INTERNATIONAL SYMPOSIUM ON SYSTEMS ENGINEERING AND ANALYSIS

Volume II - Contributed Papers

October 23-27, 1972

PURDUE UNIVERSITY
West Lafayette, Indiana

vol. II, pp. 105-110

Sponsored by:
Purdue University
Schools of Engineering and
Krannert School of Industrial Administration
National Science Foundation

Co-sponsored by:
American Society for Engineering Education
American Society of Civil Engineers
American Institute of Industrial Engineers
American Society of Agricultural Engineers
American Water Resources Association
International Association for Hydraulic Research

THE COMPOUND RANDOM SEARCH ALGORITHM

Luc P.DEVROYE
Dept.of Electrical Engineering
Catholic University of Louvain,Belgium

ABSTRACT

The optimization of a function of many variables is investigated.A new algorithm is proposed : the compound random search algorithm (CRSA).This algorithm combines the features of random search and non-random direct search.Each part of the algorithm is discussed in detail.The CRSA is compared with several other direct search methods in many different problems.The results are promising.Some modifications of the basic search scheme make the CRSA particularly useful and efficient.

1.INTRODUCTION

It is often desired to minimize a function $Q(w_1,\ldots,w_n)$ of several variables w_i.This function is denoted by $Q(\underline{w})$.This paper only deals with unimodal functions.Although only local search algorithms are treated, perhaps the most promising application of random search is found in multimodal optimization.Further,assume that this criterion $Q(\underline{w})$ is not analytically known and that it can only be measured or computed at the expense of much effort.Obviously,the behaviour of $Q(\underline{w})$ is rather unknown.The condition of unimodality isn't always fullfilled and cannot be controlled a priori ! However,suppose for simplicity that noisefree measurements of a unimodal function $Q(\underline{w})$ are available.The generalization towards multimodal functions can be found in [15] .

Several authors have shown that random search compares very well with other direct search techniques[3],[5] in high-dimensional problems with costly function evaluations.The number of function calls needed to solve a problem is roughly spoken proportional to n.A survey of the most commonly used random search algorithms can be found in[6],[8],[9],[14] .The simple random search [1-2] and the fixed step size random search [5] have been improved many times [4-5],[7].This paper is a trial to improve some of these algorithms and to broaden their field of application.

Because the cost of a function evaluation is more important than the cost of search effort,we may not be refrained from using more sophisticated search schemes.So,it is quite normal to use all the a priori available information about $Q(\underline{w})$ in order to accelerate and control the search.Moreover, new information must be gathered during the search and transformed into useful data. There is a remarkable resemblance between this search and a learning process or a game against nature.

A new algorithm is developed which performs much better,even under the worst circumstances,than the powerful methods of Gucker [13] or Rosenbrock [11].Moreover,this method is one of the fastest random optimization methods when an optimum must be localized quite accurately.In our discussion, the rate of convergence is determined in terms of the number of function calls needed to find a solution.The number of iterations is irrelevant.The CRSA combines the features of random search with the advantages of non-random search.

2.DEVELOPMENT OF A NEW METHOD

Assume that the region W,in which $Q(\underline{w})$ is defined,is closed and bounded.Further,let $\underline{w}(j)$ denote the best estimate of the minimum after j iterations.Many random search algorithms are built up as follows:a point $\underline{w}*(j+1)$,which will be called "trial",is generated using the probability density function $f(\underline{w}*(j+1)/\underline{w}(j))$.This distribution may be fixed as in the method of Brooks [1] . In general,however,this distribution is adaptive .Most commonly used is the gauszian distribution with mean $\underline{w}(j)$.Let the covariance matrix be T(j) or,for simplicity, $\sigma^2(j).I$ The trial point is compared with the old "best estimate" (or:basepoint) $\underline{w}(j)$ on the base of the measured or computed values $Q(\underline{w}(j))$ and $Q(\underline{w}*(j+1))$.Obviously,$\underline{w}*(j+1)$ is a better estimate if $Q(\underline{w}*(j+1)) < Q(\underline{w}(j))$. This yields that an auxiliary variable $y(\underline{w}(j),\underline{w}*(j+1))$,which denotes the outcome of the comparison,is put to one.If,on the other hand,$Q(\underline{w}*(j+1)) > Q(\underline{w}(j))$ or $\underline{w}*(j+1)$ violates an inequality constraint or $Q(\underline{w}*(j+1))$ isn't even measurable,then : $y(\underline{w}(j),\underline{w}*(j+1))=0$ .The trial is considered as a failure.Notice that the comparison is a success if $Q(\underline{w}*(j+1))=Q(\underline{w}(j))$,in order to allow random basepoint displacements when $Q(\underline{w})$ has staircase parts,for instance.Without this sign,each point of such a level region would be an isolated local minimum.

THE COMPOUND RANDOM SEARCH ALGORITHM

Luc P.DEVROYE
Dept.of Electrical Engineering
Catholic University of Louvain,Belgium

ABSTRACT

The optimization of a function of many variables is investigated.A new algorithm is proposed : the compound random search algorithm (CRSA).This algorithm combines the features of random search and non-random direct search.Each part of the algorithm is discussed in detail.The CRSA is compared with several other direct search methods in many different problems.The results are promising.Some modifications of the basic search scheme make the CRSA particularly useful and efficient.

1.INTRODUCTION

It is often desired to minimize a function $Q(w_1,\ldots,w_n)$ of several variables w_i.This function is denoted by $Q(\underline{w})$.This paper only deals with unimodal functions.Although only local search algorithms are treated, perhaps the most promising application of random search is found in multimodal optimization.Further,assume that this criterion $Q(\underline{w})$ is not analytically known and that it can only be measured or computed at the expense of much effort.Obviously,the behaviour of $Q(\underline{w})$ is rather unknown.The condition of unimodality isn't always fullfilled and cannot be controlled a priori ! However,suppose for simplicity that noisefree measurements of a unimodal function $Q(\underline{w})$ are available.The generalization towards multimodal functions can be found in [15] .

Several authors have shown that random search compares very well with other direct search techniques[3],[5] in high-dimensional problems with costly function evaluations.The number of function calls needed to solve a problem is roughly spoken proportional to n.A survey of the most commonly used random search algorithms can be found in[6],[8],[9],[14] .The simple random search [1-2] and the fixed step size random search [5] have been improved many times [4-5],[7].This paper is a trial to improve some of these algorithms and to broaden their field of application.

Because the cost of a function evaluation is more important than the cost of search effort,we may not be refrained from using more sophisticated search schemes.So,it is quite normal to use all the a priori available information about $Q(\underline{w})$ in order to accelerate and control the search.Moreover, new information must be gathered during the search and transformed into useful data. There is a remarkable resemblance between this search and a learning process or a game against nature.

A new algorithm is developed which performs much better,even under the worst circumstances,than the powerful methods of Gucker [13] or Rosenbrock [11].Moreover,this method is one of the fastest random optimization methods when an optimum must be localized quite accurately.In our discussion, the rate of convergence is determined in terms of the number of function calls needed to find a solution.The number of iterations is irrelevant.The CRSA combines the features of random search with the advantages of non-random search.

2.DEVELOPMENT OF A NEW METHOD

Assume that the region W,in which $Q(\underline{w})$ is defined,is closed and bounded.Further,let $\underline{w}(j)$ denote the best estimate of the minimum after j iterations.Many random search algorithms are built up as follows:a point $\underline{w}*(j+1)$,which will be called "trial",is generated using the probability density function $f(\underline{w}*(j+1)/\underline{w}(j))$.This distribution may be fixed as in the method of Brooks [1] . In general,however,this distribution is adaptive .Most commonly used is the gauszian distribution with mean $\underline{w}(j)$.Let the covariance matrix be T(j) or,for simplicity, $\sigma^2(j).I$ The trial point is compared with the old "best estimate" (or:basepoint) $\underline{w}(j)$ on the base of the measured or computed values $Q(\underline{w}(j))$ and $Q(\underline{w}*(j+1))$.Obviously,$\underline{w}*(j+1)$ is a better estimate if $Q(\underline{w}*(j+1)) < Q(\underline{w}(j))$. This yields that an auxiliary variable $y(\underline{w}(j),\underline{w}*(j+1))$,which denotes the outcome of the comparison,is put to one.If,on the other hand,$Q(\underline{w}*(j+1)) > Q(\underline{w}(j))$ or $\underline{w}*(j+1)$ violates an inequality constraint or $Q(\underline{w}*(j+1))$ isn't even measurable,then : $y(\underline{w}(j),\underline{w}*(j+1))=0$ .The trial is considered as a failure.Notice that the comparison is a success if $Q(\underline{w}*(j+1))=Q(\underline{w}(j))$,in order to allow random basepoint displacements when $Q(\underline{w})$ has staircase parts,for instance.Without this sign,each point of such a level region would be an isolated local minimum.

The basepoint or best estimate moves towards $\underline{w}*(j+1)$ when $y(\underline{w}(j),\underline{w}*(j+1))=1$;

$$\underline{w}(j+1) = \begin{cases} \underline{w}(j) & \text{if } y(\underline{w}(j),\underline{w}*(j+1))=0 \\ \underline{w}*(j+1) & \text{otherwise} \end{cases} \quad (1)$$

Let us briefly recall the advantages of this random search :

- ease of implementation
- large field of application since $f(\underline{w}*(j+1)/\underline{w}(j))$ can be made adaptive. The variance $\sigma(j)$ or the covariance matrix $T(j)$ are often updated during the search [13]. Sometimes, an adaptive preferable direction $\underline{d}(j)$ is introduced, which denotes that the mean of the distribution of $\underline{w}*(j+1)$ is no longer $\underline{w}(j)$ but $\underline{w}(j)+\underline{d}(j)$ [4] . Special schemes can be found in [3], [6] and [7] .
- when $\sigma(j)$ is fixed, a great variety of functions $Q(\underline{w})$ may be treated, such as staircase functions or highly nonlinear functions. By virtue of algorithm (1), a global optimum of a multimodal function will be localized. Convergence is guaranteed under rather general conditions on $Q(\underline{w})$.
- when $\sigma(j)$ is fixed, the rate of convergence doesn't depend upon the size of the gradient in each point.
- when n is large, this method compares well with other direct search techniques if the number of function calls, that is needed to solve the problem, is taken as the criterion of comparison [3], [5].

In the sequel, we shall assume that $\underline{w}*(j+1)$ is generated according to a normal gauszian distribution with mean $\underline{w}(j)$ and with an adaptive covariance matrix $\sigma^2(j).I$. Unfortunately, the performance of this simple method is not quite excellent when $\underline{w}(j)$ is near the optimum [3] . Some authors suggested that the ideal optimization technique perhaps consisted of a combination of random search and non-random direct search. Therefore, a somewhat different algorithm is proposed which combines the features of random search and non-random direct search (accuracy of the solution).

In each iteration, two trials are made : a random trial in

$$\underline{w}*(j+1) = \underline{w}(j) + \underline{\rho}(j)$$

and a non-random trial in

$$\underline{w}**(j+1) = \underline{w}(j) + \varepsilon(j).\underline{b}(j+1)$$

where $\varepsilon(j)$ and $\underline{b}(j+1)$ are deterministically determined. The algorithm (1) is modified now. Indeed, the new best estimate $\underline{w}(j+1)$ is the best among the points $\underline{w}(j)$, $\underline{w}*(j+1)$ and $\underline{w}**(j+1)$. Clearly, three special problems must be solved: the control of $\sigma(j)$, the control of $\varepsilon(j)$ and the estimation of $\underline{b}(j+1)$. This latter n-dimensional vector has the meaning of a "preferable direction" . It will be on-line computed. This computation necessitates the gathering of some information about $Q(\underline{w})$. Since a function call is rather expensive in the class of problems with which we deal, the necessary information must be provided by the "first stage" search. In general $\underline{b}(j+1)$ depends upon $\underline{b}(j)$, $\underline{\rho}(j)$, $Q(\underline{w}(j))$, $Q(\underline{w}*(j+1))$ and the auxiliary variable $y(\underline{w}(j),\underline{w}*(j+1))$.

Therefore, define

$$\tilde{Q}(\underline{w},\sigma) \triangleq \int p(\underline{u}/\underline{w},\sigma).Q(\underline{u}).d\underline{u} \quad (2)$$

where $\underline{u}$ is an n-dimensional random vector and $p(\underline{u}/\underline{w},\sigma)$ denotes a normal gauszian probability density function with mean $\underline{w}$ and covariance matrix $\sigma^2 I$. $\tilde{Q}(\underline{w},\sigma)$ is a smoothed form of $Q(\underline{w})$. The gradient of (2) with respect to $\underline{w}$ equals :

$$\begin{aligned} \nabla_{\underline{w}}\tilde{Q}(\underline{w},\sigma) &= \int \nabla_{\underline{w}} p(\underline{u}/\underline{w},\sigma).Q(\underline{u}).d\underline{u} \\ &= -\sigma^{-2}.\int(\underline{u}-\underline{w}).p(\underline{u}/\underline{w},\sigma).Q(\underline{u}).d\underline{u} \end{aligned} \quad (3)$$

(3) can, for each $\underline{w}$, be estimated by averaging the measurements $\underline{\rho}.Q(\underline{u}).\sigma^{-2}$ over all the realizations $\underline{u}$, that are generated according to $p(\underline{u}/\underline{w},\sigma)$ and when $\underline{\rho}$ denotes $(\underline{u}-\underline{w})$. (3) can be estimated too by averaging the expressions

$$\sigma^{-2}.\underline{\rho}.(Q(\underline{u})-Q(\underline{w})) \quad (4)$$

since the average over all $\underline{\rho}$ of $\sigma^{-2}.\underline{\rho}.Q(\underline{w})$ equals zero. Assume that $\underline{b}(j)$ continuously tracks the negative gradient. The computation may be performed by (5) :

$$\underline{b}(j+1) = \underline{b}(j) + \frac{1}{\tau}.\left\{\sigma^{-2}(j).\underline{\rho}(j).\left[Q(\underline{w}(j))-Q(\underline{w}(j+1))\right] - \underline{b}(j)\right\} \quad (5)$$

An exponential filter is proposed with adaptation time $\tau>1$ instead of a simple "averaging" filter since the problem is nonstationary : indeed, on the one hand, the basepoint continuously moves. Clearly, the gradient continuously changes too ! On the other hand, the parameter $\sigma(j)$ is often variable too.

This method is only useful, therefore, when the statistics of $\underline{\rho}(j)$ are nearly stationary, when the base point moves slowly, and when the sequence $Q(\underline{w}(j))-Q(\underline{w}*(j+1))$ behaves well. Thus, highly nonlinear functions $Q(\underline{w})$, staircase functions and functions with many imposed inequality constraints cannot be treated.

Another "preferable direction estimator" is defined

$$\begin{aligned} \underline{b}(j+1) &= \underline{b}(j) + \frac{1}{\tau}.(\underline{\rho}(j)-\underline{b}(j)) \\ &\qquad \text{if } y(\underline{w}(j),\underline{w}*(j+1))=1 \\ \underline{b}(j+1) &= \underline{b}(j) + \frac{1}{\tau}.(-h.\underline{\rho}(j)-\underline{b}(j)) \\ &\qquad \text{otherwise} \end{aligned} \quad (6)$$

with $1> h>0$ and $\tau> 1$; $\underline{b}(j+1)$ is a weighted average of all the vectors $\underline{\rho}(1),\ldots,\underline{\rho}(j)$. The two main advantages of (6) over (5) are that $\underline{b}(j+1)$ doesn't depend upon the size of the gradient and that a broader class of functions $Q(\underline{w})$ can be allowed.

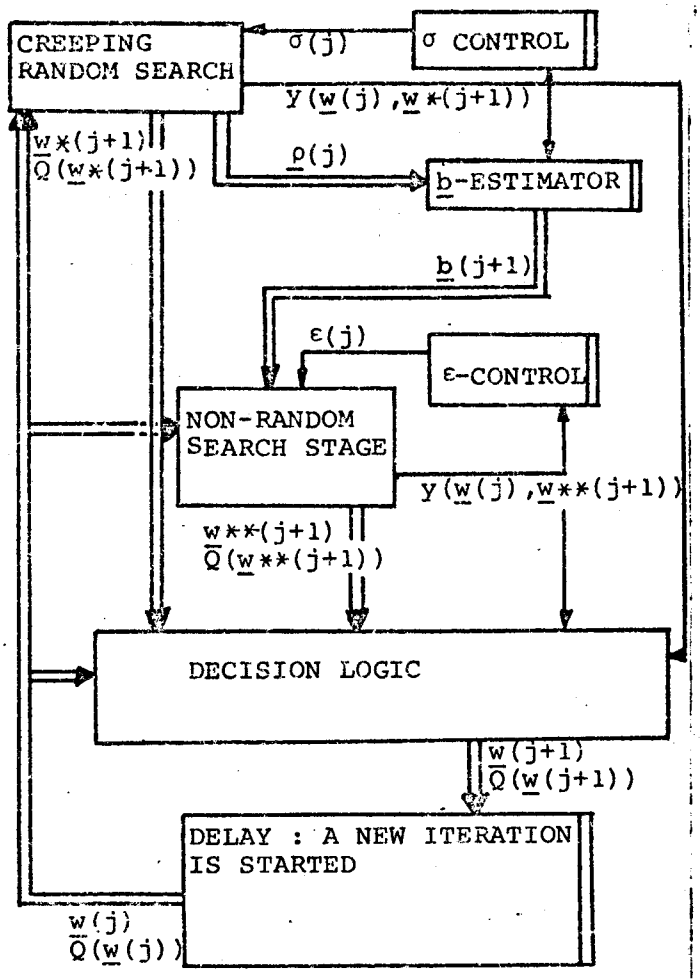

FIG.1

Let us now investigate the σ(j)-controller. Obviously, there may be no interference or correlation between the information gathering process (w*(j+1)) and the information handling process (5-6). This condition of independence is best satisfied when the control of the statistics of ρ(j) is performed only on the base of the outcomes y(w(j),w*(j+1)). No feedback from the stage, in which w**(j+1) is generated and Q(w**(j+1)) is computed, is allowed. Further, the statistics of ρ(j) are best only slowly varying.
One scheme is proposed here. It is discussed in detail in [7]:

$$\sigma(j+1) = \sigma(j).(1+\alpha) \quad \text{if } y(w(j),w^*(j+1))=1$$
$$\sigma(j+1) = \sigma(j).(1-\beta) \quad \text{otherwise} \qquad (7)$$

with : $\alpha > 0$; $1>\beta>0$; $\sigma(0)$=arbitrary ;

$$p_g \triangleq \beta/(\alpha+\beta) < 0,5 \quad \text{or} \quad \alpha > \beta \ ;$$

It is shown in [12] that, if W is closed and bounded, and under general conditions on Q(w) , σ(j) tracks a nonzero positive finite value $\sigma_o(w(j))$ which satisfies :

$$\int_{u \in W} p(u/w(j),\sigma).y(w(j),u).du = p_g \qquad (8)$$

Unfortunately, the root of (8) is nonstationary since w(j) continuously moves. Nevertheless, a stable nonnegative finite root always exists if $\alpha>0$, $1>\beta>0$, and $p_g<0.5$.

Thus, 100 p_g denotes the expected percent of times that w*(j+1) is better than w(j). By increasing p_g , the value of $\sigma_o(w(j))$, will, generally spoken, decrease: p_g is often chosen between 0,15 and 0,35. It may be compared with "the probability of a successful trial", which was first used in [5].

The condition of independence implies that α and β have to be small. Notice further that most of the search parameters become adaptive : σ(j), ε(j) and b(j) . The search becomes more "self-organizing". No a priori knowledge about Q(w) is any longer required except for the localization and the boundedness of W. The constants h(6), α(7), β(7), τ(5-6) have no such important influence on the performance of the algorithm. Experimental results have shown that there exists some optimal compromise between h and p_g : a small value h necessitates a great value p_g , since, roughly spoken, the informational amount, received by the estimators (5) or (6) increases for increasing h and decreases for increasing p_g. A value of h near 0,1 .. 0,3 is often successful when p_g varies from 0,35 to 0,15 . The constant τ, on the other hand, is best chosen larger when n increases or when Q(w) isn't a well-behaved function.

Let us investigate ε(j), defined in:

$$w^{**}(j+1) = w(j) + \varepsilon(j).b(j+1) \qquad (9)$$

The coefficient ε(j) perhaps has the greatest influence on the rate of convergence. A similar adaptation scheme is used as in (7) :

$$\varepsilon(j+1) = \varepsilon(j).(1+\eta) \quad \text{if } y(w(j),w^{**}(j+1))=1$$
$$\varepsilon(j+1) = \varepsilon(j).(1-\theta) \quad \text{otherwise} \qquad (10)$$

where

$\eta > 0$; $1>\theta>0$; $\varepsilon(0)$=arbitrary ;

The values η and θ may be greater than α and β since they determine the speed of the adaptation of $\varepsilon(j)$ to new situations. Typical values are 1 and 0,4 .It can further be verified that $\varepsilon(j)$ is a measure of the efficiency of the non-random part in the search. Indeed, $\| \underline{b}(j+1) \|^2$ is, on the average, proportional to $n.\sigma(j)^2$ when the statistics of $\underline{\rho}(j)$ are almost stationary. The displacement

$$\| \underline{w}^{**}(j+1) - \underline{w}(j) \|^2$$

is then, on the average, proportional to $n.\varepsilon(j)^2.\sigma(j)^2$. On the contrary, the displacement

$$\| \underline{w}^{*}(j+1) - \underline{w}(j) \|^2$$

grows as $n.\sigma(j)^2$. When $\varepsilon(j)$ is much greater than one, the displacement of the basepoint, caused by (9) is by far more important than the displacement obtained by means of the first trial $\underline{w}^*(j+1)$. Consequently, a large value of $\varepsilon(j)$ often corresponds to the existence of long straight valleys or ridges in that region of W where $\underline{w}(j)$ belongs to.

Let us conclude that the random stage, e.g. the first stage, assures the convergence for the global algorithm since only the best trial of both $\underline{w}^*(j+1)$ and $\underline{w}^{**}(j+1)$ is taken into account as a candidate for the new "best estimate" $\underline{w}(j+1)$. When accidently $\underline{w}^{**}(j+1)$ is always worse than $\underline{w}^*(j+1)$, the performance of the algorithm is comparable with that of creeping random search. The second -non-random- stage, on the contrary, determines the rate of convergence. The information, gathered in the first stage is efficiently used in order to speed up the rate of convergence.

3. COMPARISON WITH OTHER ALGORITHMS

Several authors tried yet to combine random search and non-random direct search in a single algorithm :

Schumer and Steiglitz [5] made a second trial in each iteration in $\underline{w}^{**}(j+1)$ which was defined as

$$\underline{w}^{**}(j+1) = \underline{w}(j) + (1+a).\underline{\rho}(j)$$

with $a > 0$.
This reduces to the CRSA when :

$$\varepsilon(j) = 1+a$$
$$\underline{b}(j+1) = \underline{\rho}(j)$$

However, the statistics of $\underline{\rho}(j)$ and the control of these statistics were somewhat different in their, so called, adaptive step size random search : ASSRS .

Gucker [13] in his compound stochastic gradient algorithm, defined the quantities $\varepsilon(j)$ and $\underline{b}(j+1)$ as :

$$\varepsilon(j) = \lambda(j)$$
$$\underline{b}(j+1) = \sigma(j)^{-2}.(Q(\underline{w}(j)) - Q(\underline{w}^*(j+1))).\underline{\rho}(j)$$

which reduces to (5) when $\tau=1$ .Although he uses the same statistics for $\underline{\rho}(j)$, his control section differs from the one, proposed in this paper. The experiments have shown that his algorithm is quite useful in accurate search problems.

4. MODIFICATIONS OF THE BASIC SEARCH SCHEME

The principal modification concerns the use of more function calls per iteration. When the function calls are not too costly, some additional evaluations could be spent in the information gathering stage. However, the most promising results were obtained when the additional calls were made in the non-random stage. Many trials $\underline{w}_i^{**}(j+1)$, $i=1,..K$ could be made on a straight line through $\underline{w}(j)$ in the direction $\underline{b}(j+1)$ .The non-random part reduces to a linear search. A proper control of the search parameters is necessary [12] in order to minimize the number of function calls in each iteration without slowing down the rate of convergence.

Random search is extremely useful in multimodal optimization. By adding one supplementary device, the CRSA can get global search properties. Several schemes were developed in [12] and [15] which work fairly well in the presence of several minima. Unfortunately, this goes beyound the size of this paper.

In some problems, the initialization of $\sigma(j)$ becomes cumbersome. Indeed, the CRSA is senseless when $\sigma(j)$ is mismatched: the random stage will fail when $\sigma(j)$ is by far too great ; the estimators (5) and (6) will deteriorate and the non-random stage is clearly superfluous. Therefore, a fast initialization of $\sigma(j)$ is desired, e.g. an auxiliary search is performed in the beginning of the optimization without spending many function calls. The initialization of $\varepsilon(j)$ and $\underline{b}(j)$, on the contrary, is quite simple :

$$\varepsilon(0) = 1$$
$$\underline{b}(0) = \underline{0}$$

Finally, when ridges or narrow valleys can be expected, an acceleration technique could be used, which is similar to an iterated partan method. After a fixed number of iterations and periodically an "acceleration-iteration" is introduced, just as in the pat tern search (Hooke and Jeeves ; Rosenbrock ; Powell ; Zangwill) .It has to be pointed out, however, that a too large basepoint move sabotages the CRSA since the degree of non-stationarity of the parameters to be estimated, increases.

5. EXPERIMENTS

The results of the experiments are averaged over many runs and many values for the search parameters h, α, β, η , and θ .Most of the results of other authors concern only the best set of search parameters. The reader is in that case completely ignorant about the sensitivity of this algorithms to changes in search parameters. This incited us to av-

erage the results over a wide range of values for the search parameters. The experimental load clearly increases, but the comparison between several techniques becomes more senseful [12]. The following direct search techniques were compared :

1°) Adaptive creeping random search (CREEP). The first stage (1) of the CRSA is tested using a similar σ-control section as in (7). The comparison with the CRSA, when a non-random second stage is added, will be interesting.

2°) The method of Matyas [4], which is completely random too. It makes use, however, of a preferable direction.

3°) The adaptive creeping random search with an acceleration device of the type "pattern search" (CREEP-AC). The algorithm is described in [12].

4°) The CRSA. The performance of this algorithm with or without linear search system is almost the same.

5°) The compound stochastic gradient algorithm of Gucker (CSGA : [13]).

6°) Pattern search of Hooke and Jeeves [10].

7°) The algorithm of Rosenbrock [11]

These techniques were compared on the base of several functions $Q(\underline{w})$. The results are given in Table 1. The tests are now briefly described :

1°) The function of Rosenbrock [11] is chosen as first test function. The starting point was always (-1,2 ; +1,0). The symbols AC1 and AC2 in Table 1 denote the mean value of $Q(\underline{w}(j))$ over a lot of runs after 600 and 2000 function calls respectively. A small value AC1 or AC2 indicates that the algorithm is capable of localizing an optimum quite accurately with almost no effort.

2°) Under CS (coarse search) in Table 1, the mean number of function calls is given that was needed for the rough localization of the minimum of a six-dimensional convex and well-behaved test function.

3°) Under CO (optimization in the presence of constraints) in Table 1, the mean number of function calls is given that was needed to get $Q(\underline{w}(j))$ less than a certain threshold value Q_0. The test function was quite simple and n was 12. The constraints are highly nonlinear and are not a priori known, e.g. they define a region in W, in which $Q(\underline{w})$ isn't measurable. Each trial in this region is regarded as a function call.

4°) The same test function and the same criterion as under CO were used in the test "PS" (see Table 1). The a priori unknown constraints, however, are tighter and define a narrow nonlinear pipeline in which the gradient of $Q(\underline{w})$ is almost perpendicular to the boundaries of the forbidden region for each $\underline{w}$. The optimization is quite difficult.

5°) Multimodal performance (MM) : + means that a global minimum will be found anyway ; 0 denotes that this global minimum of the multimodal test function will only be reached when the variance in the random part is bounded from below ; - means that the convergence to a global minimum isn't guaranteed at all.

From table 1, we see that the method of Matyas performs better in all the situations than CREEP since more information is used during the search, although the general structure of the algorithms was the same. The accelerated creeping random search per- only performs better than the simple CREEP when narrow ridges or valleys are met, as near the optimum of the Rosenbrock-function (AC2) or in the pipeline search (PS).

The partially non-random methods (CRSA, method of Gucker) are superior in "accurate search" problems (AC1, AC2) with respect to the completely random techniques (CREEP, Matyas).

Furthermore, the CRSA needs much less function evaluations in the problem of Rosenbrock than the powerful methods of Rosenbrock, Gucker or the pattern search of Hooke and Jeeves.

When the expected number of function calls, that are needed to solve the problem, is not quite large, as in coarse search (CS), the CRSA obviously is inferior in comparison with the other techniques. Indeed, this method needsa long adaptation time before an appropriate σ(j) and a useful $\underline{b}(j)$ are computed.

AC1	AC2	MM	
$0,34.10^{0}$	$0,67.10^{-2}$	0	CREEP
$0,54.10^{-1}$	$0,17.10^{-2}$	0	Matyas
$0,11.10^{-1}$	$0,47.10^{-6}$	0	CREEP-AC
$0,20.10^{-5}$	$0,10.10^{-9}$	0	CRSA
$0,82.10^{-2}$	$0,30.10^{-5}$	+	CSGA-Gucker
$0,10.10^{-1}$		-	Hooke-Jeeves
	$0,22.10^{-4}$	-	Rosenbrock

Table 1, part A

CS	CO	PS	
40,1	182,6	401,4	CREEP
37,8	154,6	249,3	Matyas
40,1	192,9	158,4	CREEP-AC
105,3	152,9	237,9	CRSA

Table 1,part B

CONCLUSION

The problem was to design a suitable method which deals with the optimization of a unimodal function Q(w) on the condition that the function evaluations are very costly. This function may contain anomalities.It can be regarded as an unknown environment in which the basepoint tries to find its optimal value.The convergence to the minimum must be assured,the sought method must be rather insensitive to the search parameter choice and needs only a few function calls to reach the purpose.
It was quite normal to use a random search technique.But until now,these techniques weren't able to localize an optimum quite accurately.The features of non-random search(its accuracy) and random search are combined in a new algorithm : the compound random search algorithm (CRSA).
During the search,information about Q(w) is gathered in order to accelerate and control the search.
The experimental results confirmed our ideas.A comparison with other random search techniques and direct search methods is made.The CRSA turns out to be superior in a large class of problems.Finally,some suggestions are made about the use of the CRSA in special problems.Some auxiliary devices may be added for this purpose.

BIBLIOGRAPHY

[1] S.H.BROOKS "A discussion of random methods for seeking maxima.",Oper.Res.6, No.2,1958

[2] S.H.BROOKS "A comparison of maximum seeking methods.",Oper.Res.7,No.4,1959

[3] L.A.RASTRIGIN "The convergence of the random search method in the extremal control of a many-parameter system.", Aut.and Rem.Control,vol.24,No.11,1963

[4] J.MATYAS "Random optimization.",Aut.and Rem.Control,vol.26,No.2,1965

[5] M.A.SCHUMER,K.STEIGLITZ "Adaptive step size random search.",IEEE,Trans.on Aut. Control,vol.AC-13,No.3,1968

[6] L.A.RASTRIGIN "Random search in optimization problems for multiparameter systems.",Air Force Systems Command,Foreign Technology Division,August 1967

[7] G.T.HEYDT "Random search using hyperconical search regions.",Ph.D.Thesis Proposal,Purdue University,February 1969

[8] M.A.SCHUMER "Optimization by adaptive random search.",Ph.D.Dissertation;,Princeton University,Nov.1967

[9] L.J.WHITE,R.G.DAY "An evaluation of adaptive step size random search.",IEEE, Trans.on Aut.Control,vol.AC-16,No.5,1971

[10] R.HOOKE,T.A.JEEVES "Direct search solution of numerical statistical problems." Journal of the Assoc.of Comp.Machinery, vol.8,No.2,1961

[11] H.H.ROSENBROCK "An automatic method for finding the greatest or least value of a function.",Comp.Journal,vol.3,1965

[12] L.P.DEVROYE "The compound random search algorithm.",Techn.Rept.KES-D-0272,Catholic Univ.of Louvain,Belgium,April 1972

[13] G.R.GUCKER "Stochastic gradient algorithms for searching multidimensional multimodal surfaces.",Techn.Rept.TR-6778-7, Stanford University,1969

[14] G.J.McMURTRY "Adaptive optimization procedures.",in : "Adaptive,learning and pattern recognition systems.",J.M.Mendel and K.S.Fu Ed.,Ac.Press,N.Y.,1970

[15] L.P.DEVROYE "MARS:Multimodal adaptive random search.",Techn.Rept.KES-D-0372, Catholic University of Louvain,Belgium, May 1972