points. This local-global problem is investigated by a great number of authors, e.g. Hill [26], Mc. Cormick [27], and many others, but still all too little information has been published. It usually believed that the random methods in this connection have some potentials not yet fully investigated.

The techniques used by Hill [26] is based on a combination of global and local search technique: initially the search space is divided into a large number of small hyper volumes (e. g. hyper cubes) and in the global part of the search the values of the objective function are evaluated in the centers of each hyper volume. It is then assumed that the objective function is unimodal within any hyper volume so that a local search based on a steep-ascent hill limping technique can be used on some hyper volumes selected by the global search. The number of hyper volumes being searched are selected by assigning to each hyper volume a probability that it contains the global extremum completely different way. The main idea behind the strategy is the use of *local* and *global steps*. The intention with the local steps is to obtain a nigh degree of convergence to the nearest minimum not depending on whether it is a global or a local minimum. So the step sizes for the local steps are “rather small”. On the contrary the global steps are not limited in size as their purpose only is to escape from a local minimum by making sufficiently large steps. The global steps must be arranged in such a way that any point in the feasible region can be reached by a global step from the point in question.

The main idea is thus the division in local and global steps, which not only differ by step size but also in the way the step-directions are chosen.

First a starting point is chosen at random in the feasible region by use of a random number generator. It is not very important to use more than one starting point because of the presence of global steps, but of course the probability of obtaining global estimates is increased if the search is initiated from several starting points.

When a starting point has been chosen the strategy is based on local and global steps in the following way. First some local trial steps are taken with step size and directional adaptation as described in section 5. When a prescribed (small) number of successful local steps are carried out a global trial step follows. The intention with the global steps is to be able to reach any point in the feasible region, so no information on any "good" direction for global steps can be given in general. Therefore the step size is chosen random in a predetermined interval with an upper bound greater than the length of the diagonal of the feasible region. By the same reason the directions of the global steps are chosen "random" by using a uniform probability distribution over a given interval for each of the parameters.

In this way a pre-described number of global trial steps are made until an improvement may happen. In such a case the new point is then used as starting point for local steps. Otherwise further local trial steps are made from the same point but now with new directions. The procedure described above continues until a pre-described number of successful steps have occurred or until no further improvements by local steps can be found under the restrictions given on the number of trial steps.

In Thoft-Christensen & Hartmann [1] some experimental results with this new strategy on unimodal and multimodal test functions with two and three parameters are presented. Recently the strategy has been used on test functions with up till 15 parameters.

An example from [1] with the following figures from a typical experiment with the function

$$F(x,y,z) = 2 - \exp\left(-\frac{1}{4\pi}\sqrt{x^2+y^2+z^2}\right)\cos x \cos y \cos z$$

shows the importance of the global steps:

Step no:	*x*	*y*	*z*	*F(x,y,z)*
0	-1.000	10.000	10.000	1.877
8	-3.317	5.916	-2.771	1.522
17	-3.122	6.161	-3.160	1.458
18	-3.203	0.094	3.131	1.304
29	-3.099	0.007	3.071	1.296
30	-0.074	0.577	-0.448	1.289
53	0.000	0.005	0.015	1.001
60	0.001	-0.001	-0.000	1.000

After 8 steps the search is trapped by a local minimum and the next 10 steps are used to improve the local minimum value (F_5 = 1.458). Step no. 18 is a global step that brings the search to a better local minimum that is reached by step no. 29 (F_5 = 1.296). Finally the global step no. 30 makes it possible to reach the global minimum (F_5 = 1.000). A CDC CYBER 72 computer belonging to Aalborg University is used and the random numbers are generated with the following linear congruential generator

$$x_{i+1} = 9901\, x_i \pmod{2^{39}+1}.$$

Normal variates with mean 0 and a standard deviation equal to 1 is generated by the formula

$$x = \sqrt{-2\ln\ r_1}\cos(2\pi r_2)$$

where r_1 and r_2 are independent random variables in the interval [0; 1] calculated by the random number generator described above (see Newmann & Odell [28]).

7. DISCUSSION AND CONCLUSIONS

One must remember that for many users of optimization algorithms the only important thing is to be able to find in the computer library an algorithm that is reliable, such a user is not concerned at all about computer time. Having this in mind random methods may be worth working with in particular for multimodal functions although they often results in an enormous number of function evaluations. But for some users it is very important to think of efficiency e.g. when optimization algorithms are used very often. In such a case it is useful to know whether one method is "better" than the others. As long as theoretical investigations are as incomplete as to day the only way to collect such information is by numerical experimentation. Examination of algorithms from a theoretical point of view is only possible for very few simple problems, but nonetheless of great importance. It is to be hoped that the intensive research in this field for non-random methods will be transferred to the random methods.

Testing of algorithms by numerical experiments is of course a difficult work if reliable results shall come out as a lot of different parameters can be changed.

Objective functions with both a small and a large number of parameter should be tested and the degree of non-linearity is important. As usually when doing numerical experimentation one can only hope that the numerical results can be generalized to more general situations. Himmelblau [29] summaries that the criteria to be considered in the evaluation of unconstrained algorithms are:

1. Robustness - success in obtaining an optimal solution (to within a certain precision);
2. Number of function (and derivative) evaluations;
3. Computer time to termination (to within the desired degree of precision).

In the same paper a great number of non-random algorithms are compared, but the same kind of evaluations of random methods is still missing.

8. REFERENCES

[1] Thoft-Christensen, P. & U. Hartmann. *A Strategy of Optimization Based on Random Search*. Stochastic Control Symposium, Budapest 1974, pp. 297-303.

[2] Pierre, D. A. *Optimization Theory with Applications*. John Wiley and Sons, N. Y., 1969.

[3] Lavi, A. & T.P.Vogl. *Recent Advances in Optimization Techniques*. John Wiley and Sons, N. Y., 1966.

[4] Lootsma, F. A. *Numerical Methods for Non-Linear Optimization*. Academic Press, London, 1972.

[5] Kuester, J. L. & J. H. Mize. *Optimization Techniques with Fortran*. McGraw-Hill, N. Y., 1973.

[6] Wilde, D. J. *Optimum Seeking Methods*. Prentice-Hall, Englewood Cliffs, N. J., 1964.

[7] Wilde, D. J. & C. S. Beightler. *Foundations of Optimization*. Prentice-Hall, Englewood Cliffs, N.J., 1969.

[8] Kowalik, J. & M. R. Osborne. *Methods For Unconstrained Optimization Problems*. American Elsevier, N. Y., 1968.

[9] Sargent, R. W. H. & D. J. Sebastian. *Numerical Experience with Algorithms for Unconstrained minimization*. pp. 45-68 in [4].

[10] Davidson, W. C. *Variable metric method for minimization*. A.E.C. Research and Development Report, ANL-5990, 1959.

[11] Fletcher, R. & M. J. D. Powelt. *A Rapidly Convergent Descent Method for Minimization*. Computer J., Vol. 6, 1963, pp. 163-168.

[12] Jacoby, S.L.S. J.S.Kowalik & J. T. Pizzo. *Iterative Methods for Nonlinear Optimization Problems.* Prentice Hall. Englewood Cliffs, 1971.

[13] Fletcher, R. & C.M.Reeve. *Function Minimization by Conjugate Gradients*. Computer J., Vol. 7. 1964, pp. 149-154.

[14] Brooks, S.H. *A Discussion of Random Methods for Seeking Maxima*. Operational Research, Vol. 6, 1958, pp. 244-251.

[15] White, R. C. *A Survey of Random Methods for Parameter Optimization*. Dept. Electrical Eng., Techn. Univ., Eindhoven, Netherlands, 1973.

[16] Spang, H. A. *A Review of Minimization Techniques for Non-Linear Functions*. SIAM Review, Vol. 4. 1961, pp. 343-365.

[17] Rastrigin, L.A. *The Convergence of the Random Search Method in the Extremal Control of a Many Parameter System*. Automation and Remote Control. Vol. 24, 1963. pp. 1337-1342.
[18] Gurin, L. S. & L. A. Rastrigin. *Convergence of the Random Search Method in the Presence of Noise*. Automation and Remote Control. Vol. 26, 1965. pp. 1505-1511.
[19] Graag, D.P. de. *Parameter Optimization Techniques for Hybrid Computers*. Sixth AICA/IFIP Conference on Hybrid Computation, Munich, Aug. 31 -Sept. 4, 1970.
[20] Matyas, J. *Random Optimization*. Automation and Remote Control, Vol. 26, 1965, pp. 224- 251.
[21] Schumer, M. A. & K. Steiglitz. *Adaptive Step Size Random Search*. IEEE Trans. Vol. AC-13, 1968. pp. 270-276.
[22] Schrack, G. & N. Borowski. *An Experimental Comparison of Three Random Searches*. pp. 137 -147 in [4].
[23] Brooks, S. H. *A Comparison of Maximum Seeking Methods*. Operations Research. Vol. 7. 1959, pp. 430-457.
[24] Weismann, J. & C. F. Wood. *The Use of 'Optimal Search' for Engineering Design*. in [3].
[25] Bocharov, I. N. & A. A. Feldbaum. *An Automatic Optimizer for the Search of the Smallest of Several Minima (a Global Optimizer)*. Automation and Remote Control. Vol. 23, 1961, pp. 260-270.
[26] Hill, J. D. *A Search Technique for Multimodal Surfaces*. IEEE Trans., Vol. SSC-5. 1969, pp. 2-8.
[27] McCormick, G. P., *Attempts to Calculate Global Solutions of Problems that may have Local Minima.* pp. 209 -221 in [4].
[28] Newmann, T. G. and P. L. Odell. *The Generation of Random Variates*. London, 1971.
[29] Himmelblau, D. M. *A Uniform Evaluation of Unconstrained Optimization Techniques*. pp. 69- 97 in [4].