Journal of Global Optimization **10:** 1–16, 1997.

A Theoretical Approach to Restart in Global Optimization

MARCO MUSELLI
Istituto per i Circuiti Elettronici, CNR, 16149 Genova, Italy
(email: muselli@ice.ge.cnr.it)

(Received: 12 December 1995; accepted: 23 February 1996)

Abstract. While searching for the global minimum of a cost function we have often to decide if a restart from a different initial point would be more advantageous than continuing current optimization. This is a particular case of the efficiency comparison between repeated minimizations and single extended search having the same total length.

A theoretical approach for the treatment of this general problem forms the subject of the present paper. A fundamental role is played by the probability of reaching the global minimum, whose asymptotical behavior allows to provide useful information on the efficiency of repeated trials.

The second part of this work is devoted to a detailed analysis of three optimization algorithms whose evolution is independent of the cost function to be minimized: pure random search, grid search and random walk. These three examples give an interesting validation of the theoretical results and provide a general procedure which can be employed in the study of more complex optimization problems.

Key words: Optimization problem, restart, repeated searches, convergence probability.

1. Introduction

The global optimization of a cost function with several variables is a valuable way to obtain a satisfactory solution for a wide range of real-world problems (Bazaraa and Shetty, 1979). For this aim a huge number of different algorithms has been proposed that attempt to establish a good trade-off between search complexity and computational burden (Törn and Žilinskas, 1989).

Most of them contain a stochastic component which increases the robustness of the method and the adaptability to the characteristics of the cost function (Aluffi-Pentini et al., 1985; Baba, 1979; Boender et al., 1982; Corana et al., 1988; Muselli and Ridella, 1992). In fact, it can be shown that, under mild hypotheses, a random optimization algorithm is always able to reach the global minimum (within any arbitrary precision) with probability one when the computing time increases indefinitely (Devroye, 1976; Solis and Wets, 1981). However, quantitative estimates of the number of iterations required to obtain a satisfactory result are only given in particular cases (Dorea, 1983).

A typical execution of a stochastic minimization algorithm generally shows two consecutive phases. At first the sampled points are characterized by a cost function

value which decreases fast, but subsequently, if the convergence has not already been reached, the current minimum is rarely updated and apparently the search continues with poor efficiency.

It is then interesting to determine, with reference to the problem we are solving, if and when the optimization process must be restarted from a new initial point. Clearly, such an action is to be undertaken according to the corresponding probability of reaching the global minimum of the given cost function. Let us denote with n_1 the iteration in which the algorithm execution will be suspended; then it is advantageous to begin a new search with length n_2 if the probability $R^{(2)}(n_1, n_2)$ of finding the global minimum during the two consecutive trials is greater than the probability $R(n_1 + n_2)$ of reaching the convergence by continuing the initial search for other n_2 iterations.

More generally, it can be useful to evaluate the efficiency of executing t trials having lengths $n_1, \ldots, n_t$ with respect to a single search with $n_1 + \cdots + n_t$ iterations. This analysis has interesting implications in the parallelization of optimization algorithms. In fact, the simplest way of executing any global minimum search on a parallel machine is to carry out several different minimizations by assigning to each processor a different starting point. Thus, if we have theoretically proved that many consecutive searches are more favorable than a single extended trial, we can conclude that this simple parallel approach is really efficient.

A theoretical study of this problem is reported in (Kolen, 1988) with reference to the employment of optimization in neural network training. Through a questionable proof the general advantage of repeated trials is asserted, whatever optimization algorithm is considered. Unfortunately, this result is not valid in some practical cases: consider for example the grid method, in which a set of equidistributed samples in the cost function domain is analyzed one at a time to find the global minimum. We can directly conclude that any restart of this algorithm (before the complete inspection of the set) leads to a loss of information and consequently to a reduction of the total convergence probability. Thus, a single extended search is always more efficient when the grid method is applied.

A coherent theoretical approach which allows to obtain general conditions on the efficiency of repeated trials is the object of the present paper. Particular attention will be devoted to the influence of the asymptotical behavior of the probability $R(n)$ that the considered optimization algorithm converges to the global minimum of the given cost function.

An adequate framework for the problem above is contained in section 2 together with the reduction of the generic analysis for t repeated trials to the special case $t = 2$. Section 3 reports the main results and the study of some typical asymptotical behaviors of the convergence probability $R(n)$. The thorough analysis of three optimization methods which do not depend on the cost function to minimize will be the subject of section 4 and allows to validate the theoretical assertions.

2. The General Problem

Consider a measurable function $f(\boldsymbol{x})$ defined on a compact set $D \subset \mathcal{R}^m$, and denote with V_D the Lebesgue measure of its domain D. Let f^* be the global minimum of $f(\boldsymbol{x})$ in D; we introduce the following convergence set H:

$$H = \{\boldsymbol{x} \in D : f(\boldsymbol{x}) < f^* + \varepsilon\}$$

being $\varepsilon > 0$ a given tolerance in the determination of the global minimum. Since the cost function $f(\boldsymbol{x})$ is measurable by hypothesis, the convergence set H will also be measurable; let us denote with V_H its Lebesgue measure.

As one will note in the following, the results contained in the present paper holds for different definitions of the convergence set H provided that its measurability is ensured. It is also possible to employ a measure μ different from the Lebesgue's one; the only basic hypothesis to be verified is that both the domain D and the convergence set H have non null measure.

Then consider a generic optimization algorithm $\mathcal{A}$; at every iteration it samples a point $\boldsymbol{x}_n \in D$ according to proper choice criteria which depend on the cost function values in the points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_{n-1}$ previously reached. If $\boldsymbol{x}_n \in H$ we say that the algorithm $\mathcal{A}$ has converged in the nth iteration and the sampling is stopped. By hypothesis, the first point $\boldsymbol{x}_1$ is randomly chosen with uniform probability in the whole domain D.

Let us call *optimization problem* $\Omega = (\mathcal{A}, f)$ the search for the global minimum f^* (or, equivalently, a point of the convergence set H) of the cost function $f(\boldsymbol{x})$ done by the algorithm $\mathcal{A}$. We want to find under what conditions it is more efficient making several repeated trials (starting from different randomly chosen initial points in D) than executing a single search with extended length.

For this aim let us denote with $R_\Omega(n)$, $n \geq 0$, the probability that the optimization algorithm $\mathcal{A}$ reaches the convergence set H of the cost function f within the nth iteration. From this definition we obtain that $R_\Omega(n)$ is a monotonically increasing sequence of real numbers with the following initial values:

$$R_\Omega(0) = 0\,, \qquad R_\Omega(1) = V_H/V_D \tag{1}$$

since the first point $\boldsymbol{x}_1$ is randomly chosen in the domain D by the algorithm $\mathcal{A}$ with uniform probability.

Then given t positive integers $n_1, \ldots, n_t$, we want to compare the following quantities:

- the probability $R_\Omega^{(t)}(n_1, \ldots, n_t)$ that the algorithm $\mathcal{A}$ converges during t consecutive optimizations having length $n_1, \ldots, n_t$ respectively and
- the probability $R_\Omega(n_1 + \cdots + n_t)$ that the algorithm $\mathcal{A}$ converges in a single global minimum search with $n_1 + \cdots + n_t$ iterations.

In this quantitative analysis of repeated trials it is assumed that on the average every iteration requires the same computing time. When this condition is not

satisfied proper corrective factors should be applied to the following theoretical results.

The comparison above can be executed by writing the quantities $R_\Omega^{(t)}(n_1, \ldots, n_t)$ and $R_\Omega(n_1 + \cdots + n_t)$ in terms of the probability $Q_\Omega(n)$ that the optimization algorithm $\mathcal{A}$ does not reach the convergence set H within the nth iteration:

$$Q_\Omega(n) = P_\Omega(\boldsymbol{x}_1 \notin H, \ldots, \boldsymbol{x}_n \notin H) = 1 - R_\Omega(n) \tag{2}$$

where the subscript Ω remembers that every probability depends on the optimization problem we are considering.

Since two subsequent searches for the global minimum are independent each other, the probability that the convergence is not reached in t trials is equal to the product of the probabilities of not finding a point of H in every single search. Thus we have

$$R_\Omega^{(t)}(n_1, \ldots, n_t) = 1 - Q_\Omega(n_1) \cdots Q_\Omega(n_t) = 1 - \prod_{i=1}^{t} Q_\Omega(n_i). \tag{3}$$

Furthermore, equation (2) gives

$$R_\Omega(n_1 + \cdots + n_t) = 1 - Q_\Omega(n_1 + \cdots + n_t) = 1 - Q_\Omega\left(\sum_{i=1}^{t} n_i\right) \tag{4}$$

Then the difference $R_\Omega^{(t)}(n_1, \ldots, n_t) - R_\Omega(n_1 + \cdots + n_t)$ can be written in the following way:

$$\begin{aligned} R_\Omega^{(t)}(n_1, \ldots, n_t) - R_\Omega(n_1 + \cdots + n_t) &= Q_\Omega\left(\sum_{i=1}^{t} n_i\right) - \prod_{i=1}^{t} Q_\Omega(n_i) = \\ &= \sum_{j=1}^{t-1} \prod_{k=1}^{j-1} Q_\Omega(n_k) \left(Q_\Omega\left(\sum_{i=j}^{t} n_i\right) - Q_\Omega(n_j) Q_\Omega\left(\sum_{i=j+1}^{t} n_i\right)\right) = \\ &= \sum_{j=1}^{t-1} \prod_{k=1}^{j-1} Q_\Omega(n_k) \left(R_\Omega^{(2)}\left(n_j, \sum_{i=j+1}^{t} n_i\right) - R_\Omega\left(n_j + \sum_{i=j+1}^{t} n_i\right)\right). \end{aligned} \tag{5}$$

But $Q_\Omega(n_k) > 0$ for every k; thus, the sign of the terms in the summation is given by the differences $R_\Omega^{(2)}(n_j, \sum_{i=j+1}^{t} n_i) - R_\Omega(n_j + \sum_{i=j+1}^{t} n_i)$. In particular, if for every j we have

$$R_\Omega^{(2)}\left(n_j, \sum_{i=j+1}^{t} n_i\right) > R_\Omega\left(n_j + \sum_{i=j+1}^{t} n_i\right)$$

we also obtain

$$R_\Omega^{(t)}(n_1, \ldots, n_t) > R_\Omega(n_1 + \cdots + n_t)$$

and consequently t repeated trials are more efficient. The opposite is true when for every j

$$R_{\Omega}^{(2)}\left(n_j, \sum_{i=j+1}^{t} n_i\right) < R_{\Omega}\left(n_j + \sum_{i=j+1}^{t} n_i\right).$$

The original problem can be therefore reduced to the comparison between the convergence probabilities of two consecutive trials with length n_1 and n_2 respectively and a single extended search having length $n_1 + n_2$.

3. Efficiency Evaluation of Two Consecutive Searches

During the execution of an optimization algorithm it is sometimes necessary to stop the current trial and restart a new search from a different initial point since further iterations do not sufficiently improve the probability of reaching the global minimum of the cost function. Thus, the problem of evaluating the efficiency of two consecutive searches can be expressed in the following way: at what iteration n_1 is it advantageous to stop an optimization algorithm and start a new search with length n_2 rather than doing a single trials having length $n_1 + n_2$?

Such a problem is difficult to solve in a general case; a thorough theoretical treatment can be obtained when we focus on the asymptotical behavior of the convergence probability $R_{\Omega}(n)$. This analysis has however great practical interest; in fact, it allows to determine if the execution of an optimization algorithm is to be continued by examining the convergence speed to the global minimum.

For this aim let us introduce two basic definitions:

DEFINITION 1. An optimization problem Ω will be called *multistart suited* if there exists a positive integer n such that for every $n_1 \geq n$ and every n_2 we have $R_{\Omega}(n_1 + n_2) \leq R_{\Omega}^{(2)}(n_1, n_2)$. It will be called *singlestart suited* if $R_{\Omega}(n_1 + n_2) \geq R_{\Omega}^{(2)}(n_1, n_2)$.

Therefore, when we are facing with a multistart suited optimization problem it is more efficient, after a given iteration n, to stop the execution of the algorithm $\mathcal{A}$ and begin a new search starting from a different initial point. On the contrary, in singlestart suited optimization problems it is better to continue the current search if a sufficient number n of iterations has already been carried out.

From (3) and (4) in the case $m = 2$ we obtain the corresponding definitions in terms of the non-convergence probability $Q_{\Omega}(n)$:

$$R_{\Omega}(n_1 + n_2) \leq R_{\Omega}^{(2)}(n_1, n_2) \iff Q_{\Omega}(n_1 + n_2) \geq Q_{\Omega}(n_1)Q_{\Omega}(n_2)$$
$$R_{\Omega}(n_1 + n_2) \geq R_{\Omega}^{(2)}(n_1, n_2) \iff Q_{\Omega}(n_1 + n_2) \leq Q_{\Omega}(n_1)Q_{\Omega}(n_2)$$

Because of their immediate applicability the inequalities containing the probability $Q_{\Omega}(n)$ will be preferred for determining the characterization of a given problem Ω.

We can also classify optimization problems with regard to the limit value of the corresponding convergence probability $R_\Omega(n)$:

DEFINITION 2. An optimization problem Ω will be called *certain* if $\lim_{n\to+\infty} R_\Omega(n) = 1$; otherwise it will be called *uncertain*.

With these premises let us analyze the asymptotical behavior of the non-convergence probability $Q_\Omega(n)$ to obtain general information on the efficiency of two repeated searches for the global minimum. For sake of simplicity, let us remove from the notation for the probabilities $R(n)$ and $Q(n)$ the explicit dependence on the optimization problem Ω we are considering. From the general properties of the convergence probability $R(n)$ we obtain that $Q(n)$ is a decreasing sequence of real numbers. Moreover, by setting $\alpha = 1 - V_H/V_D$ equations (1) gives the following initial values for $Q(n)$:

$$Q(0) = 1\,, \qquad Q(1) = \alpha$$

it should be observed that $0 \leq \alpha < 1$ since by hypothesis $V_H > 0$.

Consider at first the class of optimization problems for which two repeated trials and a single extended search have the same efficiency; in these cases we obtain

$$Q(n_1)Q(n_2) = Q(n_1 + n_2).$$

If such a relation is valid independently of the values assumed by n_1 and n_2 we have

$$Q(n) = Q(1)Q(n-1) = (Q(1))^2 Q(n-2) = (Q(1))^n = \alpha^n. \tag{6}$$

This non-convergence probability α^n characterizes the only class of optimization problems for which the execution of two consecutive trials is always equivalent to a single continuative search. As we will find in Section 4, every time the pure random search method is applied an optimization problem of this kind arises.

It can also be easily shown that if $Q(n) = a^n$ (where a can be different from α) for every n greater than a given positive integer k, repeated trials and single search are equivalent when both n_1 and n_2 are greater than k. Thus we can reasonably suppose that the non-convergence probability $Q(n) = a^n$ forms in one sense a boundary between multistart and singlestart suited optimization problems. The following lemma provides a theoretical validation of this intuitive assertion:

LEMMA 1. *If* $(Q(n_2))^{1/n_2} > \limsup_{n\to+\infty} Q(n+1)/Q(n)$ *there exists a positive integer* k *such that for every* $n_1 \geq k$ *we have* $Q(n_1 + n_2) < Q(n_1)Q(n_2)$.

If $(Q(n_2))^{1/n_2} < \liminf_{n\to+\infty} Q(n+1)/Q(n)$ *there exists a positive integer* k *such that* $Q(n_1 + n_2) > Q(n_1)Q(n_2)$ *for every* $n_1 \geq k$.

Proof. If we set $\lambda = \limsup_{n\to+\infty} Q(n+1)/Q(n)$ then for every $\varepsilon > 0$ there exists $n_\varepsilon \geq 1$ such that $Q(n+1) < (\lambda + \varepsilon)Q(n)$ for every $n \geq n_\varepsilon$. Take $\varepsilon = ((Q(n_2))^{1/n_2} - \lambda)/2$ and set $k = n_\varepsilon$; for every $n_1 \geq k$ we have

$$Q(n_1 + n_2) < (\lambda + \varepsilon)^{n_2} Q(n_1) < Q(n_1)Q(n_2)$$

since by hypothesis

$$(Q(n_2))^{1/n_2} > \left((Q(n_2))^{1/n_2} + \lambda\right)/2 = \lambda + \varepsilon.$$

The second assertion of the lemma can be proved in a similar way.

Unfortunately this lemma binds the evaluation of the efficiency of two consecutive searches to the length n_2 of the second one. In some cases of great practical interest such a dependence can be removed.

THEOREM 1. *An optimization problem is singlestart suited if* $\lim_{n\to+\infty} Q(n+1)/Q(n) = 0$ *and multistart suited when* $\lim_{n\to+\infty} Q(n+1)/Q(n) = 1$.

Proof. Consider at first the case $\lim_{n\to+\infty} Q(n+1)/Q(n) = 0$; if $Q(n_2) > 0$, from lemma 1 we obtain $Q(n_1 + n_2) < Q(n_1)Q(n_2)$ for n_1 not less than a given integer k. On the other hand, when $Q(n_2) = 0$ we have also $Q(n_1+n_2) = 0$, since $Q(n)$ is a decreasing sequence.

A similar reasoning holds in the case $\lim_{n\to+\infty} Q(n+1)/Q(n) = 1$; in fact lemma 1 gives $Q(n_1 + n_2) > Q(n_1)Q(n_2)$ since $Q(n_2) < 1$ for every $n_2 \geq 1$.

A direct application of this theorem gives the following general result:

COROLLARY 1. *Every uncertain optimization problem is multistart suited.*

Thus, when we are facing with an uncertain optimization problem it is more efficient to execute two consecutive trials if one of them runs for a time long enough. A situation of this kind occurs when we employ a local optimization algorithm to search for the minimum of a multivariate cost function. Note that corollary 1 is in agreement with practical experience: the global minimum can generally be found by executing repeated trials starting from different initial points.

Our theoretical study can then be restricted to certain optimization problems: in this case a complete theoretical analysis of some general types of asymptotical behaviors of the non-convergence probability $Q(n)$ is possible. Information on the efficiency of repeated trials in a real application can be obtained by comparing an estimate of the asymptotical behavior of the current non-convergence probability $Q(n)$ with proper reference sequences like that contained in the following two examples.

EXAMPLE 1. Consider the class of optimization problems whose probability $Q(n)$ has the following property

$$Q(n) \sim n^{-\gamma} \qquad \text{with } \gamma > 0 \tag{7}$$

where the symbol '$\sim$' denotes equal asymptotical behavior:

$$a_n \sim b_n \qquad \Longleftrightarrow \qquad \lim_{n\to+\infty} a_n/b_n = 1$$

Note that (7) implies $\lim_{n\to+\infty} Q(n) = 0$ for which the corresponding optimization problem is certain; furthermore

$$\lim_{n\to+\infty} \frac{Q(n+1)}{Q(n)} = \lim_{n\to+\infty} \frac{Q(n+1)}{(n+1)^{-\gamma}} \cdot \frac{n^{-\gamma}}{Q(n)} \cdot \frac{(n+1)^{-\gamma}}{n^{-\gamma}} = 1$$

for every $\gamma > 0$. Thus, theorem 1 allows to assert that all the problems belonging to this class are multistart suited, independently of the value of the positive exponent γ.

EXAMPLE 2. A second important class of optimization problems is characterized by the following asymptotical behavior of the non-convergence probability $Q(n)$:

$$Q(n) \sim a^{n^\gamma} \qquad \text{with } 0 < a < 1 \text{ and } \gamma > 0.$$

In this case we have again

$$\lim_{n\to+\infty} Q(n) = \lim_{n\to+\infty} a^{n^\gamma} \cdot \frac{Q(n)}{a^{n^\gamma}} = 0$$

for which we are considering certain optimization problems. Moreover

$$\lim_{n\to+\infty} \frac{Q(n+1)}{Q(n)} = \lim_{n\to+\infty} \frac{a^{(n+1)^\gamma}}{a^{n^\gamma}} = \begin{cases} 1 & \text{se } 0 < \gamma < 1 \\ a & \text{if } \gamma = 1 \\ 0 & \text{if } \gamma > 1 \end{cases}$$

Thus, three different situations can occur:

- when $0 < \gamma < 1$ the corresponding optimization problem is singlestart suited,
- when $\gamma > 1$ the corresponding optimization problem is multistart suited,
- in the case $\gamma = 1$ the relative efficiency of two repeated trials depends on the value of $Q(n_2)$ according to lemma 1.

4. Analysis of Some Optimization Algorithms

In this section we will compute the non-convergence probability $Q(n)$ for some optimization problems which are independent of the cost function to minimize. The application of the theoretical results obtained above will allow to establish when consecutive trials are more efficient than a single extended search for the global minimum.

Although the analyzed optimization algorithms are not employed for the solution of real-world problems in the simple form considered here, the practical implications (which is deferred to a following publication) can be very interesting. Furthermore, the procedure employed for this analysis can also be applied to other more complex cases.

Given a generic optimization problem $\Omega = (\mathcal{A}, f)$, from definition (2) for the probability $Q_\Omega(n)$ we obtain

$$Q_\Omega(n) = \prod_{k=1}^{n} (1 - p_\Omega(k)) \tag{8}$$

where

$$p_\Omega(k) = P_\Omega(\boldsymbol{x}_k \in H | \boldsymbol{x}_{k-1} \notin H, \ldots, \boldsymbol{x}_1 \notin H) \tag{9}$$

is the probability that the algorithm $\mathcal{A}$ converges in the kth iteration.

But

$$\lim_{n\to+\infty} \frac{Q_\Omega(n+1)}{Q_\Omega(n)} = \lim_{n\to+\infty} 1 - p_\Omega(n+1) = 1 - \lim_{n\to+\infty} p_\Omega(n) \tag{10}$$

thus, theorem 1 allows to find the following general rule: if $p_\Omega(n) \to 0$ ($p_\Omega(n) \to 1$) the optimization problem $\Omega = (\mathcal{A}, f)$ is multistart (singlestart) suited. Furthermore, by using (8) the expression for $Q_\Omega(n)$ can be obtained by computing the convergence probabilities $p_\Omega(k)$, $k = 1, 2, \ldots, n$, given by (9).

With these premises we can analyze three optimization algorithms whose evolution is independent of the cost function to minimize.

4.1. PURE RANDOM SEARCH

Pure random search is the simplest stochastic method for global optimization: it samples at random points in the domain D of the cost function $f(\boldsymbol{x})$ until an element of the convergence set H is found. Since every choice does not depend on the previous ones, we obtain

$$\begin{aligned} p_{\mathrm{RS}}(k) &= P_{\mathrm{RS}}(\boldsymbol{x}_k \in H | \boldsymbol{x}_{k-1} \notin H, \ldots, \boldsymbol{x}_1 \notin H) = \\ &= P_{\mathrm{RS}}(\boldsymbol{x}_k \in H) = V_H / V_D = 1 - \alpha \end{aligned}$$

assuming that the sampling probability is uniform in D. Then, from (8) we obtain

$$Q_{\mathrm{RS}}(n) = \prod_{k=1}^{n} (1 - p_{\mathrm{RS}}(k)) = \alpha^n$$

As already noted in section 3, the behavior (6) for the non-convergence probability $Q(n)$ corresponds thus to every optimization problem that employs the pure random search algorithm. In these cases the theoretical analysis concludes that repeated trials and single extended search have the same efficiency. This result can also be obtained by a simple reasoning: restarting a pure random search has no effect on the probability of reaching the global minimum since every sampling is independent of the others.

4.2. GRID SEARCH

In the solution of practical problems the grid search method is often employed to obtain initial coarse-grained information on the cost function to minimize. In this algorithm the domain D is subdivided into d measurable subsets $S_1, \ldots, S_d$, h of which are included in the convergence set H. At every iteration one subset S_j not

already controlled is randomly chosen and the convergence is achieved if $S_j \subset H$. In fact, it is sufficient to pick any point in S_j to have a satisfying estimate of the global minimum f^*.

It can be easily seen that the convergence probability at the kth iteration $p_{\text{GS}}(k)$ is given by

$$p_{\text{GS}}(k) = \begin{cases} \dfrac{h}{d-k+1} & \text{if } 1 \leq k \leq d-h \\ 1 & \text{if } k > d-h \end{cases}$$

since in the case $1 \leq k \leq d-h$ we have already chosen without success $k-1$ subsets. If $k > d-h$ every choice leads to the convergence.

Thus, from (8) we obtain the following expression for $Q_{\text{GS}}(n)$ when $n \leq d-h$:

$$\begin{aligned} Q_{\text{GS}}(n) &= \prod_{k=1}^{n} \left(1 - \frac{h}{d-k+1}\right) = \prod_{k=1}^{n} \frac{d-h-n+k}{d-n+k} = \\ &= \frac{(d-h)!(d-n)!}{d!(d-h-n)!} = \binom{d-n}{h} \Big/ \binom{d}{h} \end{aligned} \tag{11}$$

while $Q_{\text{GS}}(n) = 0$ for $n > d-h$.

Now, let us verify in a direct manner that a single extended search is always more efficient in the optimization problems which employ the grid search method. For this aim, let us compare the quantities $Q_{\text{GS}}(n_1)Q_{\text{GS}}(n_2)$ and $Q_{\text{GS}}(n_1+n_2)$ for generic numbers of iterations n_1 and n_2. If $n_1 > d-h$ or $n_2 > d-h$ equation (11) gives

$$Q_{\text{GS}}(n_1)Q_{\text{GS}}(n_2) = 0 = Q_{\text{GS}}(n_1+n_2)$$

consequently repeated trials and single search are trivially equivalent since both of them converge.

On the contrary, if $n_1 \leq d-h$ and $n_2 \leq d-h$ but $n_1+n_2 > d-h$, we obtain

$$Q_{\text{GS}}(n_1)Q_{\text{GS}}(n_2) > 0 = Q_{\text{GS}}(n_1+n_2)$$

for which a single extended search is more advantageous since it surely leads to the convergence set H.

Finally, in the case $n_1+n_2 \leq d-h$ we can write

$$\begin{aligned} \frac{Q_{\text{GS}}(n_1+n_2)}{Q_{\text{GS}}(n_1)Q_{\text{GS}}(n_2)} &= \frac{\binom{d-n_1-n_2}{h}\binom{d}{h}}{\binom{d-n_1}{h}\binom{d-n_2}{h}} = \\ &= \prod_{k=0}^{h-1} \frac{(d-n_1-n_2-k)(d-k)}{(d-n_1-k)(d-n_2-k)} \\ &= \prod_{k=0}^{h-1} \frac{(d-n_1-k)(d-n_2-k)-n_1n_2}{(d-n_1-k)(d-n_2-k)} < 1 \end{aligned}$$

thus, a single search is again more efficient.

Such a result is also motivated by a simple practical observation: if the execution of a grid search is restarted all the information acquired on the previous sampled subsets S_j which do not belong to the convergence set is irreparably lost. This corresponds to a reduction in the total convergence probability.

4.3. RANDOM WALK

The third optimization algorithm here considered is the random walk method; according to this technique a global minimum of the cost function $f(\boldsymbol{x})$ is searched for by performing a random walk in the domain D. Periodic boundary conditions are assumed; therefore, if the random walk leaves the domain D in a given direction, it reenters the same set at the opposite side. Although this is not a practical optimization algorithm, some minimization techniques (e.g. the simulated annealing method (Kirkpatrick et al., 1983; Corana et al., 1988)) use it as a basic component.

For sake of simplicity, let us suppose that the domain D is rectangular

$$D = \{\boldsymbol{x} \in \mathcal{R}^m | a_i \leq x_i \leq b_i,\ a_i, b_i \in \mathcal{R},\ \text{for } i = 1, \ldots, m\}$$

being x_i the ith component of the point $\boldsymbol{x}$. To obtain an explicit expression for the non-convergence probability $Q_{\text{RW}}(n)$ let us subdivide the definition interval $[a_i, b_i]$ in the ith direction into d_i (eventually different) parts.

In this way $d = d_1 \cdot d_2 \cdots d_m$ rectangular subsets S_j, $j = 1, \ldots, d$, which entirely cover the domain D have been determined. Each of them is associated to an m-dimensional integer vector $\boldsymbol{z}$ whose ith component z_i corresponds to the projection of the subset on the interval $[a_i, b_i]$. Thus we have $1 \leq z_i \leq d_i$ for every $i = 1, \ldots, m$. The vector notation $\boldsymbol{z}$ will be often used in the following to denote the corresponding subset S_j.

To avoid unnecessary complications let us suppose that the convergence set H is formed by the union of a finite number h of subsets S_j. At first let us restrict ourselves to the case $h = 1$ and denote with $\boldsymbol{z}^*$ the unique subset belonging to H. The generalization of the following results to the case $h > 1$ will be straightforward.

The random walk method starts the search by choosing at random, with uniform probability, a subset $\boldsymbol{z}_1$ in the domain D. At every subsequent iteration one of the $2m$ possible directions is selected with equal probability and the adjacent subset along this direction is visited to check its eventual belonging to the convergence set H. Obviously the search continues until the subset $\boldsymbol{z}^*$ is found.

Our theoretical analysis of the relative efficiency of repeated trials in the random walk method begins with the monodimensional case ($m = 1$). To compute the convergence probability $p^{(1)}_{\text{RW}}(k)$ at the kth iteration let us denote with $w = |z_1 - z^*|$ the (integer) distance between the subset z_1 initially chosen by the algorithm and the optimal subset z^*.

Since boundary conditions are periodic, the subset d is considered adjacent to the subset 1; we can therefore assert that the probability of reaching z^* at the kth iteration, starting from z_1, is given by the sum of two distinct components:

– the probability $u^{(1)}_{k-1}(w,d)$ that a monodimensional symmetric random walk, starting from the (integer) point w visits at step $k-1$ the extremity 0 of the interval $[0,d]$.
– the probability $u^{(1)}_{k-1}(d-w,d)$ that the same random walk visits at step $k-1$ the extremity d of the interval $[0,d]$.

Thus we obtain

$$p^{(1)}_{\text{RW}}(k) = d^{-1} \sum_{w=0}^{d-1} \left(u^{(1)}_{k-1}(w,d) + u^{(1)}_{k-1}(d-w,d) \right). \tag{12}$$

The problem of obtaining the non-convergence probability $Q^{(1)}_{\text{RW}}(n)$ is then reduced to that of finding an explicit expression for the quantities $u^{(1)}_k(w,d)$. Feller (Feller, 1968, pag. 353) gives the following formula in case of symmetric random walk when $k > 1$:

$$u^{(1)}_k(w,d) = d^{-1} \sum_{\nu=1}^{d-1} \cos^{k-1} \frac{\pi\nu}{d} \sin \frac{\pi\nu}{d} \sin \frac{\pi w\nu}{d} \tag{13}$$

from which we have

$$\begin{aligned} u^{(1)}_k(d-w,d) &= d^{-1} \sum_{\nu=1}^{d-1} \cos^{k-1} \frac{\pi\nu}{d} \sin \frac{\pi\nu}{d} \sin \frac{\pi(d-w)\nu}{d} = \\ &= d^{-1} \sum_{\nu=1}^{d-1} (-1)^{\nu+1} \cos^{k-1} \frac{\pi\nu}{d} \sin \frac{\pi\nu}{d} \sin \frac{\pi w\nu}{d}. \end{aligned} \tag{14}$$

By substituting (13) and (14) in (12) we obtain the expression for the convergence probability $p^{(1)}_{\text{RW}}(k)$ when $k > 2$

$$\begin{aligned} p^{(1)}_{\text{RW}}(k) &= d^{-2} \sum_{w=0}^{d-1} \sum_{\nu=1}^{d-1} \left(1 - (-1)^{\nu}\right) \cos^{k-2} \frac{\pi\nu}{d} \sin \frac{\pi\nu}{d} \sin \frac{\pi w\nu}{d} = \\ &= 2d^{-2} \sum_{w=0}^{d-1} \sum_{\nu=1}^{\lfloor d/2 \rfloor} \cos^{k-2} \left(\pi \frac{2\nu-1}{d} \right) \sin \left(\pi \frac{2\nu-1}{d} \right) \cdot \\ &\quad \cdot \sin \left(\pi w \frac{2\nu-1}{d} \right) \end{aligned} \tag{15}$$

having denoted with $\lfloor x \rfloor$ the integer not greater than x.

This formula can be simplified by using the identity

$$\sum_{w=0}^{d-1} \sin \gamma w = \frac{\sin \frac{\gamma d}{2} \sin \frac{\gamma(d-1)}{2}}{\sin \frac{\gamma}{2}}.$$

Thus, if we set $\gamma = \pi(2\nu-1)/d$ equation (15) becomes

$$\begin{aligned} p^{(1)}_{\mathrm{RW}}(k) &= 2d^{-2}\sum_{\nu=1}^{\lfloor d/2\rfloor}(-1)^{\nu+1}\cos^{k-2}\left(\pi\frac{2\nu-1}{d}\right)\sin\left(\pi\frac{2\nu-1}{d}\right)\cdot \\ &\quad\cdot\sin\left(\pi\frac{(2\nu-1)(d-1)}{2d}\right)\Big/\sin\left(\pi\frac{2\nu-1}{2d}\right) = \\ &= 4d^{-2}\sum_{\nu=1}^{\lfloor d/2\rfloor}(-1)^{\nu+1}\cos^{k-2}\left(\pi\frac{2\nu-1}{d}\right)\cos\left(\pi\frac{2\nu-1}{2d}\right)\cdot \\ &\quad\cdot\sin\left(\pi\frac{(2\nu-1)(d-1)}{2d}\right) \end{aligned} \tag{16}$$

since in this case

$$\sin\frac{\gamma d}{2} = \sin\pi\left(\nu-\frac{1}{2}\right) = (-1)^{\nu+1}.$$

Furthermore, for $k=0$ and $k=1$ we have

$$\begin{aligned} u^{(1)}_0(w,d) &= \begin{cases} 1 & \text{if } w=0 \\ 0 & \text{otherwise} \end{cases} \\ u^{(1)}_1(w,d) &= \begin{cases} 1/2 & \text{if } w=1 \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{17}$$

from which follows $p^{(1)}_{\mathrm{RW}}(1) = p^{(1)}_{\mathrm{RW}}(2) = 1/d$.

Now, let us employ the expression (16) for $p^{(1)}_{\mathrm{RW}}(k)$ to obtain information on the relative efficiency of two repeated trials. Note that

$$\begin{aligned} \lim_{k\to+\infty} p^{(1)}_{\mathrm{RW}}(k) &= 4d^{-2}\sum_{\nu=1}^{\lfloor d/2\rfloor}(-1)^{\nu+1}\cos\left(\pi\frac{2\nu-1}{2d}\right)\cdot \\ &\quad\cdot\sin\left(\pi\frac{(2\nu-1)(d-1)}{2d}\right)\lim_{k\to+\infty}\cos^{k-2}\left(\pi\frac{2\nu-1}{d}\right) = 0 \end{aligned}$$

since $|\cos\pi(2\nu-1)/d| < 1$ for every $\nu = 1,\ldots,\lfloor d/2\rfloor$. Thus, for theorem 1 and equation (10) every monodimensional optimization problem using the random walk method is multistart suited.

In the general case where the domain D has dimension $m>1$ it can be easily seen that the probability of reaching the subset $\boldsymbol{z}^*$ at the kth iteration, starting from the initial subset $\boldsymbol{z}_1$, is again given by the sum of two components:

- the probability $u^{(m)}_{k-1}(\boldsymbol{w},\boldsymbol{d})$ that a m-dimensional symmetric random walk, starting from $\boldsymbol{w}$ visits at step $k-1$ the point $\mathbf{0}=(0,\ldots,0)$.
- the probability $u^{(m)}_{k-1}(\boldsymbol{d}-\boldsymbol{w},\boldsymbol{d})$ that the same random walk visits at step $k-1$ the point $\boldsymbol{d}=(d_1,\ldots,d_m)$.

being $\boldsymbol{w}$ the integer vector whose generic components w_i is given by the distance $|z_{1i} - z_i^*|$. Thus, the probability $p_{\text{RW}}^{(m)}(k)$ of reaching the convergence at the kth iteration has the following form:

$$p_{\text{RW}}^{(m)}(k) = d^{-1} \sum_{w_1=0}^{d_1-1} \cdots \sum_{w_m=0}^{d_m-1} \left(u_{k-1}^{(m)}(\boldsymbol{w}, \boldsymbol{d}) + u_{k-1}^{(m)}(\boldsymbol{d} - \boldsymbol{w}, \boldsymbol{d}) \right). \tag{18}$$

Since at every iteration the choice of a new subset is done with equal probability in the $2m$ coordinate directions we obtain

$$\begin{aligned} & u_k^{(m)}(\boldsymbol{w}, \boldsymbol{d}) = \\ & = m^{-k} \sum_{k_1+\cdots+k_m=k} \frac{k!}{k_1! \cdots k_m!} u_{k_1}^{(1)}(w_1, d_1) \cdots u_{k_m}^{(1)}(w_m, d_m) \end{aligned} \tag{19}$$

where the quantities $u_{k_i}^{(1)}(w_i, d_i)$ are given by (13) for $k > 1$ and by (17) for $k = 0, 1$. The substitution of (19) in (18) provides the desired formula for the convergence probability at the kth iteration when the domain D is m-dimensional.

An analysis of the relative efficiency of two repeated trials in the general case can be executed by noting that (13) gives for $k > 1$

$$\begin{aligned} u_k^{(1)}(w, d) \;&\leq\; d^{-1} \sum_{\nu=1}^{d-1} \left| \cos^{k-1} \frac{\pi\nu}{d} \right| \cdot \left| \sin \frac{\pi\nu}{d} \right| \cdot \left| \sin \frac{\pi w \nu}{d} \right| \leq \\ &\leq\; d^{-1} \sum_{\nu=1}^{d-1} \cos^{k-1} \frac{\pi}{d} \leq \cos^{k-1} \frac{\pi}{d} \end{aligned}$$

since $|\cos(\pi\nu/d)| \leq \cos(\pi/d)$ when $1 \leq \nu \leq d-1$. This inequality is also valid for $k = 0, 1$; consequently by using (19) we can write

$$\begin{aligned} u_k^{(m)}(\boldsymbol{w}, \boldsymbol{d}) \;&\leq\; m^{-k} \sum_{k_1+\cdots+k_m=k} \frac{k!}{k_1! \cdots k_m!} \cos^{k_1-1} \frac{\pi}{d_1} \cdots \cos^{k_m-1} \frac{\pi}{d_m} \leq \\ &\leq\; m^{-k} \cos^{k-m} \frac{\pi}{d} \sum_{k_1+\cdots+k_m=k} \frac{k!}{k_1! \cdots k_m!} = \cos^{k-m} \frac{\pi}{d} \end{aligned} \tag{20}$$

being $d = d_1 \cdots d_m$. Since the right hand side does not depend on $\boldsymbol{w}$, the same upper bound is also valid for $u_k^{(m)}(\boldsymbol{d} - \boldsymbol{w}, \boldsymbol{d})$.

By substituting (20) in (18) we obtain the following inequality for $p_{\text{RW}}^{(m)}(k)$

$$p_{\text{RW}}^{(m)}(k) \leq 2d^{-1} \sum_{w_1=0}^{d_1-1} \cdots \sum_{w_m=0}^{d_m-1} \cos^{k-m-1} \frac{\pi}{d} = 2 \cos^{k-m-1} \frac{\pi}{d}.$$

Since this upper bound vanishes when k increases indefinitely we can conclude by virtue of theorem 1 that an optimization problem employing the random walk method is multistart suited also in the multidimensional case.

Such a result remains valid when the convergence set H contains a number $h > 1$ of subsets available for sampling. In this case we can consider the h optimization problems having one subset of these as convergence set; let us denote with $p_1(k), \ldots, p_h(k)$ the corresponding convergence probabilities at the kth iteration having expressions given by (18).

Then the convergence probability $p_{\mathrm{RW}}^{(m)}(k)$ of the original problem is always not greater than the sum $p_1(k) + \cdots + p_h(k)$ for which we obtain

$$\lim_{k \to +\infty} p_{\mathrm{RW}}^{(m)}(k) \leq \sum_{i=1}^{h} \lim_{k \to +\infty} p_i(k) = 0.$$

Thus the corresponding optimization problem is again multistart suited.

In the achievement of this general result we have not controlled if the considered optimization problem is certain. But, for an important theorem on random walk (Feller, 1968, pagg. 359–362) we can assert that $\lim_{n \to +\infty} Q_{\mathrm{RW}}^{(m)}(n) = 0$ only when $m = 1$ (monodimensional case) or $m = 2$ (bidimensional case), whereas for $m > 2$ (apart from very particular shapes of the set H) the convergence to the global minimum is not ensured.

5. Conclusions

The problem of giving suitable conditions for the restart of a global optimization algorithm has been approached in a theoretical way. The behavior of the convergence probability $R(n)$ (or equivalently that of its reciprocal $Q(n)$) allows to characterize different classes of optimization problems and establish the relative efficiency of repeated trials.

In particular a thorough theoretical analysis is only possible when asymptotical properties are considered. In case of certain optimization problems, for which the achievement of the global minimum is ensured when the number of iterations increases indefinitely, the limit value of the ratio $Q(n+1)/Q(n)$ allows to solve the efficiency comparison between repeated trials and single extended search. In the opposite case (uncertain problems) a general theoretical result asserts the superiority of consecutive searches.

The detailed analysis of three optimization algorithms (pure random search, grid search and random walk), whose evolution is independent of the cost function to be minimized has allowed to obtain a first validation of the theoretical approach employed. The procedure used for the efficiency evaluation of repeated trials has achieved in all the cases the correct expression for the non-convergence probability $Q(n)$ and can be applied in the study of other more complex optimization problems.

Acknowledgements

Thanks are due to dr. Elena Muselli for her valuable comments.

References

Aluffi-Pentini, F., Parisi, V. and Zirilli, F. (1985), Global Optimization and Stochastic Differential Equations, *Journal of Optimization Theory and Applications* **47**, 1–16.

Baba, N. (1979), Global Optimization of Functions by the Random Optimization Method, *International Journal of Control* **30**, 1061–1065.

Bazaraa, M.S. and Shetty, C.M. (1979), *Nonlinear Programming: Theory and Algorithms*, John Wiley & Sons, New York.

Boender, C.G.E., Rinnooy Kan, A.H.G., Stougie, L. and Timmer, G. (1982), A Stochastic Method for Global Optimization, *Mathematical Programming* **22**, 125–140.

Corana, A., Marchesi, M., Martini, C. and Ridella, S. (1988), Minimizing Multimodal Functions of Continuous Variables with the Simulated Annealing Algorithm, *ACM Transactions on Mathematical Software* **13**, 262–280.

Devroye, L.P. (1976), On the Convergence of Statistical Search, *IEEE Transactions on Systems, Man, and Cybernetics* **SMC-6**, 46–56.

Dorea, C.C.Y. (1983), Expected Number of Steps of a Random Optimization Method, *Journal of Optimization Theory and Applications* **39**, 165–171.

Feller, W. (1968), *An Introduction to Probability Theory and Its Applications, vol. 1*, John Wiley & Sons, New York, 3rd ed.

Kirkpatrick, S., Gelatt, C.D. and Vecchi, M.P. (1983), Optimization by Simulated Annealing, *Science* **220**, 671–680.

Kolen, J.F. (1988), Faster Learning through a Probabilistic Approximation Algorithm, *Proceedings of the IEEE Second International Conference on Neural Networks*, vol. I, 449–454.

Museli, M. and Ridella, S. (1992), Global Optimization of Functions with the Interval Genetic Algorithm, *Complex Systems* **6**, 193–212.

Solis, F.J. and Wets, R.J-B. (1981), Minimization by Random Search Techniques, *Mathematics of Operation Research* **6**, 19–30.

Törn, A. and Žilinskas, A. (1989), *Global Optimization*, Springer-Verlag, Berlin.