Journal of Global Optimization **13:** 25–42, 1998.

Relaxing the Assumptions of the Multilevel Single Linkage Algorithm

MARCO LOCATELLI
University of Trier, Department of Mathematics, D-54286 Trier, Germany
E-mail: locatel@uni-trier.de

(Received 17 July 1995; accepted 15 November 1997)

Abstract. In this paper we relax the assumptions of a well known algorithm for continuous global optimization, Multilevel Single Linkage (MLSL). It is shown that the good theoretical properties of MLSL are shared by a slightly different algorithm, Non-monotonic MLSL (NM MLSL), but under weaker assumptions. The main difference with MLSL is the fact that in NM MLSL some non-monotonic sequences of sampled points are also considered in order to decide whether to start or not a local search, while MLSL only considers monotonic decreasing sequences. The modification is inspired by non-monotonic methods for local searches.

Key words: Multilevel single linkage, Multistart algorithms, Non-monotonic sequences.

1. Multistart algorithms and multilevel single linkage

Multistart algorithms are well known iterative stochastic methods for the solution of the global optimization problem

$$\min_{x \in X} f(x), \quad X \subseteq \Re^d.$$

In what follows, the main concepts related to this kind of algorithm are presented.

Let C_k denote the set of points sampled up to iteration k. Let us assume that a local search procedure $LS(\cdot)$ is available, which receives a point belonging to X as an input and returns a local minimum of f over X. The main part of a Multistart algorithm is the following:

at iteration k, select a set $S_k \subseteq C_k$ of sampled points and apply the procedure LS to each of them.

In most of the Multistart algorithms presented in the literature the points are sampled from the uniform distribution over X. Multistart algorithms do not, therefore, employ clever sampling mechanisms. In this kind of algorithm the sampling mechanism only has to guarantee that the feasible region is globally explored. Their cleverness is determined by the LS starting point selection mechanism. In particular, the selection should avoid two kinds of errors. Before presenting them we need a definition. Given a local minimum $\overline{x}$, the set:

$$\mathcal{A}(\overline{x}) = \{y \in X : LS(y) = \overline{x}\},$$

is called the region of attraction of $\overline{x}$. The two kinds of errors are:

Error I. The region of attraction of a local minimum $\overline{x}$ contains at least one sampled point, but LS has never been applied to points in this region; formally

$$C_k \cap \mathcal{A}(\overline{x}) \neq \emptyset,$$

but

$$(\cup_{j=1}^{k} S_j) \cap \mathcal{A}(\overline{x}) = \emptyset.$$

Error II. LS has been applied to two or more points belonging to the region of attraction of a local minimum $\overline{x}$ so that the same local minimum is detected more than once; formally

$$| (\cup_{j=1}^{k} S_j) \cap \mathcal{A}(\overline{x}) | \geqslant 2.$$

The main differences between Multistart algorithms consist of the way they try to avoid the two kinds of errors above through the selection, at each iteration, of the set S_k's. Notice that a clever selection of S_k generally implies an increase in the computational effort. On the other hand, a clever selection also saves computational effort because good solutions can be reached within a limited number of iterations and also with a limited number of local searches.

Many papers on Multistart algorithms can be found in the literature (see e.g. [1, 3, 7, 9–13, 15]). In particular, in [10] and [11] one of the best known Multistart algorithms has been presented, Multilevel Single Linkage (MLSL).

In MLSL the mechanism of selection of the set S_k is based on the analysis of sequences of 'close' sampled points with decreasing function values. Before describing MLSL in more detail, we introduce some notation:

- ∂X is the border of X;
- $d(\cdot, \cdot)$ is the euclidean distance;
- $m(\cdot)$ is the Lebesgue measure.

We also introduce the following definition:

DEFINITION 1. *Given $r > 0$, a sequence $\{y_i\}_{i=1}^{t}$ satisfying*

$$d(y_i, y_{i+1}) \leqslant r, \quad \forall\, i = 1, \ldots, t-1,$$

and

$$f(y_{i+1}) < f(y_i),$$

is called a decreasing r-sequence.

The MLSL algorithm can be described as follows:

Initialization. Let the integer $N > 0$, $\gamma \in (0, 1]$ and $d_1, d_2 > 0$ be fixed, let $k := 1$ and

$$\alpha_k := \pi^{-1/2} \left(\Gamma\left(1 + \frac{d}{2}\right) m(X) \sigma \frac{\log kN}{kN} \right)^{1/d}; \tag{1}$$

denote with S^* the set of already detected local minima and set $S^* = \emptyset$.

Sampling phase. At step k generate a uniform random sample of size N over X.

Reduction of the sample. Sort the whole sample of kN points in order of increasing function values, and select the γkN points with the lowest values; the resulting sample R_k is called reduced sample.

Selection of the set S_k. For *every* point $x_i \in R_k$ repeat the following procedure:

Consider all the decreasing α_k-sequences $s_1, \ldots, s_r$ whose first point is x_i, made up by points in R_k and ending at a point $x_{final}(s_i)$ which satisfies one of the following conditions:

1. $d(x_{final}(s_i), \partial X) \leqslant d_1$;
2. $d(x_{final}(s_i), S^*) \leqslant d_2$;
3. LS has already been applied to $x_{final}(s_i)$;
4. no other point $y \in R_k$ can be found with

$$d(y, x_{final}(s_i)) \leqslant \alpha_k \quad \text{and} \quad f(y) < f(x_{final}(s_i)).$$

If $\exists\, i,\ 1 \leqslant i \leqslant r$, such that $x_{final}(s_i)$ satisfies conditions 1, 2 or 3 above, then do not start any local search. Otherwise apply LS to

$$y^* = \arg\min_{x_j}\{f(x_j) \ : \ \exists\, i,\ 1 \leqslant i \leqslant r, \text{ such that } x_j \in s_i\}.$$

and if a new local minimum is detected, add it to S^*.

Stopping phase. Check a stopping criterion and, if it fails, return to the sampling phase with $k := k + 1$.

An equivalent description of the 'selection of the set S_k' phase is the following.

Selection of the set S_k. Apply LS to all the points $x_i \in R_k$ which satisfy every one of the following conditions:

- no point $y \in R_k$ exists such that:

$$d(x_i, y) \leqslant \alpha_k, \quad f(y) < f(x_i);$$

- $d(x_i, \partial X) > d_1$;
- $d(x_i, S^*) > d_2$;
- LS was not previously applied to x_i.

For implementation purposes, this last description is certainly much better than the previous one. However, the first description makes easier to understand on which idea MLSL is based. In MLSL, if it can find a decreasing α_k-sequence leading either to a point to which LS has already been applied and has lead to a local minimum x^*, or to a point which is close to an already detected local minimum x^*

(i.e. $x^* \in S^*$), then no local search is started from points belonging to the sequence. This decision is based on the conjecture that the decreasing α_k-sequence can be seen as generated by a local optimization procedure leading to the already detected local minimum x^*.

There is actually another case in which LS is not applied to the points of a sequence, i.e. when the sequence gets close enough to the border of the feasible region. That is related to point *d* of Assumption 1 introduced below, which requires that no local minimum exists along the border of the feasible region.

In the description of MLSL, the stopping criterion has not been specified. In [11, p. 74] the following has been suggested:

$$\text{stop when} \quad \frac{\mid S^* \mid (\gamma kN - 1)}{\gamma kN - \mid S^* \mid -2} \leqslant \mid S^* \mid + \frac{1}{2}. \tag{2}$$

For more details about this stopping criterion see [10, p. 32] and the references therein. Essentially, the left side in (2) is the expected number of local minima of f over X and we stop when this is not greater than the current number of already detected local minima plus 0.5. In what follows, the stopping rule problem will not be further investigated and only the problem of defining a clever local search starting point selection mechanism will be addressed. Therefore, we refer to the existing literature on stopping rules (see e.g. [2, pp. 853-864]).

Besides working demonstrably well in practice, MLSL has also good theoretical properties. In particular, we mention the following properties:

Property I. Every local minimum is detected in finite time with probability one;

Property II. If $\sigma > 4$ in (1), then the expected number of local searches started by the algorithm is finite, even if the algorithm is never stopped.

The first result ensures that after a finite time with probability one there is no Error I. The second result is not a guarantee against Error II, but even if the algorithm is never stopped, the expected number of occurrences of this error is finite.

The theoretical results are obtained under the following assumptions:

ASSUMPTION 1.

a. $f \in C^2$*;*
b. X compact, convex and with non empty interior;
c. finite number of stationary points;
d. stationary points in the interior of X.

The aim of this paper is to introduce a modification of MLSL, through which the same theoretical properties described above can be obtained, but under weaker assumptions.

Many functions which may occur in global optimization do not satisfy Assumption 1. For instance, point d excludes all concave functions whose minima belong to the border of a convex set X, even if for this problem the use of deterministic

methods seems to be more appropriate (see e.g. [4]). Moreover, a simple function such as:

$$f(x,y) = \begin{cases} x^2 + ay^2 & -1 \leqslant x \leqslant 1 \quad -1 \leqslant y \leqslant 0 \\ x^2 & -1 \leqslant x \leqslant 1 \quad 0 \leqslant y \leqslant 1 \\ x^2 + b(y-1)^2 & -1 \leqslant x \leqslant 1 \quad 1 \leqslant y \leqslant 2 \end{cases} \tag{3}$$

where $a, b \neq 0$ are constants, neither satisfies point a nor point c.

In Section 2, we will introduce a new set of assumptions weaker than Assumption 1. In Section 3, a modification of MLSL is introduced and the differences between the two algorithms are described. Finally, in Section 4, it is shown that the theoretical properties of MLSL are preserved, but under the weaker assumptions introduced in Section 2.

2. Relaxing the assumptions

Now we introduce some definitions, which are a generalization of the classical definitions of maximum, minimum and saddle points. They are used to relax the assumption that only a finite number of stationary points are allowed:

DEFINITION 2. *Let $A' \subseteq X$ be a set of points such that either it contains only one point which is a saddle point or the following are satisfied:*

1. *A' is connected and maximal (with respect to the inclusion)*;
2. $\forall\, x \in A',\ f(x) = f(A') = const$;
3. *one of the following is satisfied:*

$$\forall\, x \in A'\ \exists\, \epsilon = \epsilon(x) > 0 \text{ s.t. } \forall\, y \in X,\ d(y,x) \leqslant \epsilon: \quad f(y) \geqslant f(x), \tag{4}$$

or

$$\forall\, x \in A'\ \exists\, \epsilon = \epsilon(x) > 0 \text{ s.t. } \forall\, y \in X,\ d(y,x) \leqslant \epsilon: \quad f(y) \leqslant f(x). \tag{5}$$

Let $A = cl(A')$, i.e. A is the closure of A'. Notice that if f is continuous, then $\forall\, x \in A,\ f(x) = f(A) = f(A')$. Then we define A as a

– *<u>minimum set</u> if (4) is satisfied and*

$$\exists\, \epsilon > 0:\ \forall\, y \in X \setminus A, \quad d(y,A) \leqslant \epsilon \implies f(y) > f(A);$$

– *<u>saddle set</u> if one of the following is satisfied:*

- *it contains only one point which is a saddle point;*
- *it contains more than one point, either (4) or (5) is satisfied and $\forall\, \epsilon > 0$:*

$$\exists\, y_1, y_2 \in X \setminus A,\ d(y_i, A) \leqslant \epsilon,\ i = 1,2: \quad f(y_1) < f(A) < f(y_2).$$

– *<u>maximum set</u> if (5) is satisfied and*

$$\exists\, \epsilon > 0:\ \forall\, y \in X \setminus A, \quad d(y,A) \leqslant \epsilon \implies f(y) < f(A).$$

A minimum, saddle or maximum set is also called a stationary set and a point inside a stationary set is called a representant of it.

As an instance, we can consider the function (3), where the set $A = \{(x, y) : x = 0,\ 0 \leqslant y \leqslant 1\}$ is a minimum set if $a, b > 0$, while it is a saddle set if $a < 0 < b$ or $b < 0 < a$. Notice that any stationary set contains at least one minimum, saddle or maximum point. Also notice that in this case the procedure LS will be assumed to return a representant of a stationary set.

Now we are ready to substitute assumptions 1 with some weaker ones:

ASSUMPTION 2.

a. f lipschitzian with Lipschitz constant L;
b. X compact, convex and with non empty interior;
c. there exists a finite number of stationary sets.

It will be assumed that $L \geqslant 1$. This is not true in general, but it is always true that we can take an upper bound for L not lower than 1. Therefore in what follows, $L \geqslant 1$ denotes the Lipschitz constant or an upper bound of it. Moreover, it should be noted that L is only used in the theoretical results and it is not necessary to know its precise value in the modified MLSL algorithm. The Lipschitz constant represents precious information about the problem, which is often hard to obtain, and many algorithms have been developed for the case in which it is known (for the interested reader, general books on global optimization such as [4, 5, 14, 16], besides introducing some techniques, give many references to papers on the subject).

Under Assumption 2, some of the results exploited in [10] and [11] in order to prove the theoretical results of MLSL, are lost. For instance, the proof of property II for MLSL is based on Lemma 7 in [10], but the proof of this lemma relies on the assumption $f \in C^2$. Moreover, dealing with minimum sets instead of minimum points renders inoperable the fact that if we are at a small distance $d_1 > 0$ from an already detected minimum we do not start a local search. Indeed, what we detect is not the whole minimum set but one of its representants and we are only able to compute the distance from this representant and not from the whole set. Therefore, there is a larger risk of detecting the same minimum set many times.

3. The new algorithm

We first need a definition.

DEFINITION 3. *Let $r > 0$ and $s \geqslant 0$. A sequence $\{y_i\}_{i=1}^t$ satisfying the following for any i, $i = 1, \ldots, t-1$:*

$$\begin{aligned} &d(y_i, y_{i+1}) \leqslant r, \\ &f(y_{i+1}) < \min_{j<i+1} f(y_j) + s \end{aligned} \tag{6}$$

is called a (r, s)-sequence.

Notice that decreasing r-sequences are a special case of (r, s)-sequences, in which $s = 0$. We consider the following algorithm:

Initialization. Let $k := 1$, $S^* = \emptyset$ and consider two sequences of parameters $\{\alpha_k\}$ and $\{\beta_k\}$.

Sampling phase. Sample one point x_k from the uniform distribution over X.

Selection of the set S_k. *Let $s_1, \ldots, s_t$ be all the possible $(\alpha_k, \beta_k)-$sequences $\{y_i\}$* made up by distinct sampled points (i.e. distinct points inside the set C_k), and with the following characteristics:

1. $y_1 = x_k$, i.e. the first point of these sequences is the last sampled point;
2. the final point of these sequences, denoted with $x_{final}(s_i)$, satisfies one of the following conditions:
   a) LS has already been applied to it;
   b) $d(x_{final}(s_i), S^*) \leqslant \alpha_k$;
   c) no other point in C_k different from those in the sequence and satisfying (6) with $r = \alpha_k$ and $s = \beta_k$, can be found.

If $\exists\, i,\ 1 \leqslant i \leqslant t$, such that $x_{final}(s_i)$ satisfies condition (a) or (b), then do not start any local search. Otherwise, let:

$$y^* = \arg\min_{x_j}\{ f(x_j) \ :\ \exists\, i,\ 1 \leqslant i \leqslant t,\ \text{such that } x_j \in s_i\},$$

and start a local search from y^*; if a new representant of a minimum set is detected, add it to S^*.

Notice that only sequences whose first point is the last sampled one are considered, while in this step of MLSL all sequences whose first point is in R_k are considered.

Stopping phase. Check a stopping criterion (see e.g. (2)) and, if it fails, return to the sampling phase with $k := k + 1$.

This modification of MLSL is called Non-monotonic MLSL (NM MLSL). The reason for this name is the use of (α_k, β_k)-sequences which are not necessarily strictly decreasing as in MLSL, but may be non-monotonic. The two main differences between MLSL and NM MLSL are those written in italics and underlined in the description of NM MLSL. The second one could actually easily be removed without losing the theoretical properties proved for NM MLSL in Section 4. On the other hand, it can be seen that the computational effort which is required by the analysis of (α_k, β_k)-sequences is greater than the effort for the analysis of decreasing sequences (see also [8]) and thus it seems reasonable to avoid performing this analysis for all the sequences whose first point is in R_k.

But the most important difference between NM MLSL and MLSL is certainly the use of the non-monotonic (α_k, β_k)-sequences instead of the strictly decreasing

α_k-sequences. Actually, if we set $\beta_k = 0, \ \forall\, k$ in NM MLSL, the non-monotonic (α_k, β_k)-sequences become decreasing (α_k)-sequences, which shows that NM MLSL is a generalization of MLSL. The idea of introducing non-monotonic sequences comes from algorithms for unconstrained local optimization in which some ascent steps can be made. These have been proven to perform well when applied to ill-conditioned test functions (see [6]). This also suggests that NM MLSL may perform better than MLSL when applied to ill-conditioned functions. Some limited computational experiences seem to confirm this conjecture (see [8]). More extensive computational experiences will appear elsewhere.

There are other minor differences between MLSL and NM MLSL, which can be easily removed. When compared with the description of MLSL, we notice that in NM MLSL we have:

1. the 'batchsize' N is equal to 1;
2. the value γ is equal to 1;
3. $d_1 = 0$ and d_2 is not a constant but it is equal to α_k.

Essentially, some of the MLSL parameters are fixed values in NM MLSL, but the description of NM MLSL can be modified in order to transform these fixed values into parameters. Now we are ready to prove that NM MLSL satisfies the same theoretical properties as MLSL, but under Assumption 2.

4. The theoretical properties

4.1. DETECTION OF THE MINIMUM SETS

We first need a lemma, introduced to deal with the difficulties at the boundary of X.

LEMMA 1. *Let X be convex, bounded and with non-empty interior. Then:*

$$\forall\, x \in X, \ \exists\, \beta > 0 :$$

$$\lim_{r \to 0} \frac{m(S(x,r) \cap X)}{m(S(x,r))} \geqslant \beta, \tag{7}$$

where $S(x,r)$ is the sphere of radius r and center x.

Proof. See the appendix, p. 39. □

In what follows we introduce an assumption which seems to be quite reasonable, also exploited in [10].

ASSUMPTION 3. *Given a minimum set A, $\exists\, \overline{\epsilon} > 0 : \ \forall\, x \in X$ with $d(x, A) \leqslant \overline{\epsilon}$, LS applied to x returns a representant of A.*

We also make use of the following proposition whose proof is trivial:

PROPOSITION 1. *The stationary sets are compact.*

We can then introduce the following theorem whose proof is inspired by that of Lemma 11 in [10] and Theorem 2 in [11].

THEOREM 1. *If $\alpha_k, \beta_k \to 0$, under Assumption 2, the probability that NM MLSL detects at least one representant of any minimum set in a finite time is equal to one.*

Proof. We use the following notation:

- A is a minimum set;
- $B(A, \epsilon) = \{y : 0 < d(y, A) \leqslant \epsilon\}$;
- $L(y) = \{x \in X : f(x) \leqslant y\}$;
- for $y \geqslant f(A)$, $L_A(y)$ is the connected component of $L(y)$ containing A;
- d_{min} is the minimum distance between distinct minimum sets.

We first show that $d_{min} > 0$. By contradiction, we assume $d_{min} = 0$. Then, since the number of minimum sets is finite, there exist distinct minimum sets A and A_1 with $d(A, A_1) = 0$. In view of Proposition 1, we must have $A \cap A_1 \neq \emptyset$ and $f(A) = f(A_1)$. Then $A \cup A_1$ is connected and is still a minimum set with $A, A_1 \subset A \cup A_1$, $f(A \cup A_1) = f(A) = f(A_1)$. But this is a contradiction, because A and A_1 must be maximal with respect to the inclusion.

Now we consider a value $\delta = \delta(A)$ such that $0 < \delta < \min\{d_{min}, \overline{\epsilon}\}$ ($\overline{\epsilon}$ is the same as in Assumption 3) and $\forall\, y \in B(A, \delta)$, $f(y) > f(A)$ (see the definition of minimum set). Let

$$M(A) = \min\left\{ f(y) \mid y \in X,\ \frac{\delta}{2} \leqslant d(y, A) \leqslant \delta \right\} > f(A);$$

$$f_C(A) = \frac{M(A) + f(A)}{2}.$$

We note that $L_A(f_C(A)) \subseteq B\left(A, \frac{\delta}{2}\right) \cup A$. Let

$$K_A := \min\left\{ k : \alpha_k < \frac{\delta}{2}, \quad \beta_k < f_C(A) - f(A) \right\},$$

and let $k \geqslant K_A$. Let us assume that we have never detected a representant of the minimum set A up to iteration k, and we sample a point in $L_A(f_C(A))$. Then, from (6) with $s = \alpha_k$ and $r = \beta_k$, it follows that we can only consider sequences of sampled points $\{x_i\}$ such that

$$f(x_i) \leqslant f_C(A) + \beta_k < M(A). \tag{8}$$

Since $\alpha_k < \frac{\delta}{2}$, we can not have

$$x_i \in B\left(A, \tfrac{\delta}{2}\right) \cup A \quad \text{and} \quad x_{i+1} \notin B(A, \delta) \cup A.$$

Therefore, in order to get out of $B\left(A, \frac{\delta}{2}\right)$, we have to pass across $B(A, \delta) \backslash (B\left(A, \frac{\delta}{2}\right) \cup A)$. But (8) says this is not possible. So the sequence can not get out of $B\left(A, \frac{\delta}{2}\right) \cup A$.

Therefore, if we sample a point in $L_A(f_C(A))$ and A has not been detected yet, we detect it in view of Assumption 3 and the choice of δ.
Let

$$h_A = m(L_A(f_C(A))).$$

Since f is continuous on a compact set, it is uniformly continuous. Then given $V(A) = f_C(A) - f(A)$, we have

$$\exists\, \eta = \eta(A) > 0 \;:\; \forall\, x, y \in X \;:\; d(x, y) \leqslant \eta \;=>\; |\, f(x) - f(y)\, | \leqslant V(A).$$

(Here we could also exploit lipschitzianity to give a sharper relation between η and V.) Then, for a given $z \in A$, we have

$$\forall\, y \in S(z, \eta), \quad f(y) \leqslant f_C(A),$$

i.e. $S(z, \eta) \subseteq L_A(f_C(A))$. In view of Lemma 1, for a small enough η, we have

$$h_A \geqslant \frac{\beta}{2} m(S(z, \eta)) > 0.$$

Now we consider

$$K = \max_{\text{A min.set}} K_A, \tag{9}$$

and

$$h = \min_{A \text{ min.set}} h_A, \tag{10}$$

where $h > 0$ and K is finite in view of the existence of a finite number of stationary sets (see point c of Assumption 2).

Then, given any minimum set A, for $k \geqslant K$, the probability of detecting A within the iteration k is at least

$$1 - (1 - h)^{k-K} \to 1,$$

as $k \to \infty$. □

4.2. NUMBER OF LOCAL SEARCHES

We first introduce two lemmas, which are particularly relevant to the proof of Theorem 2.

LEMMA 2. *Let Assumption 2 hold. Given $\overline{x} \in X$, $\exists$ a minimum set A, dependent on $\overline{x}$, such that $\forall\ w \in A$, there exists at least one sequence $\{y_i\}_{i=1}^{final}$ with the following characteristics:*

1. *the first point in the sequence is* $\overline{x}$*, i.e.* $y_1 = \overline{x}$;
2. *the sequence is nonincreasing, i.e.*

$$\forall\, i,\ i = 1, \ldots, final - 1: \quad f(y_{i+1}) \leqslant f(y_i);$$

3. *we have that*

$$\forall\, i,\ i = 2, \ldots, final - 1: \quad d(y_i, y_{i+1}) = \frac{\alpha_k}{2},$$

and

$$\forall\, i \neq j,\ i, j < final: \quad d(y_i, y_j) \geqslant \frac{\alpha_k}{2};$$

4. *the final point of the sequence is* w*, i.e.* $y_{final} = w$*, and is such that:*

$$d(y_{final}, y_{final-1}) \leqslant \frac{\alpha_k}{2};$$

5. *the length of the sequence is at most* $O(\frac{1}{\alpha_k^d})$.

Proof. See the appendix, p. 40. □

Of course, the order of magnitude that we have found for the length of the sequences $\{y_i\}$ is quite rough and it is probably possible to find better results, but these would not deeply affect the final results that we are trying to derive.

Now, let $\overline{x}$ and w be the same as in the previous lemma. Let $\gamma_k = \min\{\alpha_k, \beta_k\}$. We consider the family of sequences $F_k(\overline{x}, w)$ built in the following way:

$$\begin{aligned}
x_1 &= \overline{x} \\
x_2 &\in S\left(y_2, \tfrac{\delta}{2L}\gamma_k\right) \cap X \\
x_3 &\in S\left(y_3, \tfrac{\delta}{2L}\gamma_k\right) \cap X \\
&\vdots \qquad\qquad \vdots \\
x_n &\in S\left(y_n, \tfrac{\delta}{2L}\gamma_k\right) \cap X \\
&\vdots \qquad\qquad \vdots \\
x_{final} &\in S\left(y_{final}, \tfrac{\delta}{2L}\gamma_k\right) \cap X,
\end{aligned}$$

where, in view of the previous lemma

$$final \leqslant p(k) = O\left(\frac{1}{\alpha_k^d}\right), \tag{11}$$

i.e. $p(k)$ is an upper bound for the number of points in the sequences in $F_k(\overline{x}, w)$ and is independent from $\overline{x}$ and w. We can prove the following lemma:

LEMMA 3. *Let Assumption 2 hold. If* $\delta < \frac{1}{2}$*, the sequences* $\{x_i\}$ *belonging to* $F_k(\overline{x}, w)$ *have the following characteristics:*

1. $x_1 = \overline{x}$;
2. *any sequence in* $F_k(\overline{x}, w)$ *is a* (α_k, β_k)*-sequence;*
3. $d(x_{final}, w) \leqslant \frac{\alpha_k}{2}$.

Proof. See the appendix, p. 40. □

We first need an upper bound for the probability of starting a local search at iteration k. Only the case $\alpha_k, \beta_k \to 0$ will be considered. The less interesting case in which $\alpha_k \not\to 0$ or $\beta_k \not\to 0$ could easily be treated in an analogous way.

THEOREM 2. *Let Assumption 2 hold. Let* $\alpha_k, \beta_k \to 0$. *Then the probability of starting a local search in NM MLSL at iteration k has an upper bound which is asymptotic to*

$$\frac{P_1}{\alpha_k^{2d}} \exp\left\{-P_2(k-K)\gamma_k^d\right\} + (1-h)^{k-K}, \tag{12}$$

where h is defined in (10), K is an integer constant defined in (9) and $P_1, P_2 > 0$ *are constants.*

Proof. Let $k \geqslant K$. Let A be the same minimum set appearing in Lemma 2 with $\overline{x} = x_k$, the last sampled point. Finally, let us denote with

- NL_k, the event of not starting a local search at iteration k;
- $L_k = NL_k^C$ the complement of the event NL_k, i.e. the event of starting a local search at iteration k;
- E_k, the event that, at iteration k, for any $y \in A$, some of the sampled points $\{x_i\}_{i=K}^k$, i.e. the points sampled after iteration $K-1$, form at least one (α_k, β_k)-sequence $\{x_{j_i}\}_{i=1}^s$ whose final point has distance not greater than α_k from y, i.e. $d(x_{j_s}, y) \leqslant \alpha_k$;
- $E_k(y)$, the event that, at iteration k, some of the sampled points $\{x_i\}_{i=K}^k$ form at least one sequence belonging to the family $F_k(x_k, y)$, for a fixed point $y \in A$;
- D_k, the event that a representant of the minimum set A has already been detected before iteration k.

We note that

$$D_k \cap E_k \implies NL_k. \tag{13}$$

Indeed, D_k implies that a representant x of the minimum set A has been already detected. Since E_k implies that the points sampled after iteration $K-1$ form at least one (α_k, β_k)-sequence whose final point has distance not greater than α_k from x, then, from the definition of the algorithm, it follows that no local search is started.

Thus, from (13), we have

$$P[L_k] \leqslant P[E_k^C \cup D_k^C] \leqslant P[D_k^C] + P[E_k^C].$$

From the proof of Theorem 1, with h defined in (10), we have that

$$P[D_k^C] \leqslant (1-h)^{k-K},$$

which is the second term of the upper bound (12) and is independent from A. In order to derive the first term we need to find an upper bound for $P[E_k^C]$.

Since A is compact, we can find a set of points $w_r \in A$ such that

$$\forall\, x \in A, \quad \exists\, w_r : \quad d(w_r, x) \leqslant \frac{\alpha_k}{2}. \tag{14}$$

The cardinality $c(k)$ of this set of points has an upper bound $u(k)$, independent from A and such that $u(k) = O\left(\frac{1}{\alpha_k^d}\right)$.

We have that

$$\cap_{r=1}^{c(k)} E_k(w_r) \implies E_k. \tag{15}$$

Indeed, $\cap_{r=1}^{c(k)} E_k(w_r)$ means that for any $r,\ r = 1, \ldots, c(k)$, the points sampled after iteration $K-1$ form at least one sequence $s_r \in F_k(x_k, w_r)$, which, in view of Lemma 3 is a $(\alpha_k, \beta_k)-$sequence and has final point x_{final_r} such that

$$\forall\, r,\ r = 1, \ldots, c(k) : \quad d(x_{final_r}, w_r) \leqslant \frac{\alpha_k}{2}. \tag{16}$$

Therefore, from (14) and (16), the triangular inequality implies

$$\forall\, x \in A\ \ \exists\, r \ \text{ such that } \ d(x, x_{final_r}) \leqslant \alpha_k,$$

i.e. $\forall\ x \in A,\ \ \exists\ r$ such that the sequence s_r is a $(\alpha_k, \beta_k)-$sequence ending at distance not greater than α_k from x. But this means that E_k is true. Therefore, from (15) we have

$$P[E_k^C] \leqslant P[\cup_{r=1}^{c(k)} E_k^C(w_r)] \leqslant \sum_{r=1}^{c(k)} (P[E_k^C(w_r)]). \tag{17}$$

Now, let us consider the sequence $\{y_i\}$ defined in Lemma 2 for the case $\overline{x} = x_k$ and $w = w_r$. Then, we consider the following events:

$$M_i = \exists\, x_j,\ j \geqslant K : \quad x_j \in S\left(y_i, \frac{\delta}{2L}\gamma_k\right), \quad i = 1, \ldots, final.$$

If $\cap_{i=1}^{final} M_i$ is true, then the sampled points $\{x_i\}_{i=K}^{k}$ form a sequence belonging to $F_k(x_k, w_r)$. But that also implies that $E_k(w_r)$ is true. Therefore

$$\cap_{i=1}^{final} M_i \implies E_k(w_r),$$

and then

$$P[E_k^C(w_r)] \leqslant P[\cup\, M_i^C] \leqslant \sum_{i=1}^{final} P[M_i^C], \tag{18}$$

Let

$$m_k = \frac{m\left(S\left(x, \frac{\delta}{2L}\gamma_k\right)\right)}{m(X)}.$$

Remembering Lemma 1, we have that

$$P[M_i^C] \leqslant (1 - \beta m_k)^{k-K},$$

which, for any i, is independent from y_i. Therefore, by also replacing $final$ in (18) with its upper bound $p(k)$ from (11), we have that

$$P[E_k^C(w_r)] \leqslant p(k)(1 - \beta m_k)^{k-K}, \tag{19}$$

where the term on the right is independent from x_k and w_r.

From (17) and (19), we obtain

$$P[E_k^C] \leqslant \sum_{r=1}^{c(k)} \sum_{i=1}^{p(k)} (1 - \beta m_k)^{k-K} \leqslant u(k)p(k)[1 - \beta m_k]^{k-K},$$

where the term on the right is independent from the last sampled point x_k, and, since $m_k \to 0$, it is asymptotic to

$$u(k)p(k)exp\{-\beta(k - K)m_k\}.$$

Since $u(k)$, $p(k) = O\left(\frac{1}{\alpha_k^d}\right)$ and $m_k = P_0\gamma_k^d$ for some constant $P_0 > 0$, constants P_1, $P_2 > 0$ must exist such that $P[E_k^C]$ has an upper bound asymptotic to

$$\frac{P_1}{\alpha_k^{2d}} \exp\{-P_2(k - K)\gamma_k^d\},$$

as we wanted to prove. □

Now we are ready for the final theorem:

THEOREM 3. *Let Assumption 2 hold. If*

$$\alpha_k := \left(\sigma \frac{\log k}{k}\right)^{1/d} \qquad \beta_k \propto \alpha_k, \tag{20}$$

and σ is big enough, the expected number of local searches in NM MLSL is finite.

Proof. We first note that the second term in the upper bound (12) for the probability of starting a local search at iteration k is the term of a converging series. Thus we only have to worry about the first term. From the definition of α_k and β_k given in (20), constants P_3 and P_4 must exist such that the first term in the upper bound (12) is equal to

$$\frac{P_1 k^2}{\sigma^2 \log^2 k} \exp\left\{-P_4\sigma \frac{k - K}{k} \log k\right\} = \frac{P_3 k^2}{\sigma^2(\log^2 k)k^{\sigma P_4 \frac{k-K}{k}}},$$

and we note that for a big enough σ this is the term of a converging series. □

5. Conclusion

In this paper, a well known algorithm, Multilevel Single Linkage, belonging to the class of Multistart algorithms, has been introduced. Its theoretical properties have been presented together with the assumptions under which they hold. Most of these assumptions have been relaxed and a new algorithm, Non-monotonic MLSL, has been presented, showing that it shares the same theoretical properties of MLSL but under a set of weaker assumptions. The main difference between MLSL and NM MLSL is that MLSL does not start a local search from a point if a *decreasing sequence* exists which goes close to an already detected minimum, while in NM MLSL *non-monotonic sequences* are also considered. The idea is inspired by non–monotonic algorithms for unconstrained local optimization in which some ascent steps are allowed and which have been proven to perform well, especially with ill-conditioned test functions (see [6]).

Acknowledgments

I would like to thank Prof. Fabio Schoen and the anonymous referees whose comments and suggestions have been an important help in improving the quality of the paper.

Appendix A. Proof of Lemma 1

For any point in the interior of X the limit is equal to 1. We only need to consider the points on the border ∂X of X. Let $y \in X \setminus \partial X$ and $\delta > 0$ such that $S(y, \delta) \subseteq X \setminus \partial X$. Let $x \in \partial X$. We take the segment s from x to y and the hyperplane which divides the sphere in two equal semispheres and is orthogonal to s. The intersection between the hyperplane and the sphere defines a $(d-1)$-dimensional sphere $S' = S'(x)$ whose radius is δ. We take the convex envelope of $S' \cup \{x\}$ and obtain the set $C(x) \subseteq X$ because of the convexity of X. The distance $l = l(x)$ between x and S' can be bounded from above by the finite diameter of X and from below by δ, and the bounds are independent from x. For any $r \leqslant l$ we have that $\exists\, \beta > 0$:

$$\frac{m(S(x,r) \cap X)}{m(S(x,r))} \geqslant \frac{m(S(x,r) \cap C(x))}{m(S(x,r))} \geqslant \beta > 0,$$

and from this the result of the lemma follows.

Appendix B. Proof of Lemma 2

First, we note that $\forall\, \overline{x} \in X$, $\exists$ a continuous curve $T = \{T(t) :\ t \in [0,1]\} \subseteq X$, such that

- $T(0) = \overline{x}$;
- $T(1) = w \in A$, where A is a minimum set;
- $f(T(t))$ is a non increasing function for $t \in [0,1]$.

Notice that w is a generic point belonging to A.

Then we exploit the following scheme in order to build a sequence with the characteristics presented in the lemma:

1. let $y_1 = \overline{x}$ and set $i := 1$;
2. if $w \in S\left(y_i, \frac{\alpha_k}{2}\right)$, then set $y_{i+1} = w$ and STOP, otherwise go to 3;
3. set $y_{i+1} = T(t_i)$, where
$$t_i = \max\left\{t :\ T(t) \in \partial S\left(y_i, \frac{\alpha_k}{2}\right)\right\};$$
4. set $i := i + 1$ and go back to 2.

Now we want to show that

$$\forall\, j > i+1,\ j < final: \quad d(y_i, y_j) \geqslant \frac{\alpha_k}{2}. \tag{21}$$

By contradiction, we assume that

$$\exists\, j > i+1,\ j < final: \quad d(y_i, y_j) < \frac{\alpha_k}{2}.$$

Notice that in this case we must have $T(t_{j-1}) \in S^\circ\left(y_i, \frac{\alpha_k}{2}\right)$, where S° denotes the interior of the sphere. Then, two cases are possible:

1. the piece of curve $\{T(t),\ t \in [t_{j-1}, 1]\}$ has an empty intersection with $\partial S\left(y_i, \frac{\alpha_k}{2}\right)$; then, $w \in S^\circ\left(y_i, \frac{\alpha_k}{2}\right)$; but this is not possible because in this case the sequence would stop at iteration i with $y_{i+1} = w$.
2. The piece of curve $\{T(t) :\ t \in [t_{j-1}, 1]\}$ crosses $\partial S\left(y_i, \frac{\alpha_k}{2}\right)$ at a point $\overline{y} = T(\overline{t})$ with $\overline{t} > t_{j-1} \geqslant t_i$, but this too is not possible in view of the definition of t_i.

Therefore, (21) must hold and the number of y_i cannot be greater than the maximum number of points that we can place in X whose distance between each other is not lower than $\frac{\alpha_k}{2}$. Since X is bounded, this number is at most $O(\frac{1}{\alpha_k^d})$.

Appendix C. Proof of Lemma 3

The fact that $x_1 = \overline{x}$ is true by definition. The third characteristic, i.e. $d(x_{final}, w) \leqslant \frac{\alpha_k}{2}$ is true because

$$d(x_{final}, y_{final}) \leqslant \frac{\delta}{2L}\alpha_k,$$

$y_{final} = w$, $\delta < \frac{1}{2}$ and $L \geqslant 1$. About the second characteristic, we first have to show that $\forall\ i : d(x_i, x_{i+1}) \leqslant \alpha_k$. This is true in view of the triangular inequality. Indeed

$$d(x_i, y_i) \leqslant \frac{\delta}{2L}\alpha_k, \quad d(y_i, y_{i+1}) \leqslant \frac{\alpha_k}{2}, \quad d(x_{i+1}, y_{i+1}) \leqslant \frac{\delta}{2L}\alpha_k$$
$$\Downarrow$$
$$d(x_i, x_{i+1}) \leqslant \alpha_k.$$

Then, we have to show that $\forall\ i$

$$f(x_{i+1}) \leqslant \min_{j \leqslant i} f(x_j) + \beta_k. \tag{22}$$

In view of the lipschitzianity of f, we have that $\forall\ j \leqslant i$:

$$f(x_j) \geqslant f(y_j) - \frac{\delta}{2}\gamma_k,$$

and also that

$$f(x_{i+1}) \leqslant f(y_{i+1}) + \frac{\delta}{2}\gamma_k.$$

Since $\forall\ j \leqslant i :\ f(y_j) \leqslant f(y_{i+1})$, we must have that $\forall\ j \leqslant i$:

$$f(x_{i+1}) \leqslant f(y_{i+1}) + \frac{\delta}{2}\gamma_k \leqslant f(y_j) + \frac{\delta}{2}\gamma_k \leqslant f(x_j) + \delta\gamma_k \leqslant f(x_j) + \beta_k,$$

and then (22) holds.

References

1. Ali, M.N. and Storey, C. (1994), Topographical multilevel single linkage, *Journal of Global Optimization* 5, 349–358.
2. Boender, C.G.E. and Romeijn, H.E. (1995), Stochastic methods, in *Handbook of global optimization*, Kluwer Academic Publishers, pp. 829–869.
3. Byrd, R., Dert, C., Rinnooy Kan, A.H.G. and Schnabel, R. (1990), Concurrent stochastic methods for global optimization, *Mathematical Programming* 46, 1–29.
4. Horst, R. and Tuy, H. (1992), *Global Optimization-Deterministic Approaches*, 2nd edition, Springer Verlag.
5. Horst, R. and Pardalos, P. (eds.) (1995), *Handbook of global optimization*, Kluwer Academic Publishers.
6. Grippo, L., Lampariello, F. and Lucidi, S. (1990), A quasi-discrete Newton algorithm with a nonmonotone stabilization technique, *Journal of Optimization Theory and Applications* 64, 495–510.
7. Locatelli, M. and Schoen, F. (1995), Random Linkage: a family of acceptance/rejection algorithms for global optimization, submitted; *available for anonymous ftp at* **ftp-dsi.ing.unifi.it**, directory **pub/schoen/papers/randlink**.
8. Locatelli, M. (1996), A new algorithm for stochastic global optimization with local searches, Technical Report No. 155-96, Univ. di Milano, Dip. Scienze dell'Informazione.

9. Lucidi, S. and Piccioni, M. (1989), Random tunneling by means of acceptance-rejection sampling for global optimization, *Journal of Optimization Theory and Applications* 62, 255–277.
10. Rinnooy Kan, A.H.G. and Timmer, G. (1987a), Stochastic global optimization methods, Part i: clustering methods, *Mathematical Programming* 39, 27–56.
11. Rinnooy Kan, A.H.G. and Timmer, G. (1987b), Stochastic global optimization methods, Part ii: multi level methods, *Mathematical Programming* 39, 57–78.
12. Törn, A. (1977), Cluster analysis using seed points and density determined hyperspheres as an aid to global optimization, *IEEE Trans.Syst.Men and Cybernetics* 7, 610–616.
13. Törn, A. (1978), A search-clustering approach to global optimization, in *Towards Global Optimization* 2, 49–62, North-Holland Publishing Company.
14. Törn, A. and Zilinskas, A. (1987), *Global optimization*, Springer Verlag.
15. Törn, A. and Viitanen, S. (1994), Topographical global optimization using pre-sampled points, *Journal of Global Optimization* 5, 267–276.
16. Zhigljavsky, A.A. (1991), *Theory of global random search*, Kluwer Academic Publishers.