Mathematical Programming 15 (1978) 330–342.
North-Holland Publishing Company

PROGRESSIVE GLOBAL RANDOM SEARCH OF CONTINUOUS FUNCTIONS

Luc P. DEVROYE*
McGill University, Montreal, Canada

Received 25 April 1977

A sequential random search method for the global minimization of a continuous function is proposed. The algorithm gradually concentrates the random search effort on areas neighboring the global minima. A modification is included for the case that the function cannot be exactly evaluated. The global convergence and the asymptotical optimality of the sequential sampling procedure are proved for both the stochastic and deterministic optimization problem.

Key words: Global Optimization, Random Search, Convergence, Sequential Minimization, Lipschitz Functions, Stochastic Programming.

1. Introduction

The global minimum of a real-valued function Q can be found by means of random search (see [10], [16] for surveys of the literature). If $B \subseteq \mathbf{R}^d$ is the search domain in which the minimum must be located, Q can be evaluated at $X_1, \ldots, X_n$, a sequence of independent random vectors with a common distribution (e.g., if B is a hypercube, take the uniform distribution in B). The estimate X_n^* of the minimum is the X_i with the lowest value $Q(X_i)$. This method, called crude search [3, 4] is discussed by Brooks [2] and compared on a theoretical basis with uniform grid search by Anderssen and Bloomfield [1]. Both methods have the drawback that they are essentially nonsequential and thus require a very large number of samples to estimate and locate the minimum with a high probability of achieving a certain accuracy.

A combination of local hill-climbing techniques and nonsequential global search appears to be the most popular technique nowadays in global minimization problems of smooth functions [6–8, 11–12, 14–15, 20, 23]. Hartman [7] and Cockrell and Fu [5] experimentally compare several techniques. Torn [20] attempts to define the classes of functions in which one technique is theoretically more efficient than the other. The global convergence of most random search methods is covered by the theorems found in [6] and [14]. The overall convergence rate of an algorithm is mainly determined by its global search routine, while the hill-climbing component in the algorithm merely insures the accuracy of the solution.

* The research is sponsored in part by the Air Force under Grant AFOSR-72-2371.

Shubert [17] reports an interesting sequential non-random search algorithm for the global minimization of Q when Q is Lipschitz, that is, there exists a finite number C such that for all $x, y \in B$,

$$|Q(x) - Q(y)| \leq C\|x - y\|. \tag{1}$$

Requiring the knowledge of the constant C, he finds that for $d = 1$ the number of function evaluations needed to locate the minimum (relative to uniform grid search) decreases drastically. All theoretical comparisons of random search and non-random global search seem to indicate that for $d > 6$, on the average, random search is more efficient [1]. Thus, for high dimensions, a randomized sequential global search may be very effective. In this paper, a method is developed that differs from crude search only in that gradually larger and larger sections of the search domain are excluded from further search, thus concentrating the global search effort on unexplored areas and areas that, based on the available information, can possibly contain the global minimum. Brooks [3] excludes from further search small spheres with a given radius and centered at previously investigated points. However, his method of exclusion does not depend upon the values of Q at these points. In the method developed below, the radii of these spheres are adapted as the search proceeds. Two versions are studied, one for which Q satisfies (1) and C must be known, and one for which Q can be any continuous function.

The convergence and rate of convergence of the algorithm is studied in the first part of the paper. In the second part, we assume that Q cannot be exactly evaluated and we discuss the changes that are necessary to still insure the overall convergence of the algorithm.

2. Progressive global random search

Assume that $B = [0, 1]^d$ and that (1) holds. Thus, there exists a point $x_0 \in B$ with $Q(x_0) = q_{\min} = \inf_B Q(x)$. If A is a subset of $\mathbf{R}^d$, let G_A denote the uniform distribution function on $A \cap B$. Consider then the following sequential global search method:

(i) Let X_1 be a random vector with distribution function G_B; evaluate $Q(X_1)$.

(ii) Given $X_1, Q(X_1), \ldots, X_n, Q(X_n)$, let

$$M_n = \min_{1 \leq i \leq n} Q(X_i)$$

and compute $R_i = (Q(X_i) - M_n)/C$, $1 \leq i \leq n$.

(iii) If A_n is the complement of $\bigcup_{i=1}^n S(X_i, R_i)$ ($S(x, a)$ denotes the closed sphere centered at x with radius a), X_{n+1} is a random vector with distribution function G_{A_n}; evaluate $Q(X_{n+1})$ and return to (ii).

After the n-th iteration, the estimate of the global minimum is $X_n^* = X_i$ if i is the smallest integer for which

$$Q(X_i) = M_n = \inf_{1 \le j \le n} Q(X_j).$$

The region excluded from search, $S(X_1, R_1) \cup \cdots \cup S(X_n, R_n)$, contains only points y with the property that

$$Q(y) \ge Q(X_i) - CR_i \ge M_n \quad \text{all } i,$$

and is of no interest anyway. If a combination of local and global search is desired, the following model suggested by Jarvis [10, 11] and others [5, 6] is both simple and practical. Replace (iii) by

(iv) Let X_{n+1} have distribution function $\alpha_n G_{A_n} + (1 - \alpha_n) G_n^*$ where $0 \le \alpha_n \le 1$ is a given number and G_n^* is any distribution function; evaluate $Q(X_{n+1})$ and return to (ii).

The distribution function G_n^*, a measurable function of $X_1, Q(X_1), \ldots, X_n, Q(X_n)$, accounts for local hill-climbing. For example, it is not uncommon to let G_n^* be Gaussian centered at X_n^* with a small variance σ_n^2 in all directions (see Matyas [14]). Or else, G_n^* can be the distribution function corresponding to Z_n, the outcome of a few steps of gradient descent started at X_n^*.

Theorem 1 deals with the convergence of $Q(X_n^*)$ to $q_{\min}$ as n grows large. It is applicable regardless of the choice of the G_n^*. It is further assumed that in (ii) the radii are determined by

$$R_i = \gamma_n (Q(X_i) - M_n), \quad 1 \le i \le n,$$

where γ_n is a nonnegative number depending upon n only.

Theorem 1. *Let $\{\alpha_n\}$ be a sequence from $[0, 1]$ with*

$$\sum_{n=1}^{\infty} \alpha_n = \infty, \tag{2}$$

let $\gamma_n \le 1/C$ for all n large enough, and let Q be Lipschitz with constant C (see (1)). Then $Q(X_n^*) \xrightarrow{n} q_{\min}$ with probability one. In particular, for all $\epsilon > 0$ there exist constants N (depending upon $\{\gamma_n\}$, C and ϵ) and $K_1 > 0$ (depending upon C and d) such that

$$\mathbf{P}\{Q(X_n^*) > q_{\min} + \epsilon\} \le \exp\Big(-K_1 \epsilon^d \sum_{i=N+1}^{n} \alpha_i\Big) \quad \text{all } n \ge N. \tag{3}$$

If Q is continuous, if (2) holds and if

$$\gamma_n \xrightarrow{n} 0, \tag{4}$$

then $Q(X_n^*) \xrightarrow{n} q_{\min}$ with probability one as well. Furthermore, for all $\epsilon > 0$ there exist constants N and $K_1 > 0$ (both depending upon $\{\gamma_n\}$, Q, d and ϵ only) such *that*

$$\mathbf{P}\{Q(X_n^*) > q_{\min} + \epsilon\} \le \exp\Big(-K_1 \sum_{i=N+1}^{n} \alpha_i\Big) \quad \text{all } n \ge N. \tag{5}$$

Theorem 1 encompasses the case that C is unknown. The sequence $\{\gamma_n\}$ must then be picked by the designer in such a way that (4) holds. In the limit case ($\gamma_n = 0$ for all n) the method reduces to crude search provided that $\alpha_n = 1$ for all n. The larger γ_n is, the more area is excluded from search, thus increasing the sequential search effect. This shows the importance of selecting γ_n as close as possible to the true value $1/C$. In any case, it is obvious that we must let

$$\gamma_n \le \inf_{i \ne j} \|X_i - X_j\| / |Q(X_i) - Q(X_j)|. \tag{6}$$

The estimate (6) may be much larger than $1/C$ however, and blindly using it for all n may make the algorithm nonconvergent. One practical problem arises in the process of the generation of X_{n+1}. To find X_{n+1} with distribution function G_{A_n}, we generate $Z_1, Z_2, \dots$, a sequence of independent random vectors with distribution function G_B and pick the first Z_j satisfying all the inequalities

$$\|Z_j - X_i\| > R_i, \quad 1 \le i \le n.$$

If (6) holds with strict inequality, then this procedure will with probability one stop after a finite number of Z_j's have been checked.

Let us now concentrate on the study of the efficiency of the new sampling procedure relative to crude search. The relative efficiency of a search technique at the n-th iteration can be defined as the function $g(n, \epsilon)$:

$$g(n, \epsilon) = \mathbf{P}\{Q(X_1) \ge q_{\min} + \epsilon\} / \mathbf{P}\{Q(X_{n+1}) \ge q_{\min} + \epsilon\} \tag{7}$$

where X_1 has distribution function G_B as in crude search. In the next theorem we show that for all $\epsilon > 0$, $g(n, \epsilon) \xrightarrow{n} \infty$ if the sequence $\{\gamma_n\}$ satisfies the slow convergence condition

$$n\gamma_n^d \xrightarrow{n} \infty. \tag{8}$$

Theorem 2. *If $B = [0, 1]^d$ and Q is bounded on B, if in the progressive random search algorithm one lets $\alpha_n = 1$ for all n, and if $\{\gamma_n\}$ satisfies* (8) *and*

$$\gamma_n, \quad 1/\gamma_{n+1}^d - 1/\gamma_n^d \quad \text{are nonincreasing for all } n \text{ large enough,}$$

$$n\gamma_n^d(1/\gamma_{n+1}^d - 1/\gamma_n^d) \xrightarrow{n} 0, \tag{9}$$

then $Q(X_n) \xrightarrow{n} q_{\min}$ in probability whenever $Q(X_n^*) \xrightarrow{n} q_{\min}$ in probability.

Theorem 2 links the convergence of $Q(X_n)$ to the convergence of $Q(X_n^*)$. For the sampling procedure to be asymptotically optimal (that is, to concentrate all its effort on that subset of B for which $Q(y)$ is near $q_{\min}$ whenever $y \in B$), it is necessary that $n\gamma_n^d \xrightarrow{n} \infty$. It is clear that if Q is Lipschitz with constant C, and $\gamma_n = 1/C$, then (8), (9) hold. Notice however that Theorem 2 remains valid even if Q is *not continuous*. The reason why (8) must hold is because at the n-th iteration every X_i excludes an area of size proportional to $\gamma_n^d(Q(X_i) - M_n)^d$ from further search, and the total volume of the excluded search area is strictly upper bounded by $Kn\gamma_n^d$ for some constant K. Since B is compact and $n\gamma_n^d \xrightarrow{n} \infty$, we see that $Q(X_i) - M_n$ must be small for large i and n. If γ_n tends to 0 so quickly that $n\gamma_n^d \xrightarrow{n} 0$, then the sampling procedure will asymptotically approach crude search, the relative efficiency $g(n, \epsilon)$ will tend to 1, and nothing is gained by using the progressive random search algorithm.

Condition (9) insures that γ_n varies slowly enough so that the holes that are created in B by decreasing γ_n can be filled up with new X_i before they grow too large.

Example (i) If $\gamma_n^d = n^{-a}$, $1 > a > 0$, then (8) holds and $\gamma_n \xrightarrow{n} 0$, but (9) is not satisfied.

(ii) If $\gamma_n^d = n^{-a_n}$ where $a_n \xrightarrow{n} 0$ and $a_n \log n \xrightarrow{n} \infty$, then (8) and (9) hold. Taking $a_n = (\log\log n/\log n)$ gives $\gamma_n^d = 1/\log n$. With $a_n = (\log n)^{p-1}$, $0 < p < 1$, we obtain $\gamma_n^d = \exp(-(\log n)^p)$.

(iii) The conditions of Theorem 2 are trivially satisfied for constant sequences $\gamma_n = \gamma_1$.

3. Minimization of the empirical risk

The above mentioned sequential random search procedure fails to give good results if Q cannot be exactly evaluated. Assume that

$$Q(x) = \int q(x, y)\,\mathrm{d}F(y) \tag{10}$$

where $q : \mathbf{R}^d \times \mathbf{R}^m \to \mathbf{R}$ is a given or computable function, and F is an unknown distribution function on $\mathbf{R}^m$. Instead of being able to evaluate $Q(x)$, we can observe values for $Y_1, Y_2, \ldots$, a sequence of independent random vectors with distribution function F. Given $Y_1, \ldots, Y_n$, $Q(x)$ can be estimated by the *empirical risk*

$$Q_n(x) = n^{-1} \sum_{i=1}^{n} q(x, Y_i). \tag{11}$$

Since by assumption $Q(x)$ is finite for all x, we know by the strong law of large numbers [13] that $Q_n(x) \xrightarrow{n} Q(x)$ with probability one. At the n-th iteration, we replace the unknown regression function Q by its estimate Q_{λ_n} where $\{\lambda_n\}$ is a given monotone sequence of integers. The modified algorithm proceeds as follows:

(i)* Let X_1 be a random vector with distribution function G_B; obtain and store $Y_1, \dots, Y_{\lambda_1}$.

(ii)* Given $X_1, \dots, X_n, Y_1, \dots, Y_{\lambda_n}$, define $M_n = \min_{1\le i\le n} Q_{\lambda_n}(X_i)$ and compute the radii $R_i = \gamma_n(Q_{\lambda_n}(X_i) - M_n)$, $1 \le i \le n$.

(iii)* X_{n+1} has distribution function $\alpha_n G_{A_n} + (1-\alpha_n)G_n^*$ where $0 \le \alpha_n \le 1$, G_n^* is any distribution function and A_n is the complement of $S(X_1, R_1) \cup \cdots \cup S(X_n, R_n)$; obtain values for $Y_{\lambda_n+1}, \dots, Y_{\lambda_{n+1}}$ and return to (ii)*.

The estimate of the minimum of Q after n iterations is $X_n^* = X_i$ whenever i is the first integer for which

$$Q_{\lambda_n}(X_i) = \min_{1\le j\le n} Q_{\lambda_n}(X_j) = M_n.$$

For $\alpha_n = 1, \gamma_n = 0$, an empirical version of the crude random search method is obtained. It should be noted that the empirical risk minimization technique requires the storage of the sequence $Y_1, Y_2, \dots$. For applications requiring the minimization of Q (10) for special classes of functions q, the reader is referred to the work of Vapnik and Chervonenkis [21], [22]. Sysoev [19] studies the convergence of $Q(X_n^{**})$ to $q_{\min}$ where X_n^{**} is the exact minimum of the empirical risk (11). However, to exactly minimize Q_n requires a large number of q-evaluations. In situations in which evaluations are costly, a suboptimal method such as progressive global random search may be more economical. Another suboptimal method requiring the convexity of Q and the differentiability of q is studied by Sysoev [19], who at every iteration takes one step in the direction of the gradient of $q(x, Y_n)$ at X_n^* and compares X_{n+1} and X_n^* using (11).

The conditions that we will impose upon q are not as strict as the ones suggested by Sysoev [19] or Vapnik and Chervonenkis [21–22]. We say that q satisfies a *uniform Lipschitz condition* if there exists a finite C such that

$$|q(x, y) - q(z, y)| \le C\|x - z\| \quad \text{all } y \in \mathbf{R}^m \quad \text{all } x, z \in B. \tag{12}$$

Sometimes the convergence of (i)*–(iii)* can be guaranteed under the weaker condition that q is *equicontinuous*, that is, for every $x \in B$ and every $\epsilon > 0$ there exists a $\delta > 0$ such that $|q(x, y) - q(z, y)| < \epsilon$ whenever $\|x - z\| < \delta$, regardless of the value of y. Both conditions are concerned with the variation of q with regard to x. We further require, turning now to the variation of q with regard to y, that F is such that the collection of random variables $\{q(x, Y_1) | x \in B\}$ is *uniformly integrable*, i.e.

$$\lim_{s\to\infty} \sup_{x\in B} \int_{|q(x,y)-Q(x)|\ge s} |q(x, y) - Q(x)| \, \mathrm{d}F(y) = 0. \tag{13}$$

The collection of random variables $\{q(x, Y_1) | x \in B\}$ is *uniformly dominated* if $|q(x, y) - Q(x)| \le f(y)$, all $x \in B$, and $\int |f(y)| \, \mathrm{d}F(y) < \infty$, a condition that is required by Sysoev [18] to show that $\sup_{x\in B} |Q_n(x) - Q(x)| \xrightarrow{n} 0$ with probability one. Clearly, uniform domination implies uniform integrability. In the next theorem, we show that if λ_n grows unbounded, then the empirical and deterministic versions of the algorithm have the same asymptotical properties.

Theorem 3. *Let* $B=[0,1]^d$, *let* Q *and* Q_n *be defined by* (10), (11), *let* q *be equicontinuous and let* $\{q(x, Y_1) \mid x \in B\}$ *be a uniformly dominated collection of random variables. If in* (i)*–(iii)* *we let*

$$\gamma_n \xrightarrow{n} 0, \tag{14}$$

$$\lambda_n \xrightarrow{n} \infty, \tag{15}$$

and

$$\sum_{n=1}^{\infty} \alpha_n = \infty, \tag{16}$$

then $Q(X_n^*) \xrightarrow{n} q_{\min}$ *with probability one.*

The same conclusions are valid if the equicontinuity condition and (14) are simultaneously replaced by the condition that q is uniformly Lipschitz with constant C and $\limsup \gamma_n C \le \frac{1}{3}$.

Explicit expressions for upper bounds for $\mathbf{P}\{Q(X_n^*) > q_{\min} + \epsilon\}$ can only be obtained if additional assumptions are made regarding q and F such as the existence of a finite number K such that

$$\sup_{x\in B} \int |q(x,y) - Q(x)|^t \, \mathrm{d}F(y) \le K, \quad \text{some } t > 1.$$

We will not investigate this point any further in this paper. In conclusion, we state a trivial extension of Theorem 2 to the noisy (empirical) case.

Theorem 4. *Assume that* $B=[0,1]^d$, *that* Q *is bounded on* B *and that* $\{q(x, Y_1) \mid x \in B\}$ *is uniformly dominated. If in* (i)*–(iii)* *we let* $\alpha_n = 1$, $\lambda_n \xrightarrow{n} \infty$, *and* $n\gamma_n^d \xrightarrow{n} \infty$ *such that* (9) *holds, then* $Q(X_n) \xrightarrow{n} q_{\min}$ *in probability whenever* $Q(X_n^*) \xrightarrow{n} q_{\min}$ *with probability one.*

Appendix

Proof of Theorem 1. Let L_d be the constant such that the volume (Lebesgue measure) of the sphere $S(x,r)$ is given by $L_d r^d$. Let N be so large that $\gamma_n \le 1/C$ for all $n \ge N$. Assume first that Q is Lipschitz with constant C and that $x_0 \in B$ is a point for which $Q(x_0) = q_{\min}$. Then,

$$\mathbf{P}\{Q(X_n^*) > q_{\min} + \epsilon\} = \mathbf{P}\left\{\bigcap_{i=1}^{n} \{Q(X_i) > q_{\min} + \epsilon\}\right\}$$

$$= \mathbf{P}\left\{\bigcap_{i=1}^{n} \{Q(X_i) > q_{\min} + \epsilon\} \bigcap_{i=N+1}^{n} \{X_i \notin S(x_0, \epsilon/C)\}\right\}.$$

But for every $i > N$,

$$\mathbf{P}\{X_{i+1} \in S(x_0, \epsilon/C) \mid Q(X_1) > q_{\min} + \epsilon, \ldots, Q(X_i) > q_{\min} + \epsilon\} \geq$$
$$\geq \alpha_i \text{ volume}(S(x_0, \epsilon/C))/\text{volume}(B) \geq \alpha_i L_d(\epsilon/C)^d/2^d.$$

Hence,

$$\mathbf{P}\{Q(X_n^*) > q_{\min} + \epsilon\} \leq \prod_{i=N+1}^{n} (1 - \alpha_i L_d(\epsilon/2C)^d) \leq$$
$$\leq \exp\left(-L_d(\epsilon/2C)^d \sum_{i=N+1}^{n} \alpha_i\right) \xrightarrow{n} 0.$$

This proves that $Q(X_n^*) \xrightarrow{n} q_{\min}$ in probability since ϵ is arbitrary. By the monotonicity of $Q(X_n^*)$, $Q(X_n^*) \xrightarrow{n} q_{\min}$ with probability one as well.

If Q is merely continuous (and thus uniformly continuous on B since B is a compact set) and if $\delta(\epsilon) > 0$ is so small that $|Q(z) - Q(x)| \leq \epsilon$ whenever $\|z - x\| \leq \delta(\epsilon)$, then the same argument can be used if N is such that $\gamma_n \leq \delta(\epsilon)/\epsilon$ for all $n \geq N$, and if (ϵ/C) is replaced by $\delta(\epsilon)$.

Lemma A.1. *If $\{\gamma_n\}$ is a sequence of nonnegative numbers satisfying* (8), (9), *then there exists a sequence of integers $\{b_n\}$ such that $b_n \xrightarrow{n} \infty$, $b_n/n \xrightarrow{n} 0$, $b_n\gamma_n^d \xrightarrow{n} \infty$ and $n(\gamma_{n-b_n}^d - \gamma_n^d) \xrightarrow{n} 0$.*

Proof of Lemma A.1. Without loss of generality we can and do assume that $1/\gamma_{n+1}^d - 1/\gamma_n^d$ and γ_n are monotone for all n. Thus, it is possible to write $\gamma_n^d = 1/(t_1 + \cdots + t_a)$ where the t_n are nonnegative and nonincreasing, and $(t_1 + \cdots + t_n)/n \xrightarrow{n} 0$. Let b_n be the nearest integer to δ_n/γ_n^d where

$$\delta_n \xrightarrow{n} \infty, \qquad \delta_n/n\gamma_n^d \xrightarrow{n} 0, \qquad \delta_{n+1} \leq \delta_n(1 + 1/n),$$

and

$$\delta_n n\gamma_n^d(1/\gamma_{n+1}^d - 1/\gamma_n^d) \xrightarrow{n} 0.$$

Such a sequence $\{\delta_n\}$ can be found since $n\gamma_n^d \xrightarrow{n} \infty$ and $\prod_{n=1}^{\infty}(1 + 1/n) = \infty$, and because $n\gamma_n^d(1/\gamma_{n+1}^d - 1/\gamma_n^d) \xrightarrow{n} 0$.

It is easy to check that $b_n \xrightarrow{n} \infty$, $b_n/n \xrightarrow{n} 0$ and $b_n\gamma_n^d \xrightarrow{n} \infty$. Also, since

$$\delta_n \leq \delta_{n-b_n}(1 + 1/n - b_n) \cdots (1 + 1/n - 1) \leq \delta_{n-b_n} \exp\left(\sum_{i=n-b_n}^{n-1} 1/i\right)$$
$$\leq \delta_{n-b_n}/(1 - b_n/n),$$

we have

$$\begin{aligned} n(\gamma_{n-b_n}^d - \gamma_n^d) &= n(t_{n-b_n+1} + \cdots + t_n)/(t_1 + \cdots + t_n)(t_1 + \cdots + t_{n-b_n}) \\ &\leq nb_n t_{n-b_n+1}\gamma_n^d\gamma_{n-b_n}^d \\ &\leq \delta_n(n - b_n)\gamma_{n-b_n}^d(1/\gamma_{n-b_n+1}^d - 1/\gamma_{n-b_n}^d)/(1 - b_n/n) \\ &\leq \delta_{n-b_n}(n - b_n)\gamma_{n-b_n}^d(1/\gamma_{n-b_n+1}^d - 1/\gamma_{n-b_n}^d)/(1 - b_n/n)^2 \end{aligned}$$

which tends to 0 as n grows large since $b_n/n \xrightarrow{n} 0$.

Proof of Theorem 2. If $B = [0, 1]^d$, then $T_n = \mathbf{P}\{Q(X_{n+1}) > q_{\min} + \epsilon \mid X_1, \ldots, X_n\}$ is the volume of $A_n \cap B \cap \{x \mid Q(x) > q_{\min} + \epsilon\}$. Clearly,

$$\begin{aligned}\mathbf{P}\{Q(X_{n+1}) > q_{\min} + \epsilon\} &= \\ = \mathbf{E}\{T_n\} \le \mathbf{E}\{T_n I_{\{T_n \le 2\theta\}}\} &+ \mathbf{P}\{Q(X_c^*) > q_{\min} + \tfrac{1}{2}\epsilon\} \\ &+ \mathbf{P}\left\{\{Q(X_c^*) \le q_{\min} + \tfrac{1}{2}\epsilon\} \bigcap_{i=n-b_n}^{n} \{T_i \ge \theta\}\right\}\end{aligned} \tag{17}$$

where $\theta > 0$ is to be specified later and c is a specially selected integer. First we find $x_1, \ldots, x_{N_n}$ in B with the property that for every $z \in B$, there exists a j with $\|x_j - z\| < \frac{1}{4}\gamma_n\epsilon$. It is obvious that we can pick $N_n = a/\gamma_n^d$ where a depends upon d and ϵ only. Assume now that $T_{n-b_n}, \ldots, T_n \ge \theta$, then at most N_n of $X_{c+1}, \ldots, X_n$ can have $Q(X_i) > q_{\min} + \epsilon$ provided that $Q(X_c^*) \le q_{\min} + \frac{1}{2}\epsilon$. Indeed, for every such $i > c$, find an x_j with $\|X_i - x_j\| < \frac{1}{4}\gamma_n\epsilon$. The same x_j cannot be used for two different values of i, say i_1 and i_2, because that would imply that $\|X_{i_1} - X_{i_2}\| \le \|X_{i_1} - x_j\| + \|X_{i_2} - x_j\| < \frac{1}{2}\gamma_n\epsilon$, which by the monotonocity of $\{\gamma_n\}$ and the fact that from the c-th iteration on, $R_i > \frac{1}{2}\gamma_n\epsilon$ whenever $Q(X_i) > q_{\min} + \epsilon$, is impossible. If $T_n \ge \theta$, then at most $c + a/\gamma_n^d$ of the $X_1, \ldots, X_n$ have $Q(X_i) > q_{\min} + \epsilon$.

Also, whenever $c < n - b_n \le i \le n$,

$$T_n - T_i \le L_d q^{*d} n(\gamma_{n-b_n}^d - \gamma_n^d) \tag{18}$$

where L_d is a constant depending upon d only and $q^* = \sup_{x\in B} Q(x) - \inf_{x \in B} Q(x)$.

First we let θ be small enough, noting that $\mathbf{E}\{T_n I_{\{T_n \le 2\theta\}}\} \le 2\theta$. Next we pick c so large that $\mathbf{P}\{Q(X_c^*) > q_{\min} + \frac{1}{2}\epsilon\} < \theta$. Then we find b_n such that (18) is smaller than θ (which in turn explains (17)), $c < n - b_n$, $(c + a/\gamma_n^d)/b_n < \frac{1}{2}\theta$ and $\exp(-2b_n(\frac{1}{2}\theta)^2) < \theta$ (Lemma A.1). For such large n,

$$\begin{aligned}&\mathbf{P}\left\{\{Q(X_c^*) \le q_{\min} + \tfrac{1}{2}\epsilon\} \bigcap_{i=n-b_n}^{n} \{T_i \ge \theta\}\right\} \le \\ &\le \mathbf{P}\left\{\left\{\sum_{j=1}^{b_n} I_{\{Q(X_{n-b_n+j}) > q_{\min}+\epsilon\}} \le a/\gamma_n^d\right\} \bigcap_{i=n-b_n}^{n} \{T_i \ge \theta\}\right\} \\ &\le \mathbf{P}\left\{b_n^{-1} \sum_{j=1}^{b_n} (U_j - \theta) \le a/\gamma_n^d b_n - \theta\right\}\end{aligned}$$

where $U_1, \ldots, U_{b_n}$ are Bernoulli random variables with $\mathbf{P}\{U_i = 1\} = \theta$. Since $a/\gamma_n^d b_n < \frac{1}{2}\theta$, we can upper bound this last expression by $\exp(-2b_n(\frac{1}{2}\theta)^2)$ using Hoeffding's inequality [9]. The theorem follows by the arbitrariness of θ and ϵ.

Proof of Theorem 3. If q is equicontinuous and B is compact, then for every $\epsilon > 0$ there exists a $\theta(\epsilon) > 0$ such that for all $y \in \mathbf{R}^m$, $x, z \in B$, $|q(x, y) - q(z, y)| < \epsilon$ whenever $\|x - z\| < \theta$. Indeed, by the equicontinuity of q, we can find $\delta(\epsilon, x)$ such that $|q(x, y) - q(z, y)| < \epsilon$ for all $z \in S(x, \delta(\epsilon, x))$. The collection of all open spheres

$S(x, \delta(\epsilon, x))$, $x \in B$, is a cover for the compact set B. By the Heine–Borel property, it is possible to find a finite subcover $\{S(x_1, \delta(\epsilon, x_1)), \ldots, S(x_K, \delta(\epsilon, x_K))\}$. We then let $\theta(\epsilon) = \min(\delta(\epsilon, x_1), \ldots, \delta(\epsilon, x_K))$. Notice that the same $\theta(\epsilon)$ can be used for Q and Q_n, regardless of the value the sequence $Y_1, \ldots, Y_n$ takes.

Before proving Theorem 3, we need two auxiliary results, which we state as separate lemmas.

Lemma A.2. *If B is compact, q is equicontinuous, $\{q(x, Y_1) \mid x \in B\}$ is uniformly dominated and $\lambda_n \xrightarrow{n} \infty$, then $\sup_{x \in B} |Q_{\lambda_n}(x) - Q(x)| \xrightarrow{n} 0$ with probability one.*

Proof of Lemma A.2. We need only show that for all $\theta > 0$,

$$\mathbf{P}\left\{\bigcup_{j=n}^{\infty} \sup_{x \in B} |Q_j(x) - Q(x)| > \theta\right\} \xrightarrow{n} 0.$$

First, we find $N_1(\theta)$ points $x_1, \ldots, x_{N_1}$ in B with the property that for every $z \in B$, $\|z - x_i\| < \delta$ for some $1 \le i \le N_1$ where $\delta > 0$ is so small that $|Q(z) - Q(x)| < \frac{1}{3}\theta$ and $|Q_j(z) - Q_j(x)| < \frac{1}{3}\theta$, all $j \ge 1$, all $\|z - x\| < \delta$, $z, x \in B$. If (12) holds, then it suffices to let $\delta = \theta/3C$. It is easy to deduce that

$$\sup_{x \in B} |Q_n(x) - Q(x)| \le \sup_{1 \le i \le N_1} |Q_n(x_i) - Q(x_i)| + \tfrac{2}{3}\theta.$$

A quick inspection of the proof of the strong law of large numbers [12] shows that the following uniform version of the strong law is valid. If $W_1, \ldots, W_n, \ldots$ are independent zero mean random variables with distribution function F_a where a is an element of an index set $\mathscr{A}$, then for all $\epsilon > 0$,

$$\sup_{a \in \mathscr{A}} \mathbf{P}\left\{\bigcup_{j=n}^{\infty} \left| j^{-1} \sum_{i=1}^{j} W_i \right| > \epsilon\right\} \xrightarrow{n} 0$$

provided that there exists a distribution function F_0 such that

$$\sup_{a \in \mathscr{A}} \int_{|w|>s} \mathrm{d}F_a(w) \le \int_{|w|>s} \mathrm{d}F_0(w), \quad \text{all } s$$

and

$$\int |w| \,\mathrm{d}F_0(w) < \infty.$$

Since $\{q(x, Y_1) \mid x \in B\}$ is a uniformly dominated collection of random variables, this property can be used to show that

$$\mathbf{P}\left\{\bigcup_{j=n}^{\infty} \{\sup_{x \in B} |Q_j(x) - Q(x)| > \theta\}\right\} \le$$

$$\le \mathbf{P}\left\{\bigcup_{i=1}^{N_1} \bigcup_{j=n}^{\infty} \{|Q_j(x_i) - Q(x_i)| > \tfrac{1}{3}\theta\}\right\}$$

$$\le N_1 \sup_{x \in B} \mathbf{P}\left\{\bigcup_{j=n}^{\infty} \{|Q_j(x) - Q(x)| > \tfrac{1}{3}\theta\}\right\} \xrightarrow{n} 0.$$

If $\|\cdot\|$ is the maximum component norm, then N_1 can be taken smaller than K_d/δ^d+1 where K_d is a constant depending upon d only.

Lemma A.3. *If B is compact, q is equicontinuous, $\{q(x, Y_1) \mid x\in B\}$ is uniformly integrable and $\lambda_n \xrightarrow{n} \infty$, then $\sup_{x\in B}|Q_{\lambda_n}(x)-Q(x)| \xrightarrow{n} 0$ in probability.*

Proof of Lemma A.3. The proof is similar to the proof of Lemma A.2. The crucial observation, a uniform version of the weak law of large numbers, is that if $W_1,\ldots,W_n$ are independent zero mean random variables with distribution function F_a where a is an element of an index set $\mathcal{a}$, then for all $\epsilon>0$,

$$\sup_{a\in\mathcal{a}} \mathbf{P}\left\{\left|n^{-1}\sum_{i=1}^{n} W_i\right|>\epsilon\right\} \xrightarrow{n} 0$$

provided that

$$\sup_{a\in\mathcal{a}} \int_{|w|>s} |w|\,\mathrm{d}F_a(w) \xrightarrow{s\to\infty} 0.$$

Proof of Theorem 3 (*contd.*). Let $\epsilon>0$ be arbitrary and let $x_0\in B$ be such that $Q(x_0)=q_{\min}$. Find $\delta>0$ so small that $\|x-z\|<\delta$, x, $z\in B$ implies that $|Q(z)-Q(x)|<\frac{1}{3}\epsilon$ and $|Q_j(x)-Q_j(z)|<\frac{1}{3}\epsilon$ for all j (note that if (12) holds, then we can let $\delta=\epsilon/3C$). Let C_N denote the event

$$C_N=\bigcap_{n=N}^{\infty}\{\sup_{x\in B}|Q_{\lambda_n}(x)-Q(x)|\le\tfrac{1}{3}\epsilon\}.$$

By Lemma A.2, find N large enough so that the probability of C_N^c is smaller than $\frac{1}{2}\eta$ and that $\gamma_n\le\delta/\epsilon$ for all $n\ge N$. For $k>N$,

$$\mathbf{P}\left\{\bigcup_{n=k}^{\infty}\{Q(X_n^*)>q_{\min}+2\epsilon\}\right\}=\mathbf{P}\left\{\bigcup_{n=k}^{\infty}\bigcap_{i=1}^{n}\{Q(X_i)>q_{\min}+\epsilon\}\right\}+\mathbf{P}\{C_N^c\}\le$$

$$\le\mathbf{P}\left\{\bigcup_{n=k}^{\infty}\bigcap_{i=N+1}^{n}\{\{X_i\notin S(x_0,\delta)\}\cap C_N\cap\bigcap_{i=1}^{n}\{Q(X_i)>q_{\min}+\epsilon\}\right\}+\mathbf{P}\{C_N^c\}.$$

However, if C_N occurs and $Q(X_i)>q_{\min}+\epsilon$ for $1\le i\le n$, then necessarily $M_n>q_{\min}+\frac{2}{3}\epsilon$. Under the same assumption, we show that whenever $Q(z)\le q_{\min}+\frac{1}{3}\epsilon$, z cannot belong to any of the $S(X_i, R_i)$, $n\ge i\ge N$. Indeed,

$$\begin{aligned}Q(z)&\ge Q(X_i)-\epsilon(R_i/\delta)=Q(X_i)-\epsilon(Q_{\lambda_n}(X_i)-M_n)\gamma_n/\delta\\&\ge M_n+Q_{\lambda_n}(X_i)-Q(X_i)\ge M_n-\tfrac{1}{3}\epsilon\ge q_{\min}+\tfrac{1}{3}\epsilon.\end{aligned}$$

Therefore, the set $A_0=\{x\mid Q(x)<q_{\min}+\frac{1}{3}\epsilon\}$ is properly included in A_n for all $n>N$, and since $S(x_0,\delta)$ is included in A_0, it is clear that for all $n>N$,

$$\mathbf{P}\left\{X_{n+1}\in S(x_0,\delta)\,\middle|\,\bigcap_{i=1}^{n}Q(X_i)>q_{\min}+\epsilon,\ \sup_{x\in B}|Q_{\lambda_n}(x)-Q(x)|\le\tfrac{1}{3}\epsilon\right\}\ge K_d\alpha_n\delta^d$$

where K_d is a constant depending upon d only. Arguing as in the proof of Theorem 1, we thus obtain for $k > N$,

$$\mathbf{P}\left\{\bigcup_{n=k}^{\infty}\{Q(X_n^*) > q_{\min} + 2\epsilon\}\right\} \leq \prod_{i=N+1}^{k}(1-\alpha_i K_d \delta^d) + \tfrac{1}{2}\eta$$

$$\leq \exp\left(-K_d\delta^d \sum_{i=N+1}^{k}\alpha_i\right) + \tfrac{1}{2}\eta$$

which is smaller than η for k large enough. Theorem 3 follows by the arbitrariness of η.

References

[1] R.S. Anderssen and P. Bloomfield, "Properties of the random search in global optimization", *Journal of Optimization Theory and Applications* 16 (1975) 383–398.

[2] G. Bennett, "Probability inequalities for the sums of independent random variables", *Journal of* Proceedings of the Third IFAC Symposium, Ischia, Italy (1973) 298–306.

[3] S.H. Brooks, "Discussion of random methods for locating surface maxima", *Operations Research* 6 (1958) 244–251.

[4] S.H. Brooks, "A comparison of maximum-seeking methods", *Operations Research* 7 (1959) 430–457.

[5] L.D. Cockrell and K.S. Fu, "On search techniques in adaptive systems", Tech. Rept. TR-EE-70-1, Purdue University, Lafayette (1970).

[6] L.P. Devroye, "On the convergence of statistical search", *IEEE Transactions on Systems, Man and Cybernetics* 6 (1976) 46–56.

[7] J.K. Hartman, "Some experiments in global optimization", *Naval Research Logistics Quarterly* 20 (1973) 569–576.

[8] J.D. Hill, "A search technique for multimodal surfaces", *IEEE Transactions on Systems, Science and Cybernetics* 5 (1969) 2–8.

[9] W. Hoeffding, "Probability inequalities for the sums of bounded random variables", *Journal of the American Statistical Association* 58 (1963) 13–30.

[10] R.A. Jarvis, "Optimization strategies in adaptive control: a selective survey", *IEEE Transactions on Systems, Man and Cybernetics* 5 (1975) 83–94.

[11] R.A. Jarvis, "Adaptive global search by the process of competitive evolution", *IEEE Transactions on Systems, Man and Cybernetics* 5 (1975) 297–311.

[12] R. Konakovsky and Z. Binder, "A multimodal searching technique using a learning controller", Proceedings of the Third IFAC Symposium, Ischia, Italy (1973) 298–306.

[13] M. Loeve, *Probability* (Van Nostrand, Princeton, NJ, 1963).

[14] J. Matyas, "Random optimization", *Automation and Remote Control* 26 (1965) 244–251.

[15] G.J. McMurtry and K.S. Fu, "A variable structure automaton used as a multimodal searching technique", *IEEE Transactions on Automatic Control* 11 (1966) 379–387.

[16] G.J. McMurtry, "Adaptive optimization procedures", in: J.M. Mendel and K.S. Fu, eds., *Adaptive, learning and pattern recognition systems* (Academic Press, New York, 1970).

[17] B.O. Shubert, "A sequential method seeking the global maximum of a function", *SIAM Journal on Numerical Analysis* 9 (1972) 379–388.

[18] L.P. Sysoev, "Statistical learning methods based on teacher identification", *Automation and Remote Control* 31 (1970) 1733–1741.

[19] L.P. Sysoev, "Training procedures which combine stochastic approximation and minimization of the empirical risk", *Automation and Remote Control* 34 (1973) 398–411.

[20] A. Torn, "Global optimization as a combination of global and local search", Skriftserie Utgiven Av Handelshogskolan vid Abo Akademi, Abo, Finland (1974).

[21] V.N. Vapnik and A.Ya. Chervonenkis, "Ordered risk minimization. I", *Automation and Remote Control* 35 (1974) 1226–1235.
[22] V.N. Vapnik and A.Ya. Chervonenkis, "Ordered risk minimization. II", *Automation and Remote Control* 35 (1974) 1403–1412.
[23] E.M. Vaysbord and D.B. Yudin, "Multiextremal stochastic approximation", *Engineering Cybernetics* 5 (1968) 1–10.