Efficient Line Searching for Convex Functions

den Boef, E.; den Hertog, D.

Publication date:
2004

Citation for published version (APA):
den Boef, E., & den Hertog, D. (2004). *Efficient Line Searching for Convex Functions*. (CentER Discussion Paper; Vol. 2004-52). Operations research.

No. 2004–52

EFFICIENT LINE SEARCHING FOR CONVEX FUNCTIONS

By E. den Boef, D. den Hertog

May 2004

ISSN 0924-7815

Efficient Line Searching for Convex Functions

Edgar den Boef[1,2] Dick den Hertog[3]

[1]Philips Research Laboratories, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands
[2]Technische Universiteit Eindhoven, P.O. Box 513, 5600 MB Eindhoven, The Netherlands
[3]Department of Econometrics and OR, Tilburg University, P.O. Box 90153, 5000 LE Tilburg, The Netherlands
denboef@natlab.research.philips.com
D.denHertog@uvt.nl

May 12, 2004

Abstract

In this paper we propose two new line search methods for convex functions. These new methods exploit the convexity property of the function, contrary to existing methods. The first method is an improved version of the golden section method. For the second method it is proven that after two evaluations the objective gap is at least halved. The practical efficiency of the methods is shown by applying our methods to a real-life bus and buffer size optimization problem and to several classes of convex functions.

Keywords: convex optimization, golden section, line search.

1 Introduction

Line searching is an important step for many optimization methods. In practice both exact and approximate line search methods are used. Well-known line search methods are quadratic and cubic interpolation, the golden section method and backtracking. For an overview we refer to the book on optimization by Gill, Murray, and Wright [2].

The aim in such line search methods is to find a (near) optimal solution along a given direction using a minimal number of function evaluations. Especially in the case of black-box functions, where often time-consuming simulation runs, i.e., function evaluations, have to be done, it is desirable to perform as few function evaluations as possible.

Now suppose the (black-box) function is known to be convex (or concave). Then the function has exactly one optimum on a closed domain. This fact is used by above mentioned methods. However, convexity of a function gives more information. For example, if a function is convex then using the performed functions evaluations, an upper and lower bound can be constructed for the function values. This information can be used to obtain better information on the location of the optimum.

In this paper we will show how this convexity information can be used. We will show that existing methods propose new candidates which may a priori be detected as not optimal by using the information of the previous iterations and the convexity property. We will describe two new methods: the *improved golden section method* and the *triangle section method*. We will show that theoretically the improved golden section strategy is at least as good as the regular golden section

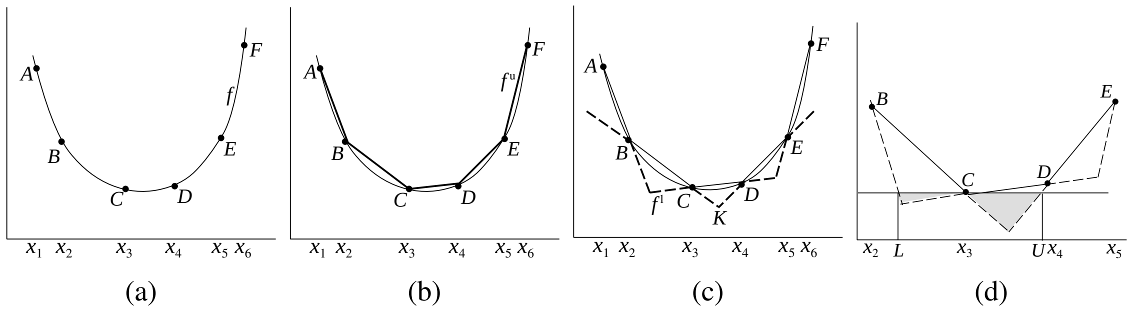


Figure 1. (a) Example of a convex function f with six function evaluations. (b) A piecewise-linear upper bound based on the convexity property. (c) A piecewise-linear lower bound based on the convexity property. (d) (Zoomed in compared with (a)–(c)) The optimum lies somewhere in the gray areas, the areas of uncertainty. The interval of uncertainty based on the convexity property is given by $[L,U]$.

strategy. For the triangle section method we will show that the range of uncertainty, i.e., the difference between the current best known objective value and the lower bound for the optimal value, at least halves after two function evaluations. It is significant that this result relates to the objective of the optimization, namely the function values, in contrast to the convergence result of the golden section method, which relates to the function domain values. We also describe the application of our methods to a real-life bus and buffer size optimization problem and to several classes of convex functions. We compare the resulting efficiency of our methods with the efficiency of the golden section method.

In Section 2 we show how convexity of a function can be used to reduce the interval in which the optimum can be. For a concave function similar methods can be used. We continue with deriving performance guarantees on the interval reduction in Section 3. In Section 4 we describe an application involving bus and buffer size optimization in which a non-differentiable function with computationally hard function evaluations has to be optimized. We describe experimental results for this application and for other classes of convex functions in Section 5. Finally, in Section 6 we give our conclusions.

2 Interval reduction using convexity

In this section we describe how to use convexity of a function to obtain upper and lower bounds for the function values. We show how they can be used to reduce the interval in the function domain in which the minimum can be, which is called the *interval of uncertainty*. For concave functions a similar method can be used.

Let $f(x)$ be a continuous, univariate, convex function on a closed domain D. Assume that for a given set of points $\mathcal{X} = \{x_1, \ldots, x_n\}$ in D the function values of f are known.

Figure 1(a) gives an example of a convex function f of which six function evaluations are known. As f is convex, $\alpha f(x_i) + (1-\alpha) f(x_j) \geq f(\alpha x_i + (1-\alpha)x_j)$ with $\alpha \in [0,1]$ for each $x_i, x_j \in \mathcal{X}$. Using this property we obtain the piecewise-linear upper bound f^{u} of f with $f^{\mathrm{u}}(x) \geq f(x)$ for all $x \in D$ and $f^{\mathrm{u}}(x) = f(x)$ if $x \in \mathcal{X}$; cf. Figure 1(b). Now we consider the line segments BC and DE and extend them until they intersect at point K, as shown in Figure 1(c). Then, the lines CK and KD give a lower bound for the function f between x_3 and x_4. This can be done for any four consecutive points, resulting in the lower bound f^{l} on f given in Figure 1(c) by the dashed line. Now let $x_k \in \mathcal{X}$ be a point with the lowest determined function evaluation, i.e., $f(x_k) \leq f(x)$ for

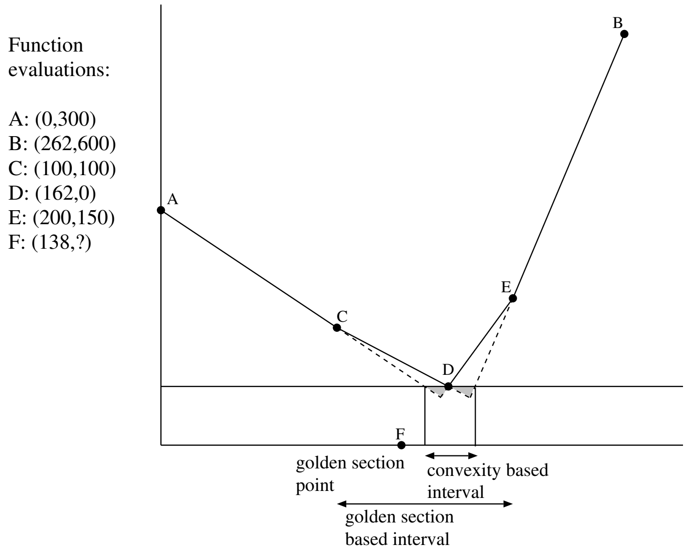


Figure 2. Example where the golden section method chooses a point outside the interval of uncertainty for function evaluation. Five function evaluations have already been made following the sequence $A,\ldots,E$. The sixth point for evaluation proposed by golden section is point F. However, the interval of uncertainty comprising the two small gray triangles is located to the right of F.

all $x \in \mathcal{X}$. Then the minimum function value of f must lie between $f(x_k)$ and the minimum of $f^1(x)$. The possible locations of the minimum are given in an enlarged view in Figure 1(d) by the gray areas, the *areas of uncertainty*. The interval of uncertainty is $[L,U]$.

This shows how the interval of uncertainty can be decreased using the convexity property. The next step in finding the minimum of f is choosing a point for evaluation. Naturally, this should be a point in the interval of uncertainty. Taking one of the existing methods to choose a point, however, does not always give a point in the interval of uncertainty. Figures 2, 3, and 4 show three examples where the golden section method, unit search, and quadratic interpolation, respectively, would evaluate the function at a point that is outside the interval of uncertainty. In Section 3 we discuss some strategies for choosing a new point. Finally, the interval of uncertainty may even be further decreased if for an evaluated point the gradient of the function is known. However, we leave this for future research.

3 Function evaluation strategies

We discuss in this section two strategies for choosing a new point. In Section 3.1 we show how to choose a new point such that the reduction of the interval of uncertainty is at least as large as the reduction which the golden section method would obtain. In Section 3.2 we show how to choose a new point such that the *range of uncertainty*, i.e., the interval consisting of all possible function values for the minimum, at least halves for each two new function evaluations. Finally, in Section 3.3 we shortly describe how known piecewise linearity of a function can be used to terminate the search procedure.

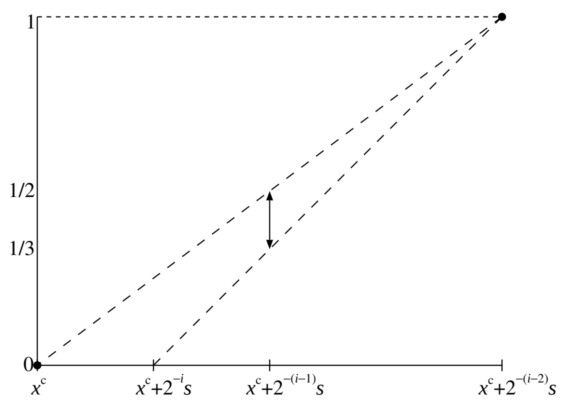


Figure 3. Example where unit search chooses a point outside the interval of uncertainty for function evaluation. The known function evaluations are $f(x^c) = 0$ and $f(x^c + 2^{-(i-2)}s) = 1$. Furthermore, we know $f(x^c + 2^{-(i-1)}s)$. For the next iteration, unit search would evaluate $x^c + 2^{-i}s$. However, if $\frac{1}{3} \leq f(x^c + 2^{-(i-1)}s) \leq \frac{1}{2}$ holds, then using convexity we know that $f(x^c + 2^{-i}s) \geq 0$, and therefore will not lead to a new minimum. Unit search can be improved by each time choosing the point in the middle of the interval of uncertainty based on convexity.

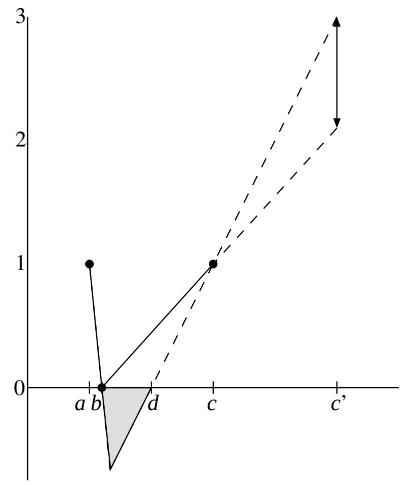


Figure 4. Example where quadratic interpolation chooses a point outside the interval of uncertainty for function evaluation. The points with known function evaluations are $a = 0$, $b = \frac{1}{10}$, and $c = 1$, with $f(a) = f(c) = 1$ and $f(b) = 0$. Furthermore, we know $f(c')$. Quadratic interpolation would take $d = \frac{1}{2}$ as the point for a new function evaluation. However, if $2\frac{1}{9} \leq f(c') \leq 3$ holds, then using convexity it is clear that $f(d) \geq 0$ and therefore will not lead to a new minimum.

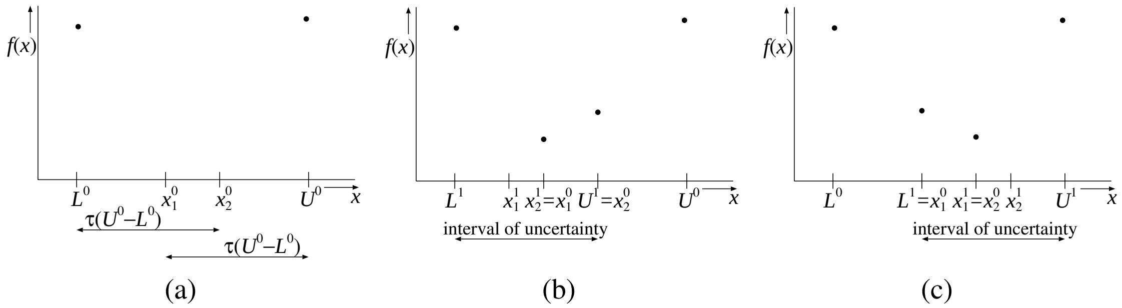


Figure 5. Starting the golden section method. (a) First two points, x_1^0 and x_2^0, are chosen in the interior of the starting interval of uncertainty, $[L^0, U^0]$. Next, the function is evaluated at these points, and depending on which one is lower, the interval of uncertainty is adjusted; see (b) and (c).

3.1 Function domain reduction

The strategy to reduce the interval of uncertainty in the function domain that we present in this paper, is basically the golden section method improved with the reduction that follows from the convexity property, as described in the previous section. The golden section method chooses new points for function evaluation in such a way that the interval of uncertainty can be decreased by a constant factor $\tau = (\sqrt{5}-1)/2$ in each iteration. Figure 5 shows an example. Let $[L^0, U^0]$ be the initial interval of uncertainty. Then golden section chooses the following two interior points x_1^0 and x_2^0 for evaluation: $x_1^0 = U^0 - \tau(U^0 - L^0)$ and $x_2^0 = L^0 + \tau(U^0 - L^0)$. Now, suppose x_1^0 has the lowest function evaluation. Then the new interval of uncertainty $[L^1, U^1]$ is equal to $[L^0, x_2^0]$. Furthermore, the new interior points to be evaluated are $x_1^1 = U^1 - \tau(U^1 - L^1)$ and $x_2^1 = L^1 + \tau(U^1 - L^1)$. However, because the interior points are chosen with the golden section factor τ, x_2^1 is the same point as x_1^0, and therefore, only x_1^1 has to be evaluated for the next step. Similarly, if x_2^0 has the lowest function evaluation, then $L^1 = x_1^0$ and $x_1^1 = x_2^0$.

The improved golden section method now works as follows. Initially, it is the same as the regular golden section method. Let $[L, U]$ be the interval of uncertainty with two interior points x_1 and x_2 such that $x_2 - L = U - x_1 = \tau(U - L)$. If $f(L) = \min\{f(L), f(x_1), f(x_2), f(U)\}$, then the new interval of uncertainty according to the golden section method is given by $[L, x_1]$. However, using the convexity property we can obtain a smaller interval of uncertainty $[L, U']$ with $U' \leq x_1$ for which two new interior points are selected. If $f(U) = \min\{f(L), f(x_1), f(x_2), f(U)\}$, then similarly a new interval $[L', U]$ can be obtained. Notice that by evaluating the function f first at the boundaries L and U, and then at the interior point closest to the boundary with the lowest function value, e.g. x_1 if $f(L) < f(U)$, the function does not need to be evaluated at the other interior point if the lowest function value is still at a boundary.

Now let the lowest function value be at x_1; if $f(x_2) < f(x_1)$, then a strategy analogous to what we describe here can be followed. The new interval of uncertainty using the golden section method is now given by $[L, x_2]$. Using the convexity property we can obtain a smaller interval of uncertainty $[L', U']$ for which $L' \geq L$ and $U' \leq x_2$ holds. For now we assume that x_1 is an interior point of $[L', U']$; the possibility that x_1 is not an interior point is handled later in this section.

Golden section would choose a new point $x_3 = x_2 - \tau(x_2 - L)$ so that $x_2 - x_3 = x_1 - L$. However, if we replace $[L, x_2]$ by $[L', U']$, and then choose $x_3 = U' - \tau(U' - L')$, $U' - x_3$ is generally not equal to $x_1 - L'$, meaning that the two interior points x_1 and x_3 do not satisfy the golden section property w.r.t. the interval of uncertainty $[L', U']$. Therefore, we will stretch the interval $[L', U']$ to a new interval $[\tilde{L}, \tilde{U}]$ such that the golden section property can be maintained for the new point to

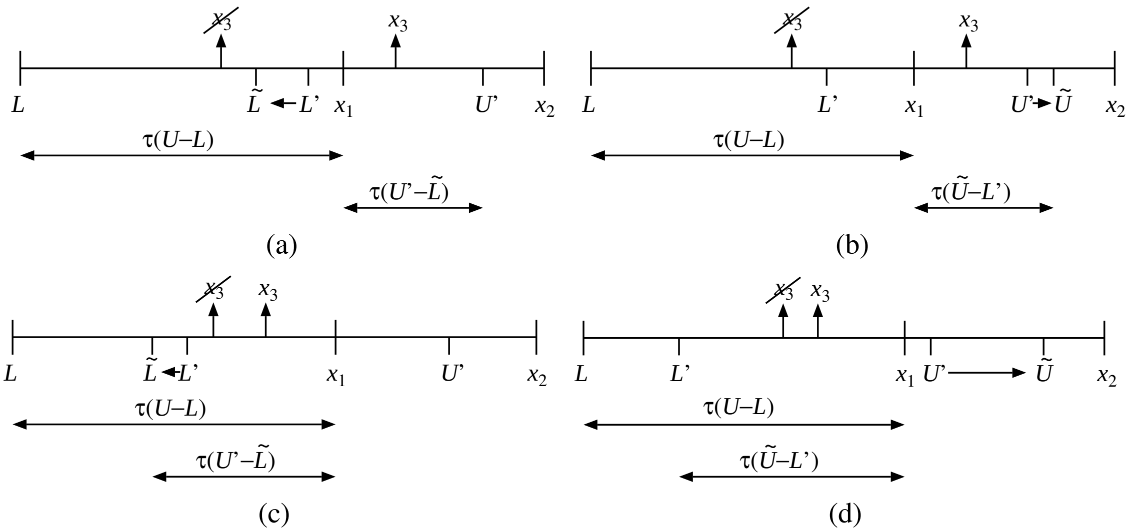


Figure 6. Stretching the interval of uncertainty obtained with the convexity property such that the golden section property is maintained for the new function evaluation. The four figures correspond to the four different possibilities.

evaluate. We distinguish four possibilities for this:

(a) $x_1 \leq U' - \tau(U' - L')$,

(b) $U' - \tau(U' - L') < x_1 < (U' + L')/2$,

(c) $(U' + L')/2 \leq x_1 < L' + \tau(U' - L')$,

(d) $x_1 \geq L' + \tau(U' - L')$.

Figure 6 shows an example of these four possibilities. The corresponding stretched intervals are the following:

(a) $\tilde{L} = U' - (U' - x_1)/\tau$, $\quad \tilde{U} = U'$,

(b) $\tilde{L} = L'$, $\quad \tilde{U} = (x_1 - \tau L')/(1 - \tau)$,

(c) $\tilde{L} = (x_1 - \tau U')/(1 - \tau)$, $\quad \tilde{U} = U'$,

(d) $\tilde{L} = L'$, $\quad \tilde{U} = L' + (x_1 - L')/\tau$.

The next lemma states that the obtained stretched interval is not larger than the interval of uncertainty obtained with the regular golden section method.

Lemma 1. $\tilde{U} - \tilde{L} \leq x_2 - L$.

Proof. For (a),(c), and (d) we prove that $\tilde{U} - \tilde{L} \leq x_2 - L$ holds by showing that $\tilde{L} \geq L$ and $\tilde{U} \leq x_2$. For (b) it is possible that $\tilde{U} > x_2$. In the following derivations we use the fact that $\frac{1-\tau}{\tau} = \tau$.

(a) As $\tilde{U} = U'$ it follows that $\tilde{U} \leq x_2$. For $\tilde{L}$ we can derive:

$$\begin{aligned}
\tilde{L} &= U' - (U' - x_1)/\tau \\
&= \left(1 - \frac{1}{\tau}\right) U' + \frac{1}{\tau}(L + \tau(x_2 - L)) \\
&= x_2 - \frac{1-\tau}{\tau}(U' - L) \\
&= x_2 - \tau(U' - L) \\
&\geq x_2 - \tau(x_2 - L) \\
&> x_2 - (x_2 - L)
\end{aligned}$$

$$= \quad L.$$

(b) From $x_1 \le \frac{1}{2}(L'+U')$ it follows that $\frac{1}{2}(x_1-L') \le \frac{1}{2}(U'-x_1)$, and therefore $x_1-L' \le U'-x_1 \le x_2-x_1$. We can now make the following derivation:

$$\begin{aligned}
\tilde{U}-\tilde{L} &= \frac{x_1-\tau L'}{1-\tau}-L' \\
&= \frac{x_1-L'}{1-\tau} \\
&\le \frac{x_2-x_1}{1-\tau} \\
&= \frac{x_2-(L+\tau(x_2-L))}{1-\tau} \\
&= x_2-L.
\end{aligned}$$

(c) As $\tilde{U}=U'$ it follows that $\tilde{U} \le x_2$. For $\tilde{L}$ we can derive:

$$\begin{aligned}
\tilde{L} &= (x_1-\tau U')/(1-\tau) \\
&= \frac{1}{1-\tau}(L+\tau(x_2-L)) - \frac{\tau}{1-\tau}U' \\
&= L+\frac{1}{\tau}(x_2-U') \\
&\ge L.
\end{aligned}$$

(d) As $\tilde{L}=L'$ it follows that $\tilde{L} \ge L$. For $\tilde{U}$ we can derive:

$$\begin{aligned}
\tilde{U} &= L'+(x_1-L')/\tau \\
&= \left(1-\frac{1}{\tau}\right)L'+\frac{1}{\tau}(L+\tau(x_2-L)) \\
&= x_2-\left(\frac{1-\tau}{\tau}\right)(L'-L) \\
&= x_2-\tau(L'-L) \\
&\le x_2.
\end{aligned}$$

□

This leads to the following strategy for choosing a new point for evaluation.

Improved golden section strategy, $f(x_1) \le f(x_2)$. Determine the stretched interval $[\tilde{L},\tilde{U}]$ as described above. Choose the new point x_3 for function evaluation as follows for the four previously distinguished possibilities.

(a) $x_1 \le U'-\tau(U'-L')$	$x_3 = \tilde{L}+\tau(\tilde{U}-\tilde{L}) = U'-\tau(U'-x_1)$.
(b) $U'-\tau(U'-L') < x_1 < \frac{1}{2}(L'+U')$	$x_3 = \tilde{L}+\tau(\tilde{U}-\tilde{L}) = \frac{1}{\tau}x_1-\tau L'$.
(c) $\frac{1}{2}(L'+U') \le x_1 < L'+\tau(U'-L')$	$x_3 = \tilde{U}-\tau(\tilde{U}-\tilde{L}) = \frac{1}{\tau}x_1-\tau U'$.
(d) $x_1 \ge L'+\tau(U'-L')$	$x_3 = \tilde{U}-\tau(\tilde{U}-\tilde{L}) = L'+\tau(x_1-L')$.

In the following theorem we show that the improved golden section strategy performs at least as good as the regular golden section method, while ensuring that the new point chosen for function evaluation lies in the interval of uncertainty.

Theorem 1. *Choosing a point according to the improved golden section strategy reduces the interval of uncertainty by at least a factor* τ*. Furthermore, the new point chosen for function evaluation lies in the interval of uncertainty, i.e.,* $x_3 \in [L', U']$.

Proof. Using the golden section method, the interval of uncertainty $[L, U]$ with function evaluations at $x_1 = U - \tau(U - L)$ and $x_2 = L + \tau(U - L)$ reduces to interval $[L, x_2]$ of size $\tau(U - L)$. The new point for function evaluation x_3 is then chosen such that the golden section property is maintained, i.e., $x_3 = x_2 - \tau(x_2 - L)$.

Using the improved golden section strategy, the new point for function evaluation is chosen either $x_3 = \tilde{U} - \tau(\tilde{U} - \tilde{L})$ or $x_3 = \tilde{L} + \tau(\tilde{U} - \tilde{L})$. Expressing the other internal point x_1 in $\tilde{L}$ and $\tilde{U}$ by rewriting the expressions for $\tilde{L}$ and $\tilde{U}$ gives $x_1 = \tilde{L} + \tau(\tilde{U} - \tilde{L})$ for (c) and (d), and $x_1 = \tilde{U} - \tau(\tilde{U} - \tilde{L})$ for (a) and (b). So for the improved golden section strategy the golden section property is maintained for the interval $[\tilde{L}, \tilde{U}]$. Lemma 1 states that $\tilde{U} - \tilde{L} \leq x_2 - L$ and thus $\tilde{U} - \tilde{L} \leq \tau(U - L)$. Therefore, the starting interval of uncertainty is reduced by at least a factor τ. Furthermore, as the new interval of uncertainty has the golden section property, the same reduction factor is guaranteed for next function evaluations.

Now we show that $x_3 \in [L', U']$. For (a), $x_3 = U' - \tau(U' - x_1)$, so $x_3 < U'$ is obvious. Furthermore, using $\tau < 1$ and $x_1 > L'$ it follows that $x_3 > L'$. Likewise, for (d) $x_3 \in [L', U']$ holds. For (b) we can make the following derivations:

$$
\begin{aligned}
x_3 &= \frac{1}{\tau}x_1 - \tau L' \\
&\leq \frac{L' + U'}{2\tau} - \tau L' \\
&= \frac{U'}{\tau} - \frac{(U' - L')}{2\tau} - \tau U' + \tau(U' - L') \\
&= U' - \frac{1 - 2\tau^2}{2\tau}(U' - L') \\
&< U'.
\end{aligned}
$$

$$
\begin{aligned}
x_3 &= \frac{1}{\tau}x_1 - \tau L' \\
&> \frac{1}{\tau}(U' - \tau(U' - L')) - \tau L' \\
&= \frac{1 - \tau}{\tau}U' + (1 - \tau)L' \\
&= \tau U' + (1 - \tau)L' \\
&> L'.
\end{aligned}
$$

In a symmetrical manner it can be shown for (c) that $x_3 \in [L', U']$. □

In case $f(x_2) < f(x_1)$ we have the following strategy:

Improved golden section strategy, $f(x_2) < f(x_1)$. Determine the stretched interval $[\tilde{L}, \tilde{U}]$. Choose the new point x_3 for function evaluation as follows.

(a) $x_2 \geq L' + \tau(U' - L')$	$x_3 = L' + \tau(x_2 - L')$.
(b) $\frac{1}{2}(L' + U') \leq x_2 < L' + \tau(U' - L')$	$x_3 = \frac{1}{\tau}x_2 - \tau U'$.
(c) $U' - \tau(U' - L') < x_2 < \frac{1}{2}(L' + U')$	$x_3 = \frac{1}{\tau}x_2 - \tau L'$.

(d) $x_2 \leq U' - \tau(U' - L')$ $\qquad\qquad$ $x_3 = U' - \tau(U' - x_2)$.

In an analogous way as above we can show that this reduces the interval of uncertainty by at least a factor τ.

Finally, we consider the possibility in which $[L', U']$, the interval of uncertainty obtained using the convexity property, does not have one of the previous interior points x_1, x_2 as an interior point. This can happen when $f(x_1) = f(x_2)$ or at least three consecutive points that have already been evaluated lie on one line; in the latter case the function f is at least partially piecewise linear. Then the interval $[L', U']$ lies between x_1 and x_2, i.e., $[L', U'] \subseteq [x_1, x_2]$ with either $L' = x_1$ or $U' = x_2$ or both. Two new function evaluations are now required to decrease the interval further by at least a factor τ. However, as $x_2 - x_1 = \tau^2(U - L)$, already a reduction of τ^2 has been obtained. Therefore, for any function evaluation the two-step reduction, for the new and last function evaluation, will be at least τ^2, giving an average reduction of at least factor τ for each function evaluation. Instead of immediately taking two new interior points for function evaluation, we choose one new point using the golden section property, i.e., either $x_3 = x_2 - \tau(x_2 - x_1)$ or $x_3 = x_1 + \tau(x_2 - x_1)$, which ensures that for following function evaluations a reduction of at least τ can be guaranteed. The resulting function evaluation of x_3 is used to update the interval of uncertainty and another point is chosen according to the described improved golden section strategy.

Now we have shown how the convexity property can be used to choose new points for evaluation such that the interval of uncertainty is reduced by at least the same factor as for the golden section method. However, in practice the reduction will be larger as we show with empirical results in Section 5. In this section we continue with a method that decreases the range of uncertainty by at least a factor 1/2 after two new function evaluations.

3.2 Function range reduction

As is shown in Figure 1(d) the convexity property can be used to obtain upper and lower bounds for the function value of each point in the interval of uncertainty. As these upper and lower bounds tighten for each new function evaluation, they can be used for a strategy that guarantees a reduction of the range of uncertainty instead of the interval of uncertainty. Figure 7 depicts the area in which the optimum can be, together with the points corresponding to the function evaluations and the interval of uncertainty.

Let M be the point with the lowest function evaluation so far, $f(M)$. Then $L' \leq M \leq U'$. Now we define Δf_1^k as the height of the triangle in the area of uncertainty between L' and M after k function evaluations, and Δf_2^k as the height of the triangle between M and U'. We can express Δf_1^k and Δf_2^k using known function values and the interval of uncertainty, as derived in Appendix A, which gives the following formulas:

$$\Delta f_1^k = \frac{(M - L')(f(L) - f(M))(f(U) - f(M))}{(f(L) - f(M))(U - M) + (f(U) - f(M))(L' - L)}, \tag{1}$$

$$\Delta f_2^k = \frac{(U' - M)(f(L) - f(M))(f(U) - f(M))}{(f(L) - f(M))(U - U') + (f(U) - f(M))(M - L)}. \tag{2}$$

The range of uncertainty is now given by the maximum height of the area of uncertainty, i.e., $\max\{\Delta f_1^k, \Delta f_2^k\}$. The point x we now choose for function evaluation lies in the middle of the area with the largest height, i.e.,

$$x = \begin{cases} \frac{1}{2}(L' + M) & \text{if} \quad \Delta f_1^k \geq \Delta f_2^k \\ \frac{1}{2}(M + U') & \text{if} \quad \Delta f_1^k < \Delta f_2^k \end{cases} \tag{3}$$

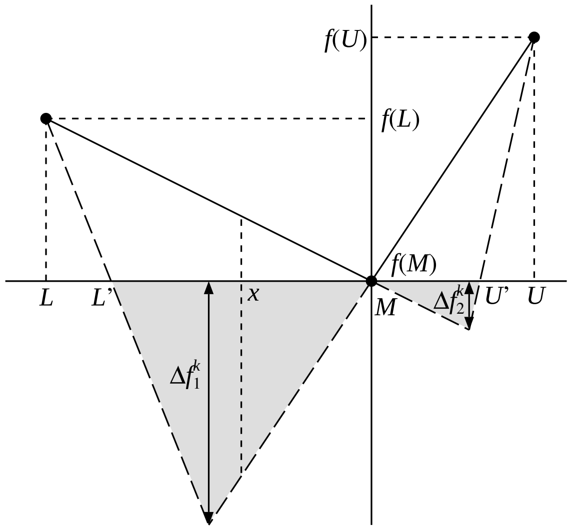


Figure 7. The areas of uncertainty. The three points with their function evaluations are given by L, M, and U. The interval of uncertainty in the function domain begins at L' and ends at U'. The height of the two areas after k function evaluations is given by Δf_1^k and Δf_2^k. The new point for function evaluation is x.

We refer to the method that chooses a new point for function evaluation according to (3) as the *triangle section method*. A greedy strategy is to take the point where the lower bound is minimal. However, it can be easily shown that the performance bounds for this greedy strategy are worse than the performance bounds for the triangle section strategy which we give in this paper.

In the remainder of this section we normalize, w.l.o.g., the function values and the size of the interval of uncertainty and the range of uncertainty in the following way:

$$M = 0, f(M) = 0, \Delta f_1^k = 1, L' = -1.$$

After substituting these values into the expression for Δf_1^k it follows that $f(U) = \frac{f(L)U}{1+L+f(L)}$, if $1+L+f(L) \neq 0$. The values of $L, f(L), U$, and U' then determine the exact situation. For the ease of notation we make the following substitutions, as shown in Figure 8:

$$\begin{aligned} A &= L \\ B &= f(L) \\ C &= U' \\ D &= U \\ E &= f(x) \end{aligned}$$

Furthermore, w.l.o.g. we assume that $\Delta f_1^k \geq \Delta f_2^k$. The new point for evaluation then is $x = -\frac{1}{2}$.

For the values of A, B, C, D, and E we can derive the following properties:

(i) The lower corner of the left area should be to the left of M, i.e., $-\frac{D}{f(D)} \leq 0$. As $f(D) = f(U) = \frac{f(L)U}{1+L+f(L)} = \frac{BD}{1+A+B}$, if $1+A+B \neq 0$, we have $-\frac{1+A+B}{B} \leq 0$ and $D \neq 0$. Since $B = f(L) > 0$, it follows that $1+A+B \geq 0$.

(ii) $L \leq L'$, i.e., $A \leq -1$ or $-1-A \geq 0$ or $1+A \leq 0$.

(iii) $U \geq U' \geq 0$, i.e., $D \geq C \geq 0$.

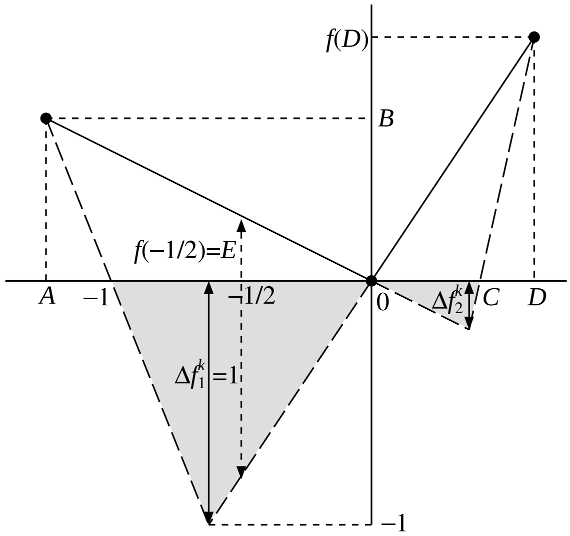


Figure 8. The areas of uncertainty with normalization of the function and interval values. The point 0 has the lowest function evaluation of 0. The size of the range of uncertainty is given by Δf_1^k, which is set to 1. The lower bound of the interval of uncertainty in the function domain is equal to -1, thus the new point for function evaluation is $-\frac{1}{2}$.

(iv) $\Delta f_2^k \leq \Delta f_1^k = 1$. Substitution of Δf_2^k gives $\Delta f_2^k = \frac{BCD}{(D-C)(1+A+B)-AD}$. So $BCD \leq (D-C)(1+A+B)-AD$.

(v) Let $f^{\mathrm{u}}(x)$ denote the upper bound for point x, i.e., the line $(A,B)-(0,0)$. Then $f(x) \leq f^{\mathrm{u}}(x)$ should hold, i.e., $E \leq -\frac{B}{2A}$ or $-2AE \leq B$.

(vi) Let $f^{\mathrm{l}}(x)$ denote the lower bound for point x, i.e., the line parts below the $y=0$-line of the lines $(A,B)-(-1,0)$ and $(0,0)-(D,f(D))$. Then $f(x) \geq f^{\mathrm{l}}(x)$ should hold. The value of $f^{\mathrm{l}}(x)$ depends on whether the lowest corner of the left area lies to the left or to the right of x, i.e., $f^{\mathrm{l}}(x) = \max\{-\frac{B}{2(1+A+B)}, -\frac{B}{2(-1-A)}\}$. This gives $B \geq -2E(1+A+B)$ and $B \geq -2E(-1-A)$. Notice that this also holds if $1+A+B=0$ or $-1-A=0$.

We now show in Lemma 2 that in the special case the area of uncertainty consists of one triangle, i.e., $\Delta f_1^k = 0$ or $\Delta f_2^k = 0$, the triangle section method at least halves the range of uncertainty for a new function evaluation.

Lemma 2. *Let a new function evaluation x_{k+1} be chosen according to (3), and let either $\Delta f_1^k = 0$ or $\Delta f_2^k = 0$. Then the range of uncertainty decreases by at least a factor $\frac{1}{2}$ after the function evaluation $f(x_{k+1})$ is known.*

Proof. W.l.o.g. we assume that $\Delta f_2^k = 0$. We distinguish four possibilities for $f(x_{k+1})$; cf. Figure 9(a).

1. $f(x_{k+1}) = f^{\mathrm{u}}(x_{k+1})$, see Figure 9(b). Then the upper bound is equal to the lower bound for all points in $[L,U]$ and $M=0$ is the optimum with function value $f(M)=0$. Thus $\Delta f_1^{k+1} = 0$.
2. $f^{\mathrm{u}}(x_{k+1}) > f(x_{k+1}) \geq 0$, see Figure 9(c). If we write Δf_1^{k+1} as an expression of A,B,C,D, and E we get

$$\Delta f_1^{k+1} = \frac{B(\frac{1}{2}B+AE)}{B(-\frac{1}{2}-A)+(B-E)(1+A+B)}. \tag{4}$$

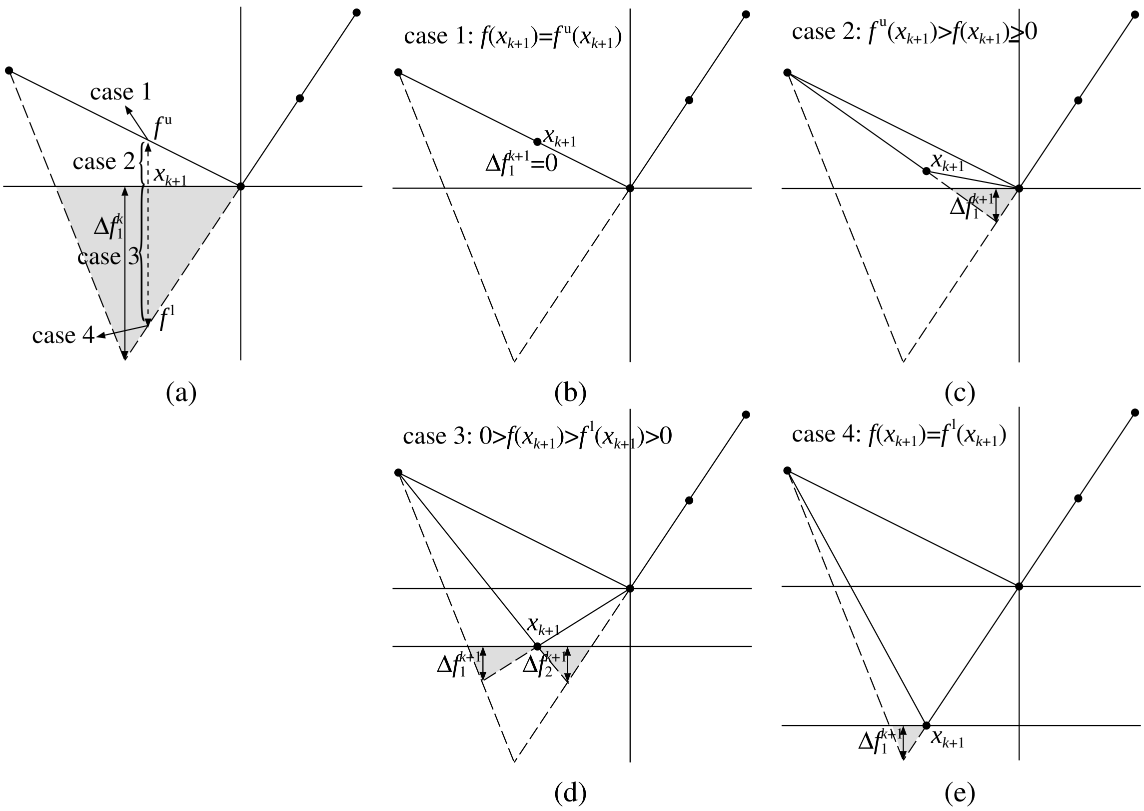


Figure 9. Decrease of range of uncertainty when $\Delta f_2^k = 0$. (a) Four possibilities are distinguished for the new function evaluation $f(x_{k+1})$. (b) $f(x_{k+1}) = f^{\mathrm{u}}(x_{k+1})$. (c) $f^{\mathrm{u}}(x_{k+1}) > f(x_{k+1}) \geq 0$. (d) $0 > f(x_{k+1}) > f^{\mathrm{l}}(x_{k+1})$. (e) $f(x_{k+1}) = f^{\mathrm{l}}(x_{k+1})$.

Now we need to show that $\Delta f_1^{k+1} \le \frac{1}{2}$ or $\frac{1}{2} - \Delta f_1^{k+1} \ge 0$. From properties (i),(ii), and (v) it follows that $B(-\frac{1}{2} - A) + (B - E)(1 + A + B) > 0$. So we need to show that $\frac{1}{2}B(-\frac{1}{2} - A) + \frac{1}{2}(B - E)(1 + A + B) - B(\frac{1}{2}B + AE) \ge 0$. Rewriting the left part of this inequality gives $\frac{1}{4}B + \frac{1}{2}E((-1 - A)(1 + B) - AB)$. Since $B, E, (-1 - A) \ge 0$ the inequality holds, and $\Delta f_1^{k+1} \le \frac{1}{2}$.

3. $0 > f(x_{k+1}) > f^1(x_{k+1})$, see Figure 9(d). Then both $\Delta f_1^{k+1} > 0$ and $\Delta f_2^{k+1} > 0$ will hold. If we write Δf_1^{k+1} as an expression of A, B, C, D, and E we get

$$\Delta f_1^{k+1} = \frac{-E(\frac{1}{2}B + E(-1 - A))}{\frac{1}{2}B - E(-1 - A)}. \tag{5}$$

Now we show that $\frac{1}{2} - \Delta f_1^{k+1} \ge 0$. As $B > 0$, $E < 0$, and $-1 - A > 0$, this can be done by showing that $\frac{1}{2}(\frac{1}{2}B - E(-1 - A)) - (-E(\frac{1}{2}B + E(-1 - A))) \ge 0$. Rewriting the left part of this inequality and using $B \ge -2E(1 + A + B)$ from property (vi) gives $(-1 - A)E^2 + \frac{1}{2}E(1 + A + B) + \frac{1}{4}B \ge (-1 - A)E^2 \ge 0$.

If we write Δf_2^{k+1} as an expression of A, B, C, D, and E we get

$$\Delta f_2^{k+1} = \frac{E(B - E)(1 + A + B) + \frac{1}{2}B(B - E)}{(B - E)(1 + A + B) + B(-\frac{1}{2} - A)}. \tag{6}$$

As $B > E$, $1 + A + B > 0$, $B > 0$, and $-\frac{1}{2} - A > -1 - A \ge 0$, it suffices to show that $\frac{1}{2}((B - E)(1 + A + B) + B(-\frac{1}{2} - A)) - (E(B - E)(1 + A + B) + \frac{1}{2}B(B - E)) \ge 0$. Rewriting the left part of this inequality and using $B \ge -2E(-1 - A)$ from property (vi) gives $-\frac{1}{2}E(1 + A) + \frac{1}{4}B - E(B - E)(1 + A + B) \ge -\frac{1}{2}E(1 + A) + \frac{1}{4}(-2E(-1 - A)) - E(B - E)(1 + A + B) = -E(B - E)(1 + A + B) \ge 0$.

4. $f(x_{k+1}) = f^1(x_{k+1})$, see Figure 9(e). Then $\Delta f_2^{k+1} = 0$, and $\Delta f_1^{k+1} = \Delta f_1^k - (f(M) - f^1(x^{k+1})) = \Delta f_1^k + f^1(x_{k+1})$. So, for $\Delta f_1^{k+1} \le \frac{1}{2}\Delta f_1^k = \frac{1}{2}$ we need to show that $-f^1(x_{k+1}) - \frac{1}{2} \ge 0$. The lower bound is given by $f^1(x_{k+1}) = \max\{-\frac{B}{2(1+A+B)}, -\frac{B}{2(-1-A)}\}$. If $f^1(x_{k+1}) = -\frac{B}{2(1+A+B)}$ we have

$$\frac{B}{2(1 + A + B)} - \frac{1}{2} = \frac{B - (1 + A + B)}{2(1 + A + B)} = \frac{-1 - A}{2(1 + A + B)} \ge 0.$$

If $f^1(x_{k+1}) = -\frac{B}{2(-1-A)}$ we have

$$\frac{B}{2(-1 - A)} - \frac{1}{2} = \frac{B - (-1 - A)}{2(-1 - A)} = \frac{1 + A + B}{2(-1 - A)} \ge 0.$$

□

We use Lemma 2 to show that the triangle section method at least halves the range of uncertainty after two new function evaluations.

Theorem 2. *Let each new function evaluation x_{k+1} be chosen according to (3). Then the range of uncertainty at least halves after two function evaluations, i.e., for all k,* $\max\{\Delta f_1^{k+2}, \Delta f_2^{k+2}\} \le \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$.

Proof. W.l.o.g. we assume that $\Delta f_1^k \ge \Delta f_2^k$. We distinguish three possibilities for the function value $f(x_{k+1})$.

1. $f(x_{k+1}) = f^{\mathrm{u}}(x_{k+1})$. Then for all $y \in [L, M]$ we have $f^{\mathrm{u}}(y) = f^{\mathrm{l}}(y)$, and $\Delta f_1^{k+1} = 0$. Furthermore, $\Delta f_2^{k+1} = \Delta f_2^k$. Now it follows from Lemma 2 that $\max\{\Delta f_1^{k+2}, \Delta f_2^{k+2}\} \leq \frac{1}{2}\Delta f_2^{k+1} \leq \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$.
2. $0 \geq f(x_{k+1}) \geq f^{\mathrm{l}}(x_{k+1})$. Then Δf_1^{k+1} and Δf_2^{k+1} are identical to those given in the proof of Lemma 2 for the corresponding value of $f(x_{k+1})$. It follows that $\max\{\Delta f_1^{k+1}, \Delta f_2^{k+1}\} \leq \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$ and thus $\max\{\Delta f_1^{k+2}, \Delta f_2^{k+2}\} \leq \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$.
3. $f^{\mathrm{u}}(x_{k+1}) > f(x_{k+1}) > 0$. The expression for Δf_1^{k+1} is identical to the one given in the proof of Lemma 2, and it follows that $\Delta f_1^{k+1} \leq \frac{1}{2}\Delta f_1^k$. For Δf_2^{k+1} we distinguish two possibilities.
   - $\Delta f_2^{k+1} \leq \frac{1}{2}\Delta f_1^k$. Then $\max\{\Delta f_1^{k+1}, \Delta f_2^{k+1}\} \leq \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$ and thus $\max\{\Delta f_1^{k+2}, \Delta f_2^{k+2}\} \leq \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$.
   - $\Delta f_2^{k+1} > \frac{1}{2}\Delta f_1^k$. Then also $\Delta f_2^{k+1} > \Delta f_1^{k+1}$. A new point x_{k+2} is now chosen for function evaluation according to (3). For the function value $f(x_{k+2})$ we can also distinguish three possibilities.
     - $f(x_{k+2}) = f^{\mathrm{u}}(x_{k+2})$. As $\Delta f_2^{k+1} > \Delta f_1^{k+1}$, it follows that $\Delta f_2^{k+2} = 0$. Thus, $\max\{\Delta f_1^{k+2}, \Delta f_2^{k+2}\} \leq \Delta f_1^{k+1} \leq \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$.
     - $0 \geq f(x_{k+2}) \geq f^{\mathrm{l}}(x_{k+2})$. Then $\max\{\Delta f_1^{k+2}, \Delta f_2^{k+2}\} \leq \frac{1}{2}\Delta f_2^{k+1} \leq \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$.
     - $f^{\mathrm{u}}(x_{k+2}) > f(x_{k+2}) > 0$. As $\Delta f_2^{k+1} > \Delta f_1^{k+1}$, it follows that $\Delta f_2^{k+2} \leq \frac{1}{2}\Delta f_2^{k+1} \leq \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$. Furthermore, $\Delta f_1^{k+2} \leq \Delta f_1^{k+1} \leq \frac{1}{2}\max\{\Delta f_1^k, \Delta f_2^k\}$.

□

Corollary 1. *Using the triangle section method, the decrease of the range of uncertainty is exponential in the number of function evaluations.*

Proof. Let the initial range of uncertainty be given by $\max\{\Delta f_1^0, \Delta f_2^0\}$, and the range of uncertainty after k new function evaluations by $\max\{\Delta f_1^k, \Delta f_2^k\}$. Then $\max\{\Delta f_1^k, \Delta f_2^k\} \leq \frac{1}{2}\max\{\Delta f_1^{k-2}, \Delta f_2^{k-2}\} \leq \left(\frac{1}{2}\right)^{\lfloor \frac{k}{2} \rfloor}\max\{\Delta f_1^0, \Delta f_2^0\}$. □

So we have shown that the range of uncertainty at least halves for each two new function evaluations. However, the average reduction of the range of uncertainty generally is much bigger, which is backed up by the empirical results that we present in Section 5.

3.3 Piecewise-linear functions

We now consider the case in which the function f is also known to be piecewise linear besides convex or concave. An example of such a function is given in Section 4.

The piecewise-linear property of f can be used as follows to terminate the improved golden section method or the triangle section method with the exact minimum as result. For a piecewise-linear function the slope of a line segment will be identified as soon as three function evaluations are made of points on the segment. Furthermore, the optimum lies at the intersection of two segments. When these two line segments have been identified, the exact minimum can be obtained by determining the intersection of the two line segments.

When dealing with piecewise-linear functions, we use this fact as follows. We evaluate the point x_k with the lowest lower bound value, i.e., $x_k = \arg\min f^{\mathrm{l}}(x)$, when a line segment containing the point with the lowest function evaluation so far has been identified, i.e., when three adjacent points lie on one line segment where the lowest function evaluation is either the first or the last point;

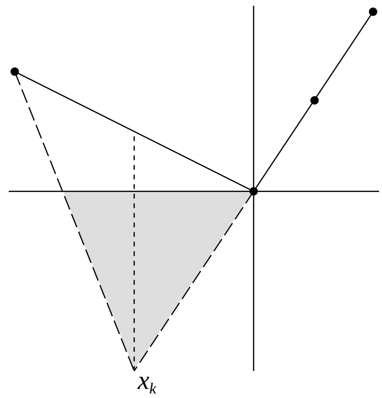


Figure 10. A line segment containing the point with the lowest function evaluation so far has been identified by three adjacent function evaluations. The next point x_k that is chosen for function evaluation is the point with the lowest lower bound value.

see Figure 10 for an example. Both the improved golden section method and the triangle section method can be adjusted to incorporate this strategy. For the improved golden section method it is then required to check before each new function evaluation if the point with the lowest function evaluation lies on one line segment with either the two points immediately before or the two points immediately after itself. For the triangle section method either $\Delta f_1^k = 0$ or $\Delta f_2^k = 0$ holds in this case. Although $\Delta f_1^k = 0$ or $\Delta f_2^k = 0$ also holds when there are two points with the lowest function evaluation, we use it to determine when to deviate from (3) and to evaluate $x_k = \arg\min f^{\mathrm{l}}(x)$ instead. If $f(x_k) = f^{\mathrm{l}}(x_k)$, then x_k is the point with the minimum and we are done. Otherwise, we continue with new function evaluations, also using the latest function evaluation $f(x_k)$.

4 Application

In previous sections we described how we can find the optimum for a concave or convex function of one variable with computationally hard function evaluations. In this section we give a real-life example of such a function, stemming from resource management in an in-home network. For an elaborate description of the problem from which this example originates we refer to Den Boef et al.[1].

Consider a sender and a receiver of data and a network or data transportation device to which both the sender and receiver are connected. At the connections of the sender and the receiver to the network, buffers are placed. Time is discretized into time units t, and for each time unit t the amount of data supplied at the sender is given by $s(t)$ and the amount of data consumed at the receiver is given by $d(t)$. All data is supplied into the buffer at the sender and is consumed from the buffer at the receiver. Reservations of the buffers and transportation device are based on the maximum usage during the time horizon. Costs of the buffers are given by c_{s} and c_{r} per unit buffer size, and cost of the transportation device is given by c_{b} per unit transportation capacity.

The problem is to determine reservations b of the transportation device, m_{s} and m_{r} of the buffers, and a feasible transmission schedule of the data given by $x(t)$ for each time unit t such that total costs are minimized. The transmission schedule has to be such that whenever data is taken from a buffer, it also is available in the buffer (no buffer underflow), and whenever data is put into a buffer, the buffer reservation is not exceeded. Also, the amount transmitted during a time unit may not exceed the transportation capacity reservation. This can be formulated as a linear program as follows.

$$
\begin{array}{ll}
\text{minimize} & c_{\mathrm{b}}b + c_{\mathrm{s}}m_{\mathrm{s}} + c_{\mathrm{r}}m_{\mathrm{r}} \\
\text{subject to} & x(t) \le b \\
& \sum_{k=1}^{t} s(k) - \sum_{k=1}^{t} x(k) \ge 0 \\
& \sum_{k=1}^{t} s(k) - \sum_{k=1}^{t-1} x(k) \le m_{\mathrm{s}} \\
& \sum_{k=1}^{t} x(k) - \sum_{k=1}^{t} d(k) \ge 0 \\
& \sum_{k=1}^{t} x(k) - \sum_{k=1}^{t-1} d(k) \le m_{\mathrm{r}}
\end{array}
\tag{7}
$$

In this LP all constraints are for every time unit t. Since a typical data stream can have a time horizon that is split up into more than 100,000 time units, the LP consists of a large amount of variables and constraints, leading to a relatively long calculation time when standard LP-methods are used. To speed up the calculation time the following problem-specific method is used which has complexity $\mathcal{O}(|T|)$ with T the set of time units. Given a transportation capacity the total buffer costs can be minimized by first minimizing the buffer with the highest cost coefficient and then minimizing the buffer with the lowest cost coefficient. So for a given value of b optimal values of m_{s} and m_{r} can be determined. This leads to the following reformulation of the problem.

Let f be a function on transportation capacity b with function values that represent the minimum total costs. So, for input b, function f determines optimal values of m_{s} and m_{r} given the cost coefficients c_{s} and c_{r} and then calculates the total costs $c_{\mathrm{b}}b + c_{\mathrm{s}}m_{\mathrm{s}} + c_{\mathrm{r}}m_{\mathrm{r}}$ which are returned as output. The problem now is to find the minimum of the function f. Since the original problem is an LP-problem, f is a continuous, piecewise-linear, convex function. So, the method described in this paper can be used to find an optimum. In Section 5 we show some results of the improved golden section method and the triangle section method concerning this application.

5 Numerical test results

In this section we present numerical test results for the improved golden section method and the triangle section method. These results are obtained by using these methods on the application described in Section 4 and on numerous mathematical functions, which can be divided into two types. Functions of type 1 are polynomial functions, given by $f(x) = a(x-b)^{2c}$, with $a = 0.5, 1, 1.5, \ldots, 9.5, 10$, $b = 1, 2, \ldots, 10$, and $c = 1, 2, \ldots, 5$. This gives a total of 1000 functions of type 1. A function f of type 2 is given by $f(x) = ae^{b(x-c)} - dx$, with $a = 1, 2, \ldots, 10$, $b = 1, 2, \ldots, 5$, $c = -5, -4, \ldots, 5$, and $d = 0.01, 0.05, 0.25, 1.25, 6.25, 31.25, 156.25, 781.25, 3906.25$. This gives a total of 4950 functions of type 2. For both functions of type 1 and functions of type 2 the objective was to find the minimum on the interval $[-10, 10]$. Finally, for the application discussed in the previous section we used 16 different video traces, each combined with a number of different cost coefficients c_{b}, c_{s}, and c_{r}, which resulted in 283 problem instances.

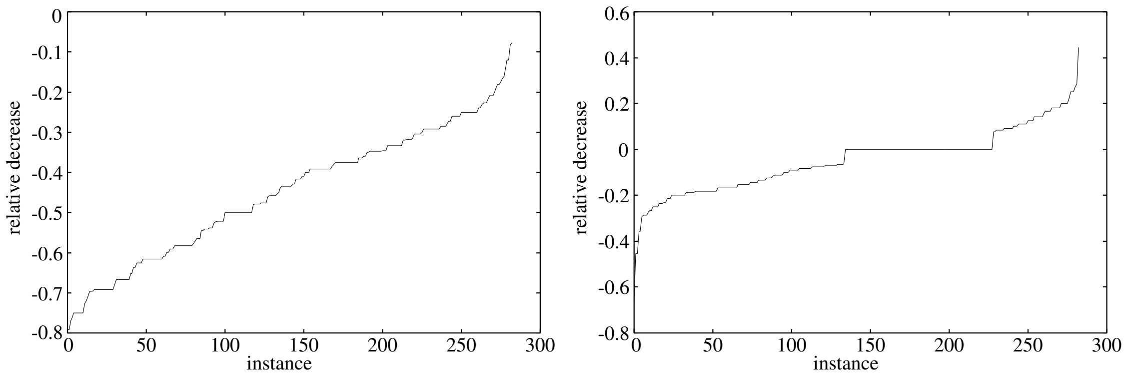
In Table 1 results are given for comparing regular golden section, improved golden section, and triangle section using the (a) 1000 functions of type 1, (b) 4950 functions of type 2, and (c) 283 application instances. The results consist of the average difference between the best found solution and the optimal solution, and the average size of the interval of uncertainty after a given number of function evaluations, with both averages over all functions or instances. These results show that for functions of type 1 and application instances the triangle section method on average gets closest to the optimal function value. For functions of type 2, however, it is outperformed by the improved golden section method, possibly due to the largely asymmetrical shape of these functions. Regarding the interval of uncertainty, we notice that the improved golden section method

(a) Type 1	Avg. deviation from optimum			Avg. int. of uncertainty		
# func.eval.	gs	igs	ts	gs	igs	ts
5	7126.103	8.087	2.054	6.111	5.733	5.638
6	3.649	3.137	0.163	3.708	3.404	2.993
7	0.200	0.192	0.008	2.265	2.076	1.678
8	0.068	0.041	0.003	1.390	1.264	1.069
9	0.030	0.013	0.001	0.855	0.770	0.799
10	0.005	0.002	0.000	0.527	0.470	0.765

(b) Type 2	Avg. deviation from optimum			Avg. int. of uncertainty		
# func.eval.	gs	igs	ts	gs	igs	ts
5	605.264	531.365	727.436	6.912	4.606	5.098
6	283.096	229.274	272.258	4.262	2.408	2.736
7	145.214	87.939	89.639	2.631	1.368	1.528
8	68.755	31.416	32.209	1.625	0.784	0.906
9	32.362	15.685	16.408	1.004	0.447	0.575
10	17.983	12.590	12.595	0.620	0.255	0.409

(c) App.	Avg. deviation from optimum			Avg. int. of uncertainty		
# func.eval.	gs	igs	ts	gs	igs	ts
5	112082	81744	70924	5079	3990	3362
6	40726	28890	23090	3100	1803	1831
7	19950	11894	10123	1900	839	1017
8	11651	7053	5585	1168	391	610
9	7301	3857	2902	719	182	407
10	4255	1945	1712	444	78	325

Table 1. Results after 5–10 function evaluations for the golden section method (gs), the improved golden section method (igs), and the triangle section method (ts). Tables (a), (b), and (c) give the results for the functions of type 1 and type 2, and for the application, respectively. In each table, the leftmost column gives the number of function evaluations. The next three columns give for each method and number of function evaluations the average over all functions or application instances of the difference between the function value of the best found solution and the optimal function value. The three rightmost columns give for each method the average over all functions or application instances of the resulting interval of uncertainty after the given number of function evaluations. Finally, improved golden section and triangle section were stopped when for a function the range of uncertainty was less than 0.01. When this happened, the last calculated result for a function was also used as a result for a higher number of function evaluations. This explains why, e.g. the average size of the interval of uncertainty for type 1 functions using triangle section is for 10 function evaluations the largest compared to regular and improved golden section, while for fewer function evaluations it is the smallest.



(a) (b)

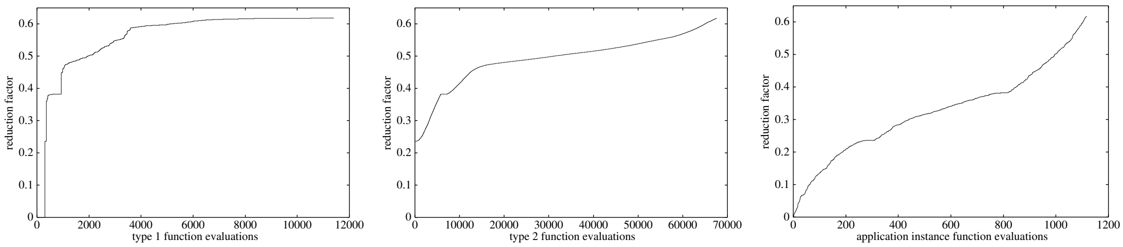
Figure 11. In these two graphs the number of function evaluations that two of the three methods require are compared. Figure (a) compares the improved golden section method with the regular golden section method for the 283 instances of the application. Both methods were stopped when the interval of uncertainty was less than 1. The graph gives for all 283 instances the relative decrease in the number of function evaluations required by improved golden section compared to regular golden section, i.e., for each application instance $(igs - gs)/gs$ is given, with igs and gs the number of function evaluations required for improved golden section and regular golden section, respectively. The instances are sorted on non-decreasing size of the relative decrease. Figure (b) compares the triangle section method to the improved golden section method, both using the procedure that uses piecewise linearity to determine the exact optimum. The graph gives for all 283 instances the relative decrease in the number of function evaluations required by triangle section compared to improved golden section, i.e., for each application instance $(ts - igs)/igs$ is given, with ts and igs the number of function evaluations required for triangle section and improved golden section, respectively.

improves upon the results of regular golden section especially for functions of type 2 and the application instances. For functions of type 1, it also returns better results than regular golden section, but here the improvement is not so impressive. This can be explained by the relatively steep slopes of the functions of type 1 surrounding both sides of the optimum.

Figure 11(a) compares improved golden section with regular golden section using the application instances and as stop criterion the size of interval of uncertainty being less than 1. It shows that the reduction in number of function evaluations can be as high as 80%, and on average around 40%. Figure 11(b) compares triangle section with improved golden section using the application instances and the piecewise-linear property, thus obtaining the exact optimum. For about one half of the instances triangle section requires fewer function evaluations than improved golden section. However, for about one quarter of the instances it requires more function evaluations. Still, the average number of function evaluations is lower for triangle section than for improved golden section.

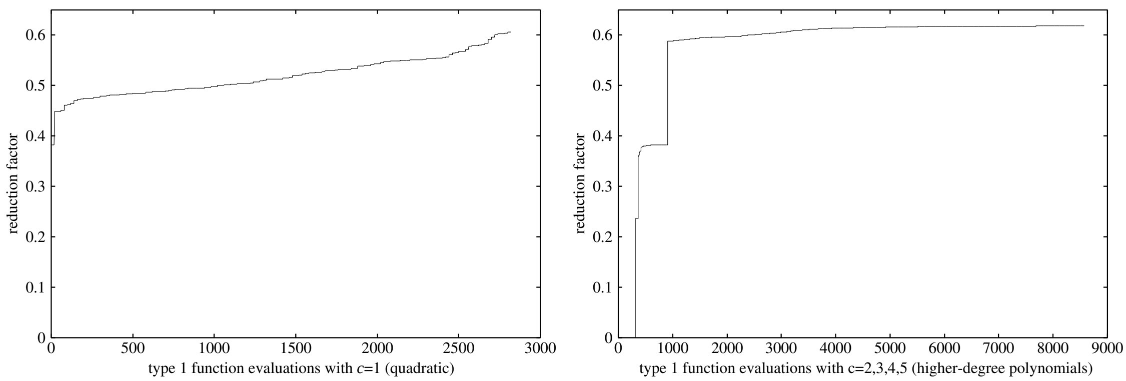
Figure 12 shows the reduction factors of the interval of uncertainty after a function evaluation, which were observed when applying the improved golden section method on all functions and instances. It shows that the reduction factor is often close to the golden section (≈ 0.618) for functions of type 1. However, for functions of type 2 and application instances the reduction is much more significant, i.e., we get much smaller intervals of uncertainty.

Figure 13 shows again the reduction factors of the interval of uncertainty using the improved golden section method for functions of type 1, but now split into quadratic functions ($c = 1$) and functions of higher degree ($c \geq 2$). It shows that the reduction factors for quadratic functions are



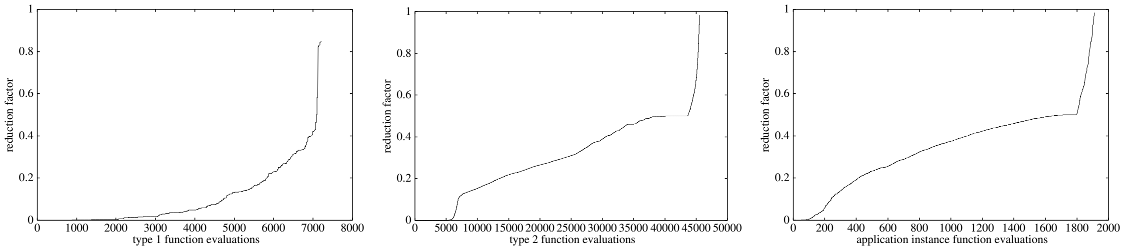
(a) (b) (c)

Figure 12. The observed reduction factors of the interval of uncertainty after a function evaluation using the improved golden section method, sorted increasingly. The number of performed function evaluations per function or application instance varies. Figure (a) gives the reduction factors for all 1000 type 1 functions, figure (b) for all 4950 type 2 functions, and figure (c) for all 283 application instances.



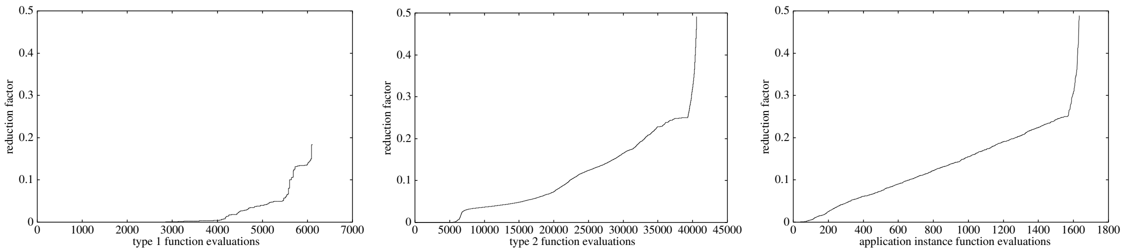
(a) (b)

Figure 13. The observed reduction factors of the interval of uncertainty after a function evaluation using the improved golden section method, sorted increasingly. The number of performed function evaluations per function or application instance varies. Figure (a) gives the reduction factors for all quadratic type 1 functions, figure (b) of all other type 1 functions.



(a) (b) (c)

Figure 14. The observed reduction factors of the range of uncertainty after a single function evaluation using the triangle section method, sorted increasingly. These reduction factors were observed after at least three functions evaluations were made for a specific function or application instance. The number of performed function evaluations per function or application instance varies. Figure (a) gives the reduction factors for all 1000 type 1 functions, figure (b) for all 4950 type 2 functions, and figure (c) for all 283 application instances.



(a) (b) (c)

Figure 15. The observed reduction factors of the range of uncertainty after two consecutive function evaluations using the triangle section method, sorted increasingly. These reduction factors were observed after at least three functions evaluations were made for a specific function or application instance. The number of performed function evaluations per function or application instance varies. Figure (a) gives the reduction factors for all 1000 type 1 functions, figure (b) for all 4950 type 2 functions, and figure (c) for all 283 application instances.

distributed evenly between 0.45 and 0.6 in contrast with the reduction factors for higher-degree functions, of which approximately 90% is close to 0.6. As the gradient of functions of higher degree at the evaluated points is larger than the gradient of quadratic functions, the lower bounds for the functions of higher degree have a steeper slope, and thus lead to a smaller reduction of the interval of uncertainty.

Figure 14 shows the reduction factors of the range of uncertainty after a single function evaluation, which were observed when applying the triangle section method on all functions and instances. It shows that more than 90% of the recorded reduction factors is spread out over the interval $[0, 0.5]$, and only a small fraction of the recorded reduction factors are between 0.5 and 1. Figure 15 shows the reduction factors of the range of uncertainty after two consecutive function evaluations, cf. Theorem 2. It shows a similar distribution, with most of the observed reduction factors between 0 and 0.25 and only a small fraction between 0.25 and 0.5.

6 Conclusion

In this paper we have considered the problem of line searching for convex functions. We have shown how the convexity property can be used to obtain upper and lower bounds on the function using the performed function evaluations. For some well-known line search methods we have shown, using these upper and lower bounds, that they may propose a candidate which is not optimal. We have presented two new line search methods which use the convexity property. The first method, the improved golden section method, uses the upper and lower bounds to improve upon the regular golden section method and always proposes a candidate which can be optimal. The second method, the triangle section method, focuses on minimizing the interval for possible objective values, the range of uncertainty, and we have shown that it at least halves the range of uncertainty after every two function evaluations.

Both methods were tested using a real-life example and two classes of convex functions. It was shown that the new methods give better approximations of the optimum than regular golden section after a fixed number of function evaluations. This also translated into a sometimes heavily reduced number of function evaluations that was required to obtain the optimum. A direct comparison of the new methods did not show a clear winner; depending on the instance either improved golden section or triangle section gave the best results.

There are several possibilities for future research in line searching methods for convex functions. The upper and lower bounds based on the convexity property can be used to adapt other well-known line search methods. They can also be used to try to estimate the complete function as efficiently as possible instead of only the optimum. Finally, it would be interesting to see how the work presented in this paper extends to multivariate, convex functions.

References

[1] Edgar den Boef, Wim F.J. Verhaegh, and Jan Korst. Smoothing streams in an in-home digital network: Optimization of bus and buffer usage. *Telecommunication Systems*, 23:273–295, 2003.

[2] Philip E. Gill, Walter Murray, and Margaret H. Wright. *Practical Optimization*. Academic Press, San Diego, California, seventh edition, 1988.

A Height of a triangle in the area of uncertainty

For the explanation of the triangle section method to obtain a guaranteed range reduction, the heights of the triangles in the area of uncertainty, Δf_1^k and Δf_2^k, need to be determined. The expressions of Δf_1^k and Δf_2^k in points L, M, and U with their function evaluations $f(L)$, $f(M)$, and $f(U)$, and the interval of uncertainty $[L', U']$ can be obtained by determining the intersection of two appropriate lines and subtracting this from the minimum found function evaluation. We first give a general expression of the intersection of two lines both determined by two points. Then we indicate for equations (1), (2) , (4), (5), and (6) how they can be obtained.

Let $(a, f(a))$ and $(b, f(b))$ define a line l_1, and $(c, f(c))$ and $(d, f(d))$ a line l_2. So, $l_1(x)$ is given by

$$\frac{f(b)-f(a)}{b-a}x+\frac{f(a)b-f(b)a}{b-a}$$

and $l_2(x)$ is given by

$$\frac{f(d)-f(c)}{d-c}x+\frac{f(c)d-f(d)c}{d-c}.$$

Let line l_1 and line l_2 intersect at $(y, f(y))$. Then using the above line equations with $x = y$, we derive from $l_1(y) = l_2(y)$ that

$$y = \frac{\frac{f(c)d - f(d)c}{d-c} - \frac{f(a)b - f(b)a}{b-a}}{\frac{f(b)-f(a)}{b-a} - \frac{f(d)-f(c)}{d-c}}.$$

Simplifying the above equation gives

$$y = \frac{(f(c)d - f(d)c)(b-a) - (f(a)b - f(b)a)(d-c)}{(f(b) - f(a))(d-c) - (f(d) - f(c))(b-a)}.$$

Substituting the above equation for y into $l_1(y)$ or $l_2(y)$ gives for $f(y)$

$$f(y) = \frac{(f(c)d - f(d)c)(f(b) - f(a)) - (f(a)b - f(b)a)(f(d) - f(c))}{(f(b) - f(a))(d-c) - (f(d) - f(c))(b-a)}.$$

Equation (1):
$(a, f(a)) = (L, f(L))$, $(b, f(b)) = (L', f(M))$, $(c, f(c)) = (M, f(M))$, and $(d, f(d)) = (U, f(U))$.
$\Delta f_1^k = f(M) - f(y)$.

Equation (2):
$(a, f(a)) = (L, f(L))$, $(b, f(b)) = (M, f(M))$, $(c, f(c)) = (U', f(M))$, and $(d, f(d)) = (U, f(U))$.
$\Delta f_1^k = f(M) - f(y)$.

Equation (4):
$(a, f(a)) = (A, B)$, $(b, f(b)) = (-\frac{1}{2}, E)$, $(c, f(c)) = (0, 0)$, and $(d, f(d)) = (D, f(D))$.
$\Delta f_1^{k+1} = 0 - f(y)$.

Equation (5):
$(a, f(a)) = (A, B)$, $(b, f(b)) = (-1, 0)$, $(c, f(c)) = (-\frac{1}{2}, E)$, and $(d, f(d)) = (0, 0)$.
$\Delta f_1^{k+1} = E - f(y)$.

Equation (6):
$(a, f(a)) = (A, B)$, $(b, f(b)) = (-\frac{1}{2}, E)$, $(c, f(c)) = (0, 0)$, and $(d, f(d)) = (D, f(D))$.
$\Delta f_1^{k+1} = E - f(y)$.