Mathematical Programming 10 (1976) 230–244.
North-Holland Publishing Company

OPTIMIZED RELATIVE STEP SIZE RANDOM SEARCHES

Günther SCHRACK* and Mark CHOIT
University of British Columbia, Vancouver, Canada

Received 29 June 1974
Revised manuscript received 8 July 1975

New results concerning the family of random searches as proposed by Rastrigin are presented. In particular, the random search with reversals and two optimized relative step size random searches are investigated. Random searches with reversals are found to be substantially better than their counterparts. A new principle of updating the step size for this family of searches is proposed.

1. Introduction

Algorithms for the numerical optimization of unconstrained objective functions may be categorized into four classes:

(i) Second-order algorithms or Newton-Raphson methods,
(ii) first-order algorithms or gradient searches,
(iii) zero-order algorithms or direct searches,
(iv) Monte Carlo algorithms or random searches.

Although, strictly speaking, random searches are direct searches, it is useful to separate them in order to contrast them with the other three classes. Classes (i) to (iii) are deterministic algorithms, whereas random searches use stochastic methods to optimize a (deterministic) objective function.

The concept of random search seems to have been proposed first by Anderson [1]. Since then, a large variety of algorithms of this class have become known [9].

This paper deals with the specific family of random searches proposed initially in 1960 by Rastrigin [5]. Rastrigin published theoretical investigations in 1963 [6] and in 1964 with Mutseniyeks [4]. Later, Schumer

* The University of British Columbia, Department of Electrical Engineering, Vancouver, B.C., V6T 1W5, Canada.

and Steiglitz [8] extended these ideas and proposed the first practical implementation of this type of search.

After a brief introduction to the concepts involved, new recursion formulas will be stated for some of the parameters of interest, to be followed by a theoretical investigation of an improved algorithm which belongs to this family. A new proposal to handle the step size updating problem will conclude the paper.

2. The step size and direction problems

Let g be the (real valued) objective function, being a function of N parameters, $g = g(x)$, $x \in \mathbf{R}^N$. Desired is a local minimum $x = x^*$, i.e., a vector x^* such that $g(x^*) \leqslant g(x)$ for all x in a suitable environment of x. Let x^0 be a given starting point. Sequential optimization algorithms are often based on the formula

$$x^{k+1} = x^k + \sigma^k d^k, \qquad k = 0, 1, ..., \tag{1}$$

where x and d are N-vectors, and σ is a scalar. Supercripts denote iteration. For convenience assume d^k, $k = 0,1, ...$ to be normalized vectors, i.e. the Euclidean norm $\|d^k\| = 1$ for all k.

Two basic problems need to be attacked in the definition of such an optimization algorithm: the *direction problem* and the *step size problem.* These are the problems of choosing the vector d^k and the scalar σ^k, representing the direction and the step size, respectively, of the next step of the search. Desired are values which will insure a fast convergence of the sequence $\{x^0, x^1, ...\}$ to x^*.

In most random searches, the direction d^k is determined by the use of random vectors. The searches then differ in the choice of certain strategies, e.g. in the choice of the distribution from which the random vectors are to be selected, in the possibility of including past information, or in the choice of the step size.

3. Rastrigin's family of random searches

To allow a theoretical analysis of a given algorithm, Rastrigin [6] approximates the arbitrary objective function g by the N-dimensional function $f(x) = x^{\mathrm{T}} x$ (superscript T denoting transposition) whose contours are hyperspherical.

He proposed a sequential random search, based on equation (1), for which unit random vectors, with a uniform distribution on the hypersphere, are chosen as search directions. Essentially, this eliminates the direction problem, and the theoretical analysis thus concentrates on the step size problem. These conditions characterize the family of random searches under discussion.

A number of proposals to handle the step size problem have been made so far: In the *fixed step size random search* [6], the step size of the normalized random vectors is never changed. In the *optimum step size random search* [4], the step size taken is the optimum step size along the random direction vector. In the *adaptive step size random search* [8], the step size is increased or decreased in a heuristic fashion according to the successes or failures experienced during the search. Other random searches based on equation (1) have been proposed as well, see e.g. [10].

For searches of this family, the probability of success $P = \mathbf{P}(f(x^{k+1}) < f(x^k))$ is derived from the ratio S_1/S_T of surfaces on the hypersphere, where S_1 is the surface of the cap subtended by the angle for which a success can still occur and S_T is the total surface of the hypersphere [6, 4]. It is given by

$$P = \frac{1}{2\,A(N)} \int_0^{\psi_m} \sin^{N-2}\psi \, \mathrm{d}\psi \,, \tag{2}$$

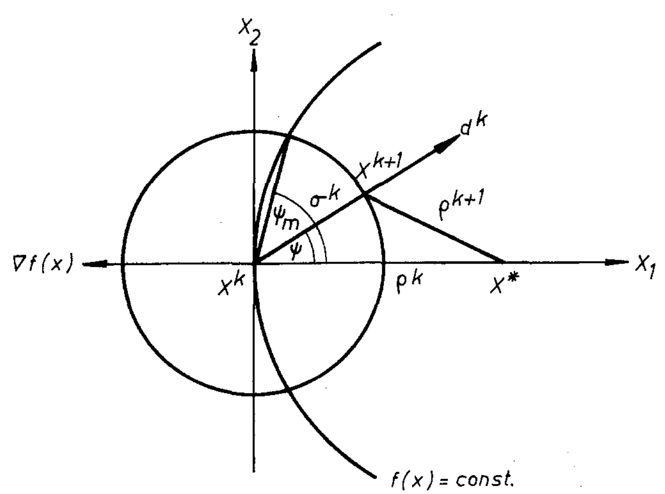

Fig. 1. Mathematical notation, $N = 2$.

where $A(N) := \int_0^{\pi/2} \sin^{N-2}\psi \, d\psi$, ψ is the angle between the random vector d and the negative gradient of f, and ψ_m is the maximum angle ψ for which a success can still occur (see Fig. 1). This angle is given by $\psi_m^k = \arccos(\sigma^k/(2\rho^k))$, where $\rho^k := \|x^k - x^*\|$ is the distance to the optimum.

Thus, the probability of success is a function of the parameters N, σ and ρ.

By introducing [8] the concept of *relative step size* $\eta := \sigma/\rho$, defined as the ratio of the step size σ to the *distance to the optimum* ρ, the probability of success P becomes a function of two parameters only, the dimension N and the relative step size η. The angle ψ_m^k becomes

$$\psi_m^k(\eta) = \arccos(\eta^k/2), \quad \text{where } \eta^k := \sigma^k/\rho^k .$$

For a success, the relative step size must observe the inequalities

$$0 < \eta^k < 2 .$$

Note, as $\eta \to 0$, $P(N, \eta) \to 0.5$.

Theorem. *The integrals $A(N)$ and $P(N, \eta)$ can both be expressed by recursion relations as follows*:

$$A(2) = \pi/2 , \qquad A(3) = 1 ,$$

$$A(N) = (N-3)\, A(N-2)/(N-2) , \qquad N \geqslant 4 \tag{3}$$

and

$$P(2,\eta) = (1/\pi) \arccos(\eta/2) ,$$

$$P(3,\eta) = (2-\eta)/4 ,$$

$$P(N,\eta) = P(N-2,\eta) - \frac{\eta}{4(N-2)\, A(N)} (1 - \eta^2/4)^{(N-3)/2},$$

$$N \geqslant 4 . \tag{4}$$

The proofs for these relations, being straightforward, will be omitted.

Since a random search employs stochastic methods for the optimization of an objective function, it becomes necessary to investigate the search with probabilistic or statistical means. Thus, its behaviour must be evaluated in terms of the *mean* of a number of optimization runs, each attempt optimizing the same objective function and starting with

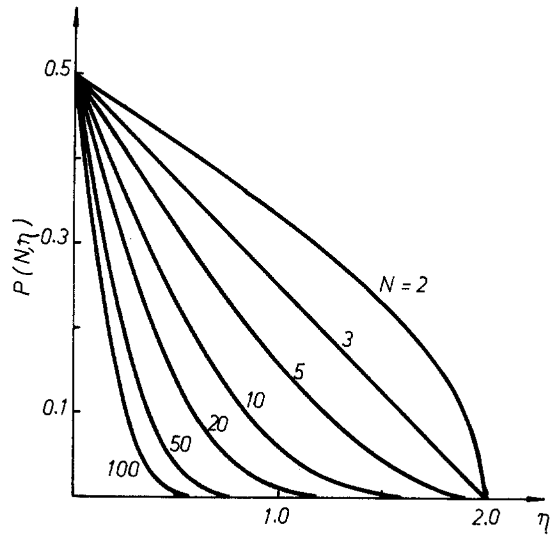

Fig. 2. Random search without reversals, probability of success vs. η, N = 2, 3, 5, 10, 20, 50, 100.

the same starting point, but using different sequences of random vectors.

In order to meet this objective as well as to allow a comparison of one algorithm with another on this basis, Rastrigin introduced an interesting concept, the *search loss.* Its reciprocal, to be called the *expected relative improvement per function evaluation* $I(N, \eta)$, is a more relevant parameter here. Let $E(\cdot)$ denote mathematical expectation. Then, for a minimization algorithm, $I(N, \eta)$ is defined by

$$I(N, \eta) := \mathrm{E}((f(x^k) - f(x^{k+1}))/f(x^k)) , \qquad k \text{ fixed.}$$

This parameter is related to the usual criterion used for the comparison of minimization algorithms, viz. the number of function evaluations required for a given reduction of the objective function value. The larger the expected relative improvement, the better the search.

For the family of random searches under discussion, it has been shown [8] that

$$I(N, \eta) = 1/(2A(N)) \int_0^{\psi_m} (2\eta \cos \psi - \eta^2) \sin^{N-2} \psi \, \mathrm{d}\psi$$

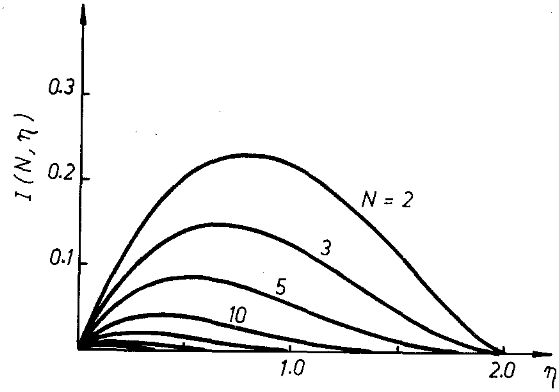

Fig. 3. Random search without reversals, expected relative improvement vs. η, N = 2, 3, 5, 10, 20, 50, 100.

from which, by integration, for $N \geqslant 2$,

$$I(N,\eta) = \begin{cases} \eta\,(1-\eta^2/4)^{(N-1)/2}/((N-1)\,A(N)) - \eta^2\,P(N,\eta)\,, \\ \quad 0 \leqslant \eta \leqslant 2\,, \\ 0 \quad \text{otherwise}\,. \end{cases} \tag{5}$$

The relations (3), (4) and (5) allow the numerical calculation of the probability of success $P(N, \eta)$ and the expected relative improvement $I(N, \eta)$ for any given η in the range $0 \leqslant \eta \leqslant 2$ and any given dimension N. Figs. 2 and 3 depict graphs of these parameters versus η for several N.

From Fig. 3 it becomes immediately obvious that there exists an optimum relative step size $\eta^*(N)$ for each dimension. Numerical values for η^* can be obtained by solving the equation

$$\mathrm{d}I(N, \eta^*)/\mathrm{d}\eta = 0\,.$$

Table 1
Optimum relative step size as a function of dimension N

N	$\eta^*(N)$	$\eta_r^*(N)$
2	0.7885	0.7490
3	0.666...	0.6235
5	0.5298	0.4897
10	0.3812	0.3494
20	0.2717	0.2480
50	0.1726	0.1572
100	0.1222	0.1112

Some of these values are shown in Table 1.
As N increases, the values for $P(N, \eta^*)$, $I(N, \eta^*)$, and $\eta^*(N)$ tend to the asymptotic expressions given in [8].

4. The optimized relative step size random search with reversals

Before entering into a discussion of the step size problem, an obvious extension of Rastrigin's random search will be investigated. This random search, called *random search with reversals,* is theoretically superior to the above random searches, as will be demonstrated in section 5. It was originally proposed by Lawrence and Steiglitz [3].

Let $\bar{x}$ be the parameter vector with associated previous best function value, i.e. $f(\bar{x}) \leqslant f(x^i)$, $i = 0, 1, ..., k$. Upon generation of a random vector and the calculation of the next step $x^{k+1} = \bar{x} + \sigma^k d^k$ and its associated function value $f(x^{k+1})$, the subsequent success or failure of the iteration will be called a *first success* or *first failure,* respectively. Upon a first failure, the direction of search is reversed, i.e. x^{k+2} becomes

$$x^{k+2} = \bar{x} - \sigma^k d^k \tag{6}$$

and the objective function is again evaluated. The success or failure following a first failure will be called a *second success* or *second failure*, respectively.

Theorem. *For the random search with reversals, the probability of succes is given by*

$$P_r(N, \eta) = 2\, P(N, \eta)/(2 - P(N, \eta))\,, \tag{7}$$

where $P(N,\eta)$ is the probability of success of the random search without reversals (eq. (2)).

Proof. Let n be the number of random vectors generated, let s_1 and s_2 be the number of first and second successes. To calculate the probability of success, examine the limit of the ratio of the total number of successes to the number of function evaluations as $n \to \infty$. Since the number of reversals is given by $n - s_1$, this becomes

$$P_r(N, \eta) = \lim_{n \to \infty} \frac{s_1 + s_2}{n + (n - s_1)}\ . \tag{8}$$

Define

$f_1 := s_1/n$, the frequency of first success,

$f_2 := s_2/(n - s_1)$, the frequency of second success.

The limit of f_1 as $n \to \infty$ is identical to the probability of success $P(N,\eta)$. The limit of f_2 is the conditional probability P_2 of a second success given that a first failure has occurred. Similar to the derivation of $P(N,\eta)$, probability P_2 is derived from the ratio S_2/S_R of surfaces on the hypersphere where S_2 is the surface of the cap subtended by the angle for which a second success can still occur and $S_R = S_T - S_1$ is the surface of the cap subtended by the angle on which a first failure occurs. Due to the symmetry of the objective function considered, i.e. since the contours of $f(x)$ are hyperspherical, the angle subtending the cap with surface S_2 is equal to the angle subtending the cap with surface S_1 and hence $S_2 = S_1$. Thus, the ratio S_2/S_R becomes $S_1/(S_T - S_1)$, from which $P_2 = P(N,\eta)/(1 - P(N,\eta))$. Of course, in general probabilistic situations, the probability of a second success given a first failure is not $P/(1 - P)$. Therefore,

$$\lim_{n\to\infty} f_2 = P(N,\eta)/(1 - P(N,\eta)) .$$

Substitution of the limits of f_1 and f_2 into (8) yields the desired probability of success.

Note, as $n \to 0$, $P_r(N,\eta) \to 2/3$.

Theorem. *For the random search with reversals, the expected relative improvement per function evaluation is given by*

$$I_r(N,\eta) = 2\, I(N,\eta)/(2 - P(N,\eta)) , \tag{9}$$

where $I(N,\eta)$ *is the expected relative improvement and* $P(N,\eta)$ *is the probability of success of the random search without reversals* (*equations* (5) *and* (2) *respectively*).

Proof. The expected relative improvements per *successful* function evaluation are identically equal for the two searches, i.e.

$$I(N,\eta)/P(N,\eta) = I_r(N,\eta)/P_r(N,\eta) .$$

From this, the theorem follows immediately.

Graphs of the parameters $P_r(N,\eta)$ and $I_r(N,\eta)$ are similar to those in Figs. 2 and 3. Again, optimum relative step sizes η_r^* exist for each N, numerical values for which (see Table 1) can be obtained by solving the equation

$$dI_r(N,\eta_r^*)/d\eta = 0 .$$

5. Comparison of the two types of algorithms

The probability of success of the random search with reversals can easily be compared to that of the search without reversals for fixed N and η:

Theorem. *For N and η fixed,*

$$P_r(N,\eta) > P(N,\eta) , \qquad N \geqslant 2 , \quad 0 \leqslant \eta < 2 . \tag{10}$$

Proof. For $0 \leqslant \eta < 2$ there follows $P^2 > 0$, whence

$$P_r - P = 2P/(2-P) - P = P^2/(2-P) > 0 .$$

Similarly, the expected relative improvements per function evaluation for the two searches can be compared.

Theorem. *Let $I_r(N,\eta)$ be the expected relative improvement per function evaluation for the random search with reversals.*

For N and η fixed,

$$I_r(N,\eta) > I(N,\eta) , \qquad N \geqslant 2 , \quad 0 < \eta < 2 . \tag{11}$$

Proof. The proof is similar to the preceding one.

Numerically, the probability of success for a high dimension (say $N > 100$), at the optimum relative step size is approximately 16% higher for the search with reversals compared to the search without,

$$P_r(N,\eta_r^*) \simeq 1.16\, P(N,\eta^*) .$$

Similarly,

$$I_r(N,\eta_r^*) \simeq 1.25\, I(N,\eta^*)\,,$$

i.e. approximately a 25% improvement.

Remark. Lawrence and Emad [2] have compared the random search with reversals with that without reversals and with the gradient search. They discuss a parameter called the search loss, $\mathrm{SL}(x)$, which is easily shown to be the inverse of the relative improvement per function evaluation: $\mathrm{SL}(x) = 1/I(N,\eta)$. They relate the search losses for the two random searches by $\mathrm{SL}_r(x)/\mathrm{SL}(x) = \frac{2}{3}$. From equation (9), however, the inequalities

$$1 \leqslant I_r(N,\eta)/I(N,\eta) < \tfrac{4}{3}$$

can be derived. The result concerning the ratio $\mathrm{SL}_r(x)/\mathrm{SL}(x)$ thus stands corrected.

6. Updating the step size

The results given so far are valid for all members of Rastrigin's family of random searches, i.e. they are independent of the strategy employed to solve the step size problem. Surely, the fixed step size random search as well as the optimum step size random search are unsatisfactory since the former will not progress fast enough in the initial and final iterations of the search and the latter requires too many function evaluations at each iteration during the one-dimensional search along the search direction. The existence of an optimum relative step size suggests the alternative approach to keep the relative step size at its optimum value at each iteration of the search, thus insuring maximum expected relative improvement in the mean. This search was introduced by Schumer and Steiglitz [8] and they named it the "optimum step size random search" (but which should not be confused with the search of the same name of Mutseniyeks and Rastrigin). Based on this search, they have defined their "adaptive step size random search", where the adapting of the step size occurs in a heuristic fashion.

The latter search as well as a number of other searches prescribe an increase in the step size after a success. But if the relative step size is to be kept constant (say, at its optimum value) throughout the search, i.e.

if

$$\eta^k = \eta^*(N) \quad \text{for all } k\,, \tag{12}$$

the step size σ must be decreased since the distance to the optimum decreases after a success as shown by $\sigma^k = \eta^*(N)\, \rho^k$. The difficulty involved in an implementation of the search is the fact that neither the distance to the optimum nor the relative step size is known at each iteration. In order to find a satisfactory solution to the step size problem, one proposal might be to update η^{k+1} in some as yet undetermined fashion after a success such that condition (12) is met. However, from the very nature of a random search, this will not be possible. Instead, it will be necessary to substitute for condition (12) an alternative one, namely to keep the relative step size constant in the mean; more precisely, to keep the mathematical expectation of η^{k+1} (given ρ^k, σ^k, σ^{k+1}, and N) constant by an appropriate choice of σ^{k+1}:

$$\mathrm{E}(\eta^{k+1} \mid \rho^k, \sigma^k, \sigma^{k+1}, N) = \eta^k \quad \text{for all } k\,. \tag{13}$$

Since $\eta^{k+1} = \sigma^{k+1}/\rho^{k+1}$ with σ^{k+1} a constant,

$$\mathrm{E}(\eta^{k+1} \mid \rho^k, \sigma^k, \sigma^{k+1}, N) = \sigma^{k+1}\, \mathrm{E}(1/\rho^{k+1} \mid \rho^k, \sigma^k, N)\,, \tag{14}$$

indicating that it is possible to employ the updating formula

$$\sigma^{k+1} = \alpha\, \sigma^k \quad \text{for all } k = 0, 1, \ldots, \tag{15}$$

where α is a scalar, to be called the *step size updating factor.* This choice of σ^{k+1} happens to have a desirable effect: the substitution of (15) into the explicit expression of $\mathrm{E}(\eta^{k+1} \mid \rho^k, \sigma^k, \sigma^{k+1}, N)$ transforms it such that it becomes a function of η^k, α and N, i.e. it no longer is dependent on ρ^k and σ^k individually.

Equation (13) then becomes

$$\mathrm{E}(\eta^{k+1} \mid \eta^k, \alpha, N) = \eta^k \quad \text{for all } k = 0, 1, \ldots\,. \tag{16}$$

Note that this equation implies that the step size updating factor α is also a function of η^k and N only, $\alpha = \alpha(\eta^k, N)$.

7. Calculation of the step size updating factor

Although it is possible to calculate the step size updating factor by equation (16), it is more convenient to factor out σ^{k+1} as in (14). The scalar α can then be calculated explicitly:

$$\alpha = 1/(\rho^k\, \mathrm{E}(1/\rho^{k+1} \mid \rho^k, \sigma^k, N))\,. \tag{17}$$

However, this expression requires the conditional expectation of the inverse distance to the optimum for given ρ^k, σ^k, and N.

Lemma. *Let* $y = 1/\rho^{k+1}$, $\rho = \rho^k$, $\sigma = \sigma^k$. *Then the mathematical expectation of the inverse distance to the optimum, given* ρ, σ, *and* N, *is given by*

$$\mathrm{E}(y \mid \rho,\sigma,N) = (1/(2\rho\sigma A(N) P(N,\eta))) \int_{1/\rho}^{1/|\rho-\sigma|} ((1-\delta^2(y)/4)^{(N-3)/2}/y^2)\,\mathrm{d}y\ ,$$

where $\delta(y) = \eta + 1/\eta - 1/(\rho\sigma y^2)$.

Proof. The conditional density distribution of the distance to the optimum can be obtained from the conditional probability of $\rho^{k+1} < w$ for fixed $w < \rho^k$. Let $z = \rho^{k+1}$, ρ, σ as before. After some manipulations, there follows

$$p(z \mid \rho, \sigma, N) = z(1-\beta^2(z)/4)^{(N-3)/2}/(2\rho\sigma A(N) P(N,\eta))\ , \qquad (18)$$

where $\beta(z) = \eta + 1/\eta - z^2/(\rho\sigma)$.

The ranges of z are given by

$$z = \rho^{k+1} \in \begin{cases} [\rho^k(1-\eta^k), \rho^k] & \text{for } \eta^k \in [0,1]\ , \\ [\rho^k(\eta^k-1), \rho^k] & \text{for } \eta^k \in [1,2]\ . \end{cases}$$

The transformation of (18) for $y = 1/\rho^{k+1}$ yields the conditional probability density for the inverse of the distance to the optimum. The desired expectation then follows immediately.

A transformation of (17) using the substitution $t = \sigma y = \sigma^k/\rho^{k+1}$ shows that $\rho\mathrm{E}(y \mid \rho,\sigma,N)$ is not dependent on ρ^k and σ^k individually, i.e. $\alpha = \alpha(\eta^k, N)$, as asserted above.

8. The optimized relative step size random searches

The proposed new updating principle, represented by equation (16) with $\eta^k = \eta^*(N)$ and equation (15), defines searches called the *optimized relative step size random search* (ORSSRS) *and optimized relative step size random search with reversals.*

For the optimum relative step size $\eta^*(N)$, a few numerical values of α are given in Table 2, as calculated with aid of eq. (17). $\alpha^*(N)$ is the optimum step size updating factor for the ORSSRS without reversals and $\alpha_r^*(N)$ that for the same with reversals.

The ORSSRS with reversals can be summarized as follows.

Table 2
Optimum step size updating factor as a function of dimension N

N	$\alpha^*(N)$	$\alpha_f^*(N)$
2	0.4471	0.4801
3	0.666...	0.6883
5	0.8212	0.8330
10	0.9183	0.9238
20	0.9609	0.9636
50	0.9848	0.9858
100	0.9924	0.9930

Algorithm

(i) Set $\bar{x} = x^0$, $\bar{\sigma} = \sigma^0$, $\bar{g} = g(x^0)$, $k = -1$.

(ii) Update $k \leftarrow k + 1$, generate d^k, a random vector with uniform distribution on the hypersphere, and calculate x^{k+1} corresponding to equation (1), i.e. by

$$x^{k+1} = \bar{x} + \bar{\sigma} d^k \ .$$

(iii) Evaluate the objective function $g(x^{k+1})$ at the new point and compare with $\bar{g}$.

(iv) In the case of success apply stopping rules. If the search is to be continued, replace $\bar{x}$ by x^{k+1}, $\bar{g}$ by $g(x^{k+1})$, and reduce the step size according to equation (15), i.e. by

$$\bar{\sigma} \leftarrow \alpha \, \bar{\sigma} \ ,$$

then continue with (ii).

(v) In the case of no success, return to (ii) if the immediately preceding step was a reversed step, else reverse the search direction according to equation (6), i.e. using

$$x^{k+2} = \bar{x} - \bar{\sigma} d^k \ ,$$

update $k \leftarrow k + 1$, and continue with (iii).

As mentioned, a value for the initial step size σ^0 must be known, without which the algorithm cannot be started. Since it cannot be assumed that σ^0 is known, it will be necessary to estimate it. Furthermore, for the optimized relative step size searches σ^0 should be such that the initial relative step size assumes its optimum value, i.e. $\eta^0 = \eta^*(N)$

A number of possibilities present themselves for this purpose. One is to obtain σ^0 from a knowledge of ρ^0, then $\sigma^0 = \eta^*(N)\ \rho^0$. Another possibility is an experimental estimation of the probability of success $P_{\mathrm{r}}(N,\eta)$ in an initialization phase of the algorithm. With an estimated step size $\hat{\sigma}$, a number of trials m_1 are carried out at the starting point x^0 and the number of successes m_2 recorded. Then the ratio m_2/m_1 represents an approximation of $P_{\mathrm{r}}(N,\hat{\eta})$, and $\hat{\eta}$ is an approximation to the relative step size $\hat{\eta} = \hat{\sigma}/\rho^0$. The implicit calculation of η from the equation

$$m_2/m_1 = P_{\mathrm{r}}(N,\hat{\eta})$$

presents no difficulties; Newton–Raphson iterations converge always. Thus,

$$\rho^0 \simeq \hat{\sigma}/\hat{\eta}\ ,$$

but desired is $\sigma^0 = \sigma^*$ from $\rho^0 = \sigma^*/\eta^*(N)$, i.e.

$$\sigma^0 \simeq \hat{\sigma}\eta^*(N)/\hat{\eta}\ ,$$

the desired initial estimate of the step size.

An implementation of this algorithm requires some additional considerations. For example, it will be necessary to take into account the possibility of too large or too small an estimated step size $\hat{\sigma}$. Furthermore, for a nonparallel implementation, a periodic reestimation of the step size is advisable.

Experimental results of the kind reported previously [7] are encouraging, particularly for the search with reversals.

References

[1] R.L. Anderson, "Recent advances in finding best operating conditions", *Journal of the American Statistical Association* 48 (1953) 789–798.

[2] J.P. Lawrence III and F.P. Emad, "An analytic comparison of random searching and gradient searching for the extremum of a known objective function", *IEEE Transactions on Automatic Control* AC-18 (1973) 669–671.

[3] J.P. Lawrence III and K. Steiglitz, "Randomized pattern search", *IEEE Transactions on Computers* C-21 (1972) 382–385.

[4] V.A. Mutseniyeks and L.A. Rastrigin, "Extremal control of continuous multi-parameter systems by the method of random search", *Engineering Cybernetics* 1 (1964) 82–90.

[5] L.A. Rastrigin, "Extremal control by the method of random scanning", *Automation and Remote Control* 21 (1960) 891–896.

[6] L.A. Rastrigin, "The convergence of the random search method in the extremal control of a many-parameter system", *Automation and Remote Control* 24 (1963) 1337–1342.
[7] G. Schrack and N. Borowski, "An experimental comparison of three random searches", in: F.A. Lootsma, Ed., *Numerical methods for non-linear optimization* (Academic Press, London, 1972) pp. 137–147.
[8] M.A. Schumer and K. Steiglitz, "Adaptive step size random search", *IEEE Transactions on Automatic Control* AC-13 (1968) 270–276.
[9] R.C. White, Jr., "A survey of random methods for parameter optimization", *Simulation* 17 (1971) 197–205.
[10] M.J. Wozny and G.T. Heydt, "Hyperconical random search", *Transactions of the ASME, Journal of Dynamic Systems, Measurements, and Control* 94G (1972) 71–78.