On Timonov's Algorithm for Global Optimization of Univariate Lipschitz Functions

PIERRE HANSEN
RUTCOR, Rutgers University, U.S.A.

BRIGITTE JAUMARD
GERAD and École Polytechnique de Montréal, Canada.

SHI-HUI LU
RUTCOR, Rutgers University, U.S.A.

(Received: 3 May 1990; accepted: 5 October 1990)

Abstract. Timonov proposes an algorithm for global maximization of univariate Lipschitz functions in which successive evaluation points are chosen in order to ensure at each iteration a maximal expected reduction of the "region of indeterminacy", which contains all globally optimal points. It is shown that such an algorithm does not necessarily converge to a global optimum.

Key words. Global optimization, univariate function, Lipschitz function.

1. Introduction

We consider the global maximization of a univariate Lipschitz function f over an interval $[a, b]$:

$$\max_{x \in [a,b]} f(x) , \tag{1}$$

the function f satisfying the condition:

$$\forall x, y \in [a, b] \quad |f(x) - f(y)| \leq L|x - y| \tag{2}$$

where L is a constant.

Many algorithms have been proposed to solve (1). The best known of them is probably that of Piyavskii [4, 5] and Shubert [7]. A survey of these algorithms is given in Hansen, Jaumard and Lu [1]; a new one and comparative computational experiments are described in [2].

Piyavskii's algorithm uses and updates a piecewise-linear upper bounding function F which coincides with f at each evaluation point. The lines of F have slopes L or $-L$. Due to its shape, we call F a *saw-tooth cover* of f. Moreover, we call the part of F between two successive evaluation points x_i, x_{i+1} on $[a, b]$ a *tooth* of which x_i and x_{i+1} are the *basis points*. The highest point of this tooth, called its *peak*, is reached at the *peak point* x_{p_i}. At each iteration of Piyavskii's algorithm the highest point of F is considered and an evaluation of f takes place at the corresponding peak point. Then F is updated. The algorithm stops when the difference between the upper bound on f given by F and the value of the best known solution does not exceed a given positive real number ε. The rules of Piyavskii's algorithm may be expressed in pseudo-code as follows:

PIYAVSKII'S ALGORITHM

Initialization

$k \leftarrow 2$;

$x_1 \leftarrow a$;

$x_2 \leftarrow b$;

$x_{\text{opt}} \leftarrow x_1$;

$f_{\text{opt}} \leftarrow f(x_{\text{opt}})$;

if $f(x_2) < f_{\text{opt}}$ then $f_{\text{opt}} \leftarrow f(x_2)$; $x_{\text{opt}} \leftarrow x_2$ endif;

$$F_{\text{opt}} \leftarrow \frac{f(x_1) + f(x_2) + L(b-a)}{2} \;;$$

$$F_2 \leftarrow \min_{i=1,2} \{f(x_i) + L|x - x_i|\} \;;$$

Piyavskii's saw-tooth cover

While $F_{\text{opt}} - f_{\text{opt}} > \varepsilon$ do

$$x_{k+1} \leftarrow \arg \max_{x \in [a,b]} F_k(x) \;;$$

if $f(x_{k+1}) > f_{\text{opt}}$ then $f_{\text{opt}} \leftarrow f(x_{k+1})$; $x_{\text{opt}} \leftarrow x_{k+1}$ endif;

$$F_{k+1}(x) \leftarrow \min_{i=1,2,\ldots,k+1} \{f(x_i) + L|x - x_i|\} \;;$$

$$F_{\text{opt}} \leftarrow \max_{x \in [a,b]} F_{k+1}(x) \;;$$

$k \leftarrow k + 1$

EndWhile.

Timonov [9] proposes an algorithm for (1) which is very close to that of Piyavskii, but based on completely different rationale. Timonov focuses on the *region of indeterminacy* which is defined as follows. Consider a piecewise-linear upper-bounding funtion F and the best known value $f_{\text{opt}} = f(x_{\text{opt}})$ of f. Draw an horizontal line of height f_{opt} (see Figure 1). Clearly, no region of $[a, b]$ where $F < f_{\text{opt}}$ may contain a globally optimal point. The complementary set, i.e., the union of the subintervals of $[a, b]$ on which $F \geq f_{\text{opt}}$ constitutes the region of indeterminacy. Timonov proposes to choose at each iteration the new evaluation point which will ensure the maximal reduction in the expected length of the region of indeterminacy. Of course, to allow to make this choice, an assumption must be made on the distribution of possible values of points in $[a, b]$. Timonov

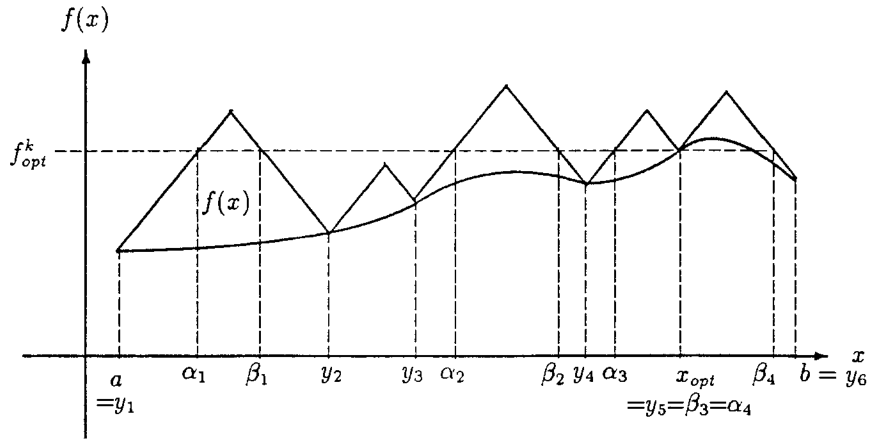

Fig. 1. Region of indeterminacy.

assumes that all values compatible with the Lipschitz condition (2) are equally probable. He states that the best choice is then to evaluate f at the peak point corresponding to the maximum of F. In the next Section, we show that this is not necessarily the case but that some peak point in the region of indeterminacy satisfies Timonov's criterion. A more serious difficulty is discussed in Section 3, where it is shown that the corrected version of Timonov's algorithm need not converge to a global optimum. Finally we discuss briefly in Section 4 a different modification to Piyavskii's algorithm also proposed by Timonov, which is both theoretically and practically an improvement: using evaluation points among those of a *passive strategy*, i.e., points of the set $\{a + (2p-1)\frac{\varepsilon}{L} \in [a, b], p \in N^+\}$.

2. Evaluation Points

Consider a saw-tooth cover of f obtained after k function evaluations at $x_1, x_2, \ldots, x_k$, which, after ranking, can be noted, $y_1 \leq y_2 \leq \cdots \leq y_k$. Let f^k_{opt} denote the best known solution, i.e. $f^k_{\text{opt}} = \max_{i=1,2,\ldots,k} f(x_i)$. The horizontal line of height f^k_{opt} intersects at most $k-1$ teeth, leading to $p \leq k-1$ intervals $[\alpha_j, \beta_j]$ (with possibly $\alpha_{j+1} = \beta_j$) where a globally optimal point may be located. The total length of the so defined "region of indeterminacy" is $l(ri^k) = \Sigma^p_{j=1} (\beta_j - \alpha_j)$ (see Figure 1).

Following Timonov's criterion, the next evaluation point x_{k+1} should be chosen in order to minimize the expected value of $l(ri^{k+1})$. This implies that an assumption is made on the distribution of possible values of f at any point x_{k+1} in $\cup_{j=1,2,\ldots,p} [\alpha_j, \beta_j]$. (Choosing x_{k+1} elsewhere cannot improve f_{opt}.) Timonov assumes a uniform distribution over the set of values compatible with the Lipschitz property of f. Thus the set of possible values for f at $x_{k+1} \in [y_i, y_{i+1}]$, where y_i and y_{i+1} are the evaluation points closest to x_{k+1} on the left and on the

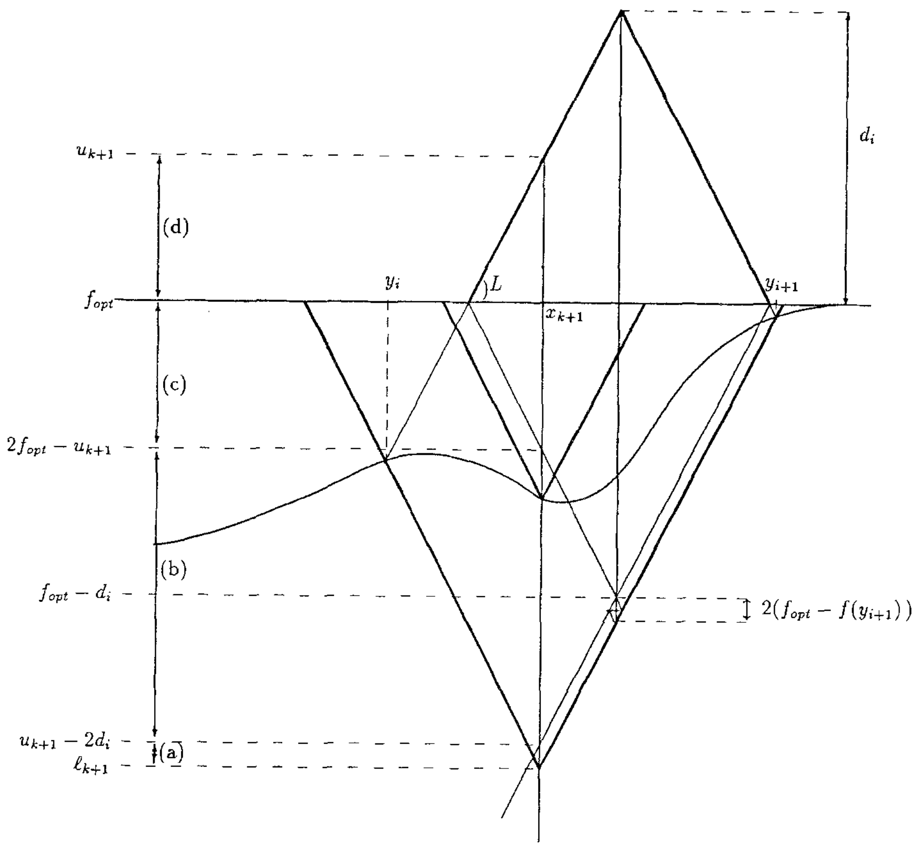

Fig. 2. Cases for reduction of an interval of indeterminacy.

right respectively, is $[l_{k+1}, u_{k+1}] = [\max\{f(y_i) - L(x_{k+1} - y_i),\ f(y_{i+1}) + L(x_{k+1} - y_{i+1})\},\ \min\{f(y_i) + L(x_{k+1} - y_i),\ f(y_{i+1}) - L(x_{k+1} - y_{i+1})\}]$ and the probability density is uniform over that interval (see Figure 2).

Timonov proves that the first three evaluation points should be at $x_1 = a$, $x_2 = b$ and $x_3 = (a + b)/2 + ((f(b) - f(a))/2L)$, i.e. at the basis points and the peak point of the first tooth. He then states that: (i) evaluation points should be peak points of teeth; (ii) the peak point of the highest tooth of the current saw-tooth cover is optimal for his criterion. This leads him in fact to advocate the use of Piyavskii's algorithm. To justify statements (i) and (ii), Timonov considers separately each interval of the region of indeterminacy. This, however, neglects the fact that if the function value at the new evaluation point is better than the incumbent, *all* intervals of the region of indeterminacy are reduced. We next derive an expression of the expected reduction in total length of the region of indeterminacy, when f is evaluated at a given point x_{k+1}. Using this expression, we then show that statement (i) is correct, but statement (ii) is not.

THEOREM 1. *Consider a saw-tooth cover of f over $[a, b]$ with p teeth of height $h_j = f_{\text{opt}} + d_j > f_{\text{opt}}$, $j = 1, 2, \ldots, p$, defining the region of indeterminacy $\bigcup_{j=1}^{p} [\alpha_j, \beta_j]$. If f is evaluated at x_{k+1}, with $x_{k+1} \in [\alpha_i, \beta_i]$ for some $i \in \{1, 2, \ldots, p\}$ and with possible values in $[l_{k+1}, u_{k+1}]$ the expected reduction in total length of the region of indeterminacy is:*

$$\Delta = \frac{1}{L(u_{k+1} - l_{k+1})} \left[2 \min\left\{ d_i, \frac{u_{k+1} - l_{k+1}}{2} \right\} \left(u_{k+1} - l_{k+1} \right.\right.$$
$$\left. - \min\left\{ d_i, \frac{u_{k+1} - l_{k+1}}{2} \right\} \right) + \sum_{j \neq i} \min\{u_{k+1} - f_{\text{opt}}, d_j\}(2(u_{k+1} - f_{\text{opt}})$$
$$\left. - \min\{u_{k+1} - f_{\text{opt}}, d_j\}) \right]. \tag{3}$$

Proof. We first consider the reduction in the interval of indeterminacy $[\alpha_i, \beta_i]$ for $i \in \{1, 2, \ldots, p\}$ in which x_{k+1} is chosen. Four cases must be considered:

(a) $l_{k+1} \leq f(x_{k+1}) \leq \max\{l_{k+1}, u_{k+1} - 2d_i\}$; the whole interval, of length $A_i = 2d_i/L$, is eliminated. (Note that this case only occurs when $l_{k+1} < u_{k+1} - 2d_i$.)
(b) $\max\{l_{k+1}, u_{k+1} - 2d_i\} \leq f(x_{k+1}) \leq 2f_{\text{opt}} - u_{k+1}$; the length of the interval is reduced by $B_i = (u_{k+1} - f(x_{k+1}))/L$.
(c) $2f_{\text{opt}} - u_{k+1} \leq f(x_{k+1}) \leq f_{\text{opt}}$; the length of the interval is reduced by $C_i = 2(f_{\text{opt}} - f(x_{k+1}))/L$.
(d) $f_{\text{opt}} \leq f(x_{k+1}) \leq u_{k+1}$; the length of the interval is reduced by $D_i = 2(f(x_{k+1}) - f_{\text{opt}})/L$.

The four cases are illustrated on Figure 2. The reduction in length of the intervals $[\alpha_j, \beta_j]$, $j = 1, 2, \ldots, p$, $j \neq i$, is zero in cases (a), (b) and (c) and, two subcases arise in case (d):

(d1) $f_{\text{opt}} \leq f(x_{k+1}) \leq f_{\text{opt}} + d_j$; the length of the interval $[\alpha_j, \beta_j]$ is reduced by $D_j^1 = 2(f(x_{k+1}) - f_{\text{opt}})/L$.
(d2) $f_{\text{opt}} + d_j \leq f(x_{k+1}) \leq u_{k+1}$; the whole interval, of length $D_j^2 = 2d_j/L$, is eliminated.

The expected reduction in total length is:

$$\Delta = \frac{1}{L(u_{k+1} - l_{k+1})} \left\{ \int_{l_{k+1}}^{\max(l_{k+1}, u_{k+1} - 2d_i)} A_i \, \mathrm{d}f + \int_{\max(l_{k+1},\, u_{k+1} - 2d_i)}^{2f_{\text{opt}} - u_{k+1}} B_i \, \mathrm{d}f \right.$$
$$+ \int_{2f_{\text{opt}} - u_{k+1}}^{f_{\text{opt}}} C_i \, \mathrm{d}f + \int_{f_{\text{opt}}}^{u_{k+1}} D_i \, \mathrm{d}f$$
$$\left. + \sum_{j, j \neq i} \left(\int_{f_{\text{opt}}}^{\min(f_{\text{opt}} + d_j,\, u_{k+1})} D_j^1 \, \mathrm{d}f + \int_{\min(f_{\text{opt}} + d_j,\, u_{k+1})}^{u_{k+1}} D_j^2 \, \mathrm{d}f \right) \right\}.$$

Integrating this expression and simplifying it leads to (3). □

COROLLARY 1. *If x_{k+1} is chosen in the interval $[\alpha_i, \beta_i]$, Δ reaches its maximum when the new evaluation point corresponds to the peak point, i.e. $x_{k+1} = x_{p_i}$.*

Proof. Let $u_{k+1} = f_{\text{opt}} + d$, where d is a non negative parameter with maximal value d_i, reached when x_{k+1} coincides with the peak point x_{p_i}. Then $l_{k+1} = f_{\text{opt}} - d - e_i$ where e_i is a non negative constant. Substituting u_{k+1} and l_{k+1} by these values in (3) leads to:

$$\Delta = \Delta_1 + \Delta_2 ,$$

where:

$$\Delta_1 = \frac{1}{L(2d + e_i)} \left[2 \min\left\{d_i, d + \frac{e_i}{2}\right\}\left(2d + e_i - \min\left\{d_i, d + \frac{e_i}{2}\right\}\right)\right], \tag{4}$$

and

$$\Delta_2 = \frac{1}{L(2d + e_i)} \sum_{j \neq i} \min\{d, d_j\}(2d - \min\{d, d_j\}) . \tag{5}$$

If $\min\{d_i, d + e_i/2\} = d_i$, then $\Delta_1 = (1/L)(2d_i - (2d_i^2)/(2d + e_i))$ and otherwise $\Delta_1 = (1/L)(d + e_i/2)$. If $\min\{d, d_j\} = d_j$, the jth term in the summation of Δ_2 is equal to $(1/L)(d_j(2d - d_j)/(2d + e_i))$, and otherwise to $(1/L)(d^2/(2d + e_i))$. Partial derivatives with respect to d of Δ_1 and Δ_2 are then easily shown to be positive in all cases. Hence, Δ increases in d, from which the corollary follows. □

COROLLARY 2. *If x_{k+1} is chosen in $\cup_{i=1,2,\ldots,p} [\alpha_i, \beta_i]$, Δ reaches its maximum when the new evaluation point corresponds to the peak point x_{p_i} of a tooth, which is not necessarily the highest one. Hence:*

$$\Delta_{\max} = \max_i \left\{ \frac{2d_i}{L} - \frac{2d_i^2}{L(2d_i + e_i)} + \frac{1}{L(2d_i + e_i)} \sum_{j \neq i} \min\{d_i, d_j\} \right.$$
$$\left. \times (2d_i - \min\{d_i, d_j\}) \right\} \tag{6}$$

where $e_i = 2(f_{\text{opt}} - \max\{f(y_l), f(y_{l+1})\})$, y_l and y_{l+1} being the basis points of the tooth whose intersection with the horizontal line of height f_{opt} is $[\alpha_i, \beta_i]$.

Proof. The first proposition is a direct consequence of Corollary 1. The expression (6) is obtained from (4) and (5) with $d = d_i$.

The following example shows that the evaluation point which maximizes Δ is not always the peak of the highest tooth of the saw-tooth cover. Consider a Lipschitz function f with $L = 1$ defined over the interval $[-8, 8]$. Assume $f(-8) = f(0) = f(4) = f(8) = 0$ and $f(-5.4) = f(-4) = f(-2.6) = 1.2$. Timonov's algorithm chooses the first seven evaluation points at $x_1 = -8$, $x_2 = 8$, $x_3 = 0$, $x_4 = -4$, $x_5 = 4$, $x_6 = -5.4$ and $x_7 = -2.6$. The saw-tooth cover after seven iterations is $F_7(x) = \min_{i=1,2,\ldots,7} \{f(x_i) + |x - x_i|\}$, represented on Figure 3. The next evaluation point would be at $x_8 = 2$ which gives $\Delta = 2.07$. However, choosing $x_8 = -6.1$ leads to $\Delta = 2.45$. □

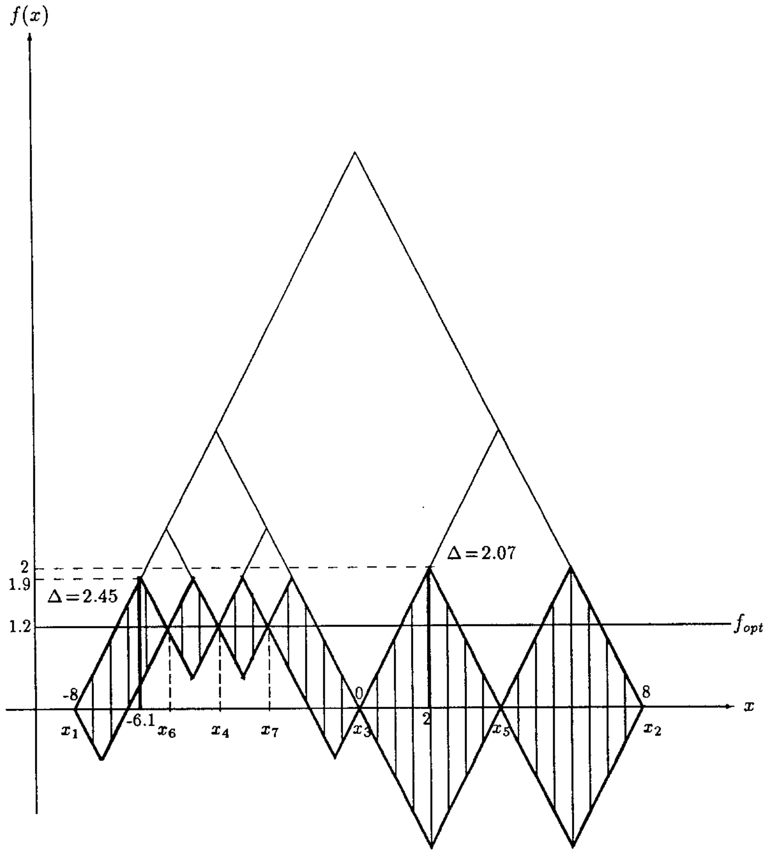

Fig. 3. The peak point where Δ is maximum is not the highest one.

3. Convergence

As shown by Corollary 2, if Timonov's criterion is adopted, one should choose as new evaluation point at each iteration the peak point of the tooth for which Δ is maximum, i.e. equal to the value Δ_{max} of (6). Such a strategy, however, may not be convergent as we now show. Consider the following example:

Let:

$$f(x) = \begin{cases} 0 & x \in [0, 18] \\ 18 - x & x \in [18, 24] \\ x - 30 & x \in [24, 31] \\ 32 - x & x \in [31, 38]\,. \end{cases}$$

The first eleven evaluation points coincide with those of Piyavskii's algorithm, i.e., they are at $x_1 = 0$, $x_2 = 38$, $x_3 = 16$, $x_4 = 8$, $x_5 = 24$, $x_6 = 4$, $x_7 = 12$, $x_8 = 2$,

$x_9 = 6$, $x_{10} = 10$ and $x_{11} = 14$. The region of indeterminacy is $[0, 18] \cup [30, 32]$ and all the ten teeth have the same height $h_i = 1$ (see Figure 4). Let us call 0th tooth the tooth with basis points at $x_5 = 24$ and $x_2 = 38$. We claim that this tooth will never be split. Hence the globally optimal value, which is equal to 1 for $x = 31$, will always remain greater by 1 than the incumbent. Indeed, no function evaluation over the interval $[0, 18]$ will change the value of f_{opt}, i.e. $f_{\text{opt}} = 0$. We have $d_0 = 1$ and $e_0 = 6$. Let $d_i = h_i - f_{\text{opt}} = h_i$ denote the height of the highest tooth over $[0, 18]$ at any iteration. We prove the claim by showing that $\Delta_i - \Delta_0 > 0$. We always have $d_0 \geq d_i \geq d_j$ for all $j = 1, 2, \ldots, k$ at any iteration k with $k > 11$. Notice that splitting a tooth different from the 0th one yields two teeth of half its height. Hence $\Sigma_{j=1}^{k} d_j = 9$ for all $k > 11$. We have:

$$\Delta_i - \Delta_0 = \left[d_i + \frac{1}{2d_i} \sum_{j \neq i,\, j=0}^{k} d_j(2d_i - d_j) \right] - \left[\frac{7}{4} + \frac{1}{8} \sum_{j=1}^{k} d_j(2 - d_j) \right] > \frac{1}{2} \, .$$

In the case where the corrected version of Timonov's algorithm does converge, the number of function evaluations will be very close to that of Piyavskii's

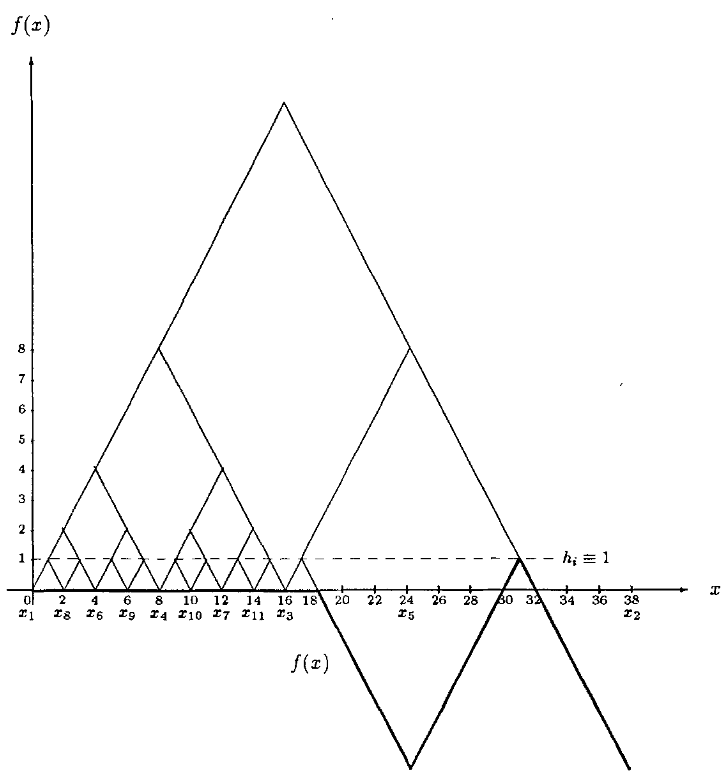

Fig. 4. Convergence to a local optimum.

algorithm. Indeed, both algorithms choose peak points of teeth as new evaluation points and even if the tooth selected following Timonov's criterion is not the highest, it will eventually be split by Piyavskii's algorithm also at a further iteration. Empirical results show that in most cases, the numbers of function evaluations do not differ by more than one even for small ε.

4. Evaluation Points from a Passive Strategy

Timonov also proposes to choose all evaluation points among those of a passive strategy, i.e., in the set $\{a + (2p-1)\frac{\varepsilon}{L} \in [a, b],\ p \in N^+\}$. (A similar suggestion was made independently by Schoen [6]). Evaluating f at all such points guarantees to obtain a point with an ε-optimal value for any f satisfying (2) and the number of function evaluations to do that is minimum (Ivanov [3], Sukharev [8]). The only modificaiton to be made in Piyavskii's algorithm to implement Timonov's idea consists in replacing the instruction

$$x_{k+1} = \arg \max_{x \in [a,b]} (F_k(x)) , \tag{7}$$

by a subroutine to find the point of a passive strategy which maximizes $F_k(x)$. This can be done by finding the points from a passive strategy closest to the peaks of $F_k(x)$ in all intervals in the region of indeterminacy. A quicker procedure, yielding very similar results, is just to consider the points of a passive strategy closest to x_{k+1} on the left and on the right and keep as new value for x_{k+1} that one of them with highest value.

Timonov's algorithm coincides with the passive algorithm in the worst case, i.e., for constant functions, and is then better than Piyavskii's algorithm. In the vicinity of a very flat global optimum, or if high precision is required, most teeth of Timonov's algorithm will have a height close to ε, instead of one between ε and $\varepsilon/2$ for those of Piyavskii's algorithm, and will hence be less numerous. Both algorithms, and others, are compared experimentally in [2]. It is found that Timonov's modification of Piyavskii's algorithm reduces the number of function evaluations by about 6%.

Acknowledgements

Research of the first and the third authors has been supported by AFOSR grants 0271 and 0066 to Rutgers University. Research of the second author has been supported by NSERC grant GP0036426 and FCAR grant 89Q4144. We thank N. Paradis for drawing some of the figures.

References

1. Hansen, P., B. Jaumard, and S. H. Lu (to appear), Global Optimization of Univariate Lipschitz Functions: Part I. Survey and Properties, *Mathematical Programming*.

2. Hansen, P., B. Jaumard and S. H. Lu (to appear), Global Optimization of Univariate Lipschitz Functions: Part II. New Algorithms and Computational Comparison, *Mathematical Programming*.
3. Ivanov, V. V. (1972), On Optimal Algorithms of Minimization in the Class of Functions with the Lipschitz Condition, *Information Processing* **71**, 1324–1327.
4. Piyavskii, S. A. (1967), An Algorithm for Finding the Absolute Minimum of a Function, *Theory of Optimal Solutions* **2**, Kiev, IK AN USSR (in Russian), 13–24.
5. Piyavskii, S. A. (1972), An Algorithm for Finding the Absolute Extremum of a Function, *U.S.S.R. Comput. Maths. Math. Phys.* **12**, 57–67. (*Zh. Výchisl Mat. mat. Fiz.* **12** (1972) 888–896).
6. Schoen, F. (1982), On a Sequential Search Strategy in Global Optimization Problems, *Calcolo* **19**, 321–334.
7. Shubert, B. O. (1972), A Sequential Method Seeking the Global Maximum of a Function, *SIAM Journal on Numerical Analysis* **9**, 379–388.
8. Sukharev, A. G. (1972), Best Sequential Search Strategies for Finding an Extremum, *U.S.S.R. Comput. Maths. Math. Phys.* **12**, 39–59 (*Zh. Vychisl. Mat. mat. Fiz* **12** (1971) 35–50).
9. Timonov, L. N. (1977), An Algorithm for Search of a Global Extremum, *Engineering Cybernetics* **15**, 38–44.