BIT 24 (1984), 366–372

AN INTERVAL VERSION OF THE SECANT METHOD

A. NEUMAIER

Institut für Angewandte Mathematik der Universität Freiburg, Hermann-Herder-Straße 10, D-7800 Freiburg i.Br., Germany

Abstract.

If an interval $[\underline{m}, \bar{m}] \notin 0$ enclosing the derivative in an initial interval is known then interval arithmetic can be used to force global convergence of any local zero finder without sacrificing convergence speed. For the secant iteration, details are given.

Let f be a real, twice continuously differentiable function which is monotone in the interval $x := [\underline{x}, \bar{x}]$ and has there a zero x^*. To determine x^* numerically, the bisection method is globally but only linearly convergent; other methods, e.g. the secant method, are superlinearly convergent but lack the global convergence property (the iterates may oscillate or leave the monotonicity interval x). In this note we use interval arithmetic to force global convergence without sacrificing the superlinear convergence property, but an interval free version of the resulting algorithm is also given. We note that there are other modifications of the secant method with global convergence (Brent [3], King [6] and references there), but the present interval arithmetic approach has the advantage of automatically taking account of the rounding errors. Moreover, an adaption of the technique developed here to the multidimensional case is possible and will be discussed elsewhere [8].

We shall denote (real) intervals by small letters, the left endpoint, midpoint, and right endpoint of an interval $x = [a, b]$ by $\underline{x} := a$, $\check{x} := \frac{1}{2}(a+b)$, and $\bar{x} := b$, respectively, and write $\varrho(x) := \frac{1}{2}(\bar{x} - \underline{x}) = \check{x} - \underline{x} = \bar{x} - \check{x}$ for the radius of x. Degenerate intervals x (with $\underline{x} = \check{x} = \bar{x}$) are identified with the point $\check{x}$.

Furthermore, we define

$$|x| := \sup\{|\tilde{x}| \mid \tilde{x} \in x\} = |\check{x}| + \varrho(x),$$
$$\langle x \rangle := \inf\{|\tilde{x}| \mid \tilde{x} \in x\} = |\check{x}| - \varrho(x) \quad \text{if } 0 \notin x.$$

Intervals can be added, subtracted, multiplied and divided (with exception of division through an interval containing zero) according to the rule

$$x * y := \square\{\tilde{x} * \tilde{y} \mid \tilde{x} \in x, \tilde{y} \in y\} = \square\{\underline{x} * \underline{y}, \underline{x} * \bar{y}, \bar{x} * \underline{y}, \bar{x} * \bar{y}\}$$

Received December 1983. Revised May 1984.

where $*$ is one of $+$, $-$, $\cdot$, and $/$, and $\square S := [\inf S, \sup S]$ denotes the smallest interval containing all points of the bounded subset S.

For the properties of interval arithmetic, and for details about their realization on a computer, see the books by Moore [7] or Alefeld and Herzberger [2]. There it is also discussed how to use interval arithmetic in order to find intervals containing the range of functions which are expressible in terms of the elementary operations $+$, $-$, $\cdot$, $/$, and certain standard arithmetic functions like $\exp x$, $\sin x$, etc. In particular, if f is a function expressed in this way then f' can also be expressed in this way, whence it is possible to find an interval m containing the range of f' in the interval x:

$$f'(\tilde{x}) \in m \quad \text{for all } \tilde{x} \in x. \tag{1}$$

From now on we shall assume that $0 \notin m$. This is a natural assumption since we required f to be monotone in x; if f does not contain saddle points then this can be enforced by preliminary bisection steps (cf. [2], pp. 100–101) or by enclosing the range of f' sufficiently narrowly (see [1] for a suitable algorithm). With such an interval at our disposal we can formulate the tools needed for the algorithm.

PROPOSITION 1. *If m satisfies* (1) *and $0 \notin m$ then the zero x^* of f in x satisfies*

$$|f(\tilde{x})| \leqq |m|\,|\tilde{x} - x^*|, \tag{2}$$

$$x^* \in \tilde{x} - f(\tilde{x})/m, \tag{3}$$

for every $\tilde{x} \in x$.

PROOF. By the mean-value theorem, $f(\tilde{x}) = f(\tilde{x}) - f(x^*) = f'(\xi)(\tilde{x} - x^*)$ for a suitable $\xi \in x$; therefore $|f(\tilde{x})| \leqq |f'(\xi)|\,|\tilde{x} - x^*| \leqq |m|\,|\tilde{x} - x^*|$ and $x^* = \tilde{x} - f(\tilde{x})/f'(\xi) \in \tilde{x} - f(\tilde{x})/m$. ∎

PROPOSITION 2. *If $0 \notin m$ then the radius of the interval*

$$z := (\tilde{x} - f(\tilde{x})/m) \cap x$$

satisfies (*provided $z \neq \emptyset$*) *the inequality*

$$\varrho(z) \leqq q \cdot \min(|x - \tilde{x}|, |f(\tilde{x})|/\langle m \rangle), \tag{4}$$

where

$$q := \frac{1}{2}\left(1 - \frac{\langle m \rangle}{|m|}\right) < \tfrac{1}{2}. \tag{5}$$

PROOF. Note first that $0 < \langle m \rangle \leqq |m|$ so that indeed $q < \frac{1}{2}$. Inequality (4) is obvious if $f(\tilde{x}) = 0$; hence we shall assume that $f(\tilde{x}) \neq 0$. Put

$$s := -1 \quad \text{if } f(\tilde{x})/m > 0, \qquad s := 1 \text{ otherwise.}$$

Then the number $w := s(z - \tilde{x})$ satisfies

$$\varrho(w) = \varrho(z - \tilde{x}) = \varrho(z),$$

and by (3) we have

$$w \subseteq s((\tilde{x} - f(\tilde{x})/m) - \tilde{x}) = -\frac{sf(\tilde{x})}{m} = \frac{|f(\tilde{x})|}{[\langle m \rangle, |m|]} = \left[\frac{|f(\tilde{x})|}{|m|}, \frac{|f(\tilde{x})|}{\langle m \rangle}\right].$$

Hence

$$\varrho(w) \leqq \frac{1}{2}\left(\frac{|f(\tilde{x})|}{\langle m \rangle} - \frac{|f(\tilde{x})|}{|m|}\right) = q \cdot \frac{|f(\tilde{x})|}{\langle m \rangle}.$$

This implies (4) unless $|f(\tilde{x})|/\langle m \rangle > |x - \tilde{x}|$. But in this case,

$$\underline{w} \geqq \frac{|f(\tilde{x})|}{|m|} = \frac{|f(\tilde{x})|}{\langle m \rangle}(1 - 2q) \geqq |x - \tilde{x}|(1 - 2q),$$

$$\bar{w} \leqq |w| = |z - \tilde{x}| \leqq |x - \tilde{x}|,$$

whence $\varrho(w) = \frac{1}{2}(\bar{w} - \underline{w}) \leqq q|x - \tilde{x}|$, and (4) holds again. ■

PROPOSITION 3. *With assumptions and notation of Propositions* 1 *and* 2,

$$\varrho(z) \leqq |\tilde{x} - x^*| \varrho(m)/\langle m \rangle. \tag{6}$$

PROOF. By (4) and (2), we have

$$\varrho(z) \leqq q|f(\tilde{x})|/\langle m \rangle \leqq q|m|\,|\tilde{x} - x^*|/\langle m \rangle = \frac{|m| - \langle m \rangle}{2\langle m \rangle}|\tilde{x} - x^*| = \frac{\varrho(m)}{\langle m \rangle}|\tilde{x} - x^*|,$$

which implies (6). ■

Now suppose that $\tilde{x}_0, \tilde{x}_1, \ldots, \tilde{x}_i, \ldots$ is a convergent sequence of real numbers $\in x$ with limit x^* (the required zero). We associate with it the interval sequence defined by

$$x_0 := x, \qquad x_{i+1} := (\tilde{x}_i - f(\tilde{x}_i)/m) \cap x_i. \tag{7}$$

By (3), $x^* \in x_i$ for all $i \geqq 0$, and by (6),

$$\varrho(x_{i+1}) \leqq |\tilde{x}_i - x^*| \cdot \varrho(m)/\langle m \rangle, \tag{8}$$

so that $\varrho(x_i) \to 0$ for $i \to \infty$. Therefore, the interval sequence x_i converges to the zero x^* and each x_i contains x^* by Proposition 1. Moreover, the radii of the x_i converge to zero with the same speed as the approximations $\tilde{x}_i$ converge towards x^*.

The iteration (7) is discussed e.g. in Alefeld and Herzberger [2]. They show that

$$\varrho(x_{i+1}) \leqq 2q\varrho(x_i) \qquad \text{if } \tilde{x}_i \in x_i, \tag{9}$$

and, sharper,

$$\varrho(x_{i+1}) \leqq q\varrho(x_i) < \tfrac{1}{2}\varrho(x_i) \qquad \text{if } \tilde{x}_i = \check{x}_i. \tag{10}$$

This follows immediately from (4) since $|x_i - \tilde{x}_i| \leqq 2\varrho(x_i)$ if $\tilde{x}_i \in x_i$, and $= \varrho(x_i)$ if $\tilde{x}_i = \check{x}_i$. To guarantee $\tilde{x}_i \in x_i$, the sequence $\tilde{x}_i$ must be generated simultaneously with the sequence x_i. As suggested by equation (8) it is sensible to pick $\tilde{x}_i$ as an approximation of the root x^*. We choose here the secant method to estimate $\tilde{x}_i$, but other estimators leading to fast local convergence (e.g. Newton's method, or one of the many formulae discussed in Traub [9]) can be used as well. We proceed according to the following algorithm.

Interval secant method (valid if (1) holds with $0 \notin m$):
Put $\tilde{x}_0 := \underline{x}, \tilde{x}_1 := \bar{x}, x_1 := x$.
If $f(\tilde{x}_0)f(\tilde{x}_1) > 0$ then stop with message "no zero exists";
For $i = 1, 2, 3, \ldots$ do:
$x_{i+1} := (\tilde{x}_i - f(\tilde{x}_i)/m) \cap x_i$;

If $x_{i+1} = x_i$ then stop with message "zero in x_i (limit accuracy)";

$$r_i := \tilde{x}_i - \frac{\tilde{x}_{i-1} - \tilde{x}_i}{f(\tilde{x}_{i-1}) - f(\tilde{x}_i)} * f(\tilde{x}_i);$$

$$\tilde{x}_{i+1} := \begin{cases} \underline{x}_{i+1} & \text{if } r_i < x_{i+1}, \\ \bar{x}_{i+1} & \text{if } r_i > x_{i+1}, \\ r_i & \text{otherwise.} \end{cases}$$

Note that since $f(\tilde{x}_{i-1})$ is available from the previous iteration step, only one function evaluation and a few arithmetic operations are performed in each step except the first. The number r_i is the approximation of the zero obtained by linear interpolation of f at $\tilde{x}_{i-1}$ and $\tilde{x}_i$, and if $r_i \in x_{i+1}$ for all i, the sequence x_i

is just that defined by the secant method. But we use instead an endpoint as approximation whenever r_i leaves the current optimal inclusion interval x_{i+1}.

THEOREM. *The iterates x_i, $\tilde{x}_i$ of the interval secant method satisfy the relations*

$$\varrho(x_{i+1}) \leqq 2q\varrho(x_i), \tag{11}$$

$$|\tilde{x}_{i+1}-x^*| \leqq C|\tilde{x}_i-x^*|\,|\tilde{x}_{i-1}-x^*|, \tag{12}$$

with $q<\frac{1}{2}$ given by (5), *and a suitable constant $C>0$. In particular, the interval secant method is globally convergent, and*

$$\tilde{x}_i \to x^*, \qquad \varrho(x_i)\to 0$$

superlinearly, with R-order $\geqq \frac{1}{2}(1+\sqrt{5}) \approx 1.618$.

PROOF. By the remarks above, $\tilde{x}_i \in x_i$ for all $i \geqq 0$; therefore (9) applies which gives (11). Since $2q<1$ by (5), we have $\varrho(x_i)\to 0$, whence $\tilde{x}_i \to x^*$ because $\tilde{x}_i$, $x^* \in x_i$ for all $i \geqq 0$. To show (12) we note that the classical analysis of the secant method (see e.g. [9]) shows the existence of a constant $C>0$ such that $|r_i-x^*| \leqq C|\tilde{x}_i-x^*|\cdot|\tilde{x}_{i-1}-x^*|$. Since $x^* \in x_{i+1}$, the construction of $\tilde{x}_{i+1}$ guarantees that $|\tilde{x}_{i+1}-x^*| \leqq |r_i-x^*|$ so that (12) follows. Now let i_0 be such that $4C|\tilde{x}_i-x^*| \leqq 1$ for $i \geqq i_0$. Then with $\kappa := \frac{1}{2}(1+\sqrt{5})$, the positive solution of $\kappa^2=\kappa+1$, we have

$$|\tilde{x}_i-x^*| \leqq \frac{1}{C}\cdot(\tfrac{1}{2})^{\kappa^{i-i_0}} \tag{13}$$

for $i=i_0$ and $i=i_0+1$, and inductively for all $i \geqq i_0$. By (7) and Proposition 3, this implies

$$\varrho(x_i) \leqq \frac{\varrho(m)}{\langle m\rangle C}\cdot(\tfrac{1}{2})^{\kappa^{i-i_0}} \tag{14}$$

for all $i \geqq i_0$, and (13) and (14) imply superlinear convergence and R-order $\geqq \kappa$. ■

REMARK. In a similar way it can be shown that if r_i is computed by a higher order formula satisfying

$$|r_i-x^*| \leqq C|\tilde{x}_i-x^*|^{p_0}|\tilde{x}_{i-1}-x^*|^{p_1}\dots|\tilde{x}_{i-s}-x^*|^{p_s}$$

then the resulting interval method is globally convergent with R-convergence order $\geqq \kappa$, the positive solution of the equation

$$\kappa^{s+1} = p_0\kappa^s + p_1\kappa^{s-1} + \dots + p_s.$$

Thus, any local high order formula can be globalized by interval arithmetic without sacrificing the convergence order.

In finite precision rounded interval arithmetic (with proper directed rounding) the algorithm stops after a finite number of steps with an interval $x_{i+1} = x_i$ containing the zero x^*, since clearly $x \supseteq x_1 \supseteq x_2 \supseteq \ldots \supseteq x_i \ni x^*$ and x contains only a finite number of machine intervals. To take account properly of the rounding errors involved in the computation of $f(\tilde{x}_i)$ the $\tilde{x}_i$ must be declared as (degenerate) intervals. On the other hand, the computation of r_i can be done in ordinary arithmetic, with, say, the midpoints of $f(\tilde{x}_{i-1})$ and $f(\tilde{x}_i)$. The decision whether $f(\tilde{x}_0)f(\tilde{x}_1) > 0$ or not is impossible if the interval computed contains zero properly; in this case the rounding errors prevent a decision whether f contains a zero in x or not.

The speed of the iteration, and the limit accuracy in finite precision arithmetic, can be increased by updating m during the iteration; this decreases the factor $\varrho(m)/\langle m\rangle$ in (8), but at the cost of considerably more computational work for recomputing m.

Recently two other interval modifications of the secant method have been proposed by Cornelius [4] and Herzberger [5]. These also have global convergence but a disadvantage compared with the present method is their need for a bound of some higher derivative over the initial interval which may be difficult to obtain.

We now reformulate the algorithm so that we can dispense with interval arithmetic. We treat only the case of increasing f (the decreasing case can be handled by using $-f$ in place of f), and assume the knowledge of constants $\underline{m}$, $\bar{m}$ with

$$0 < \underline{m} \leqq f'(\tilde{x}) \leqq \bar{m} \qquad \text{for all } \tilde{x} \in [\underline{x}, \bar{x}]. \tag{15}$$

We represent the interval x_i explicitly by its endpoints $\underline{x}_i$, $\bar{x}_i$, and simulate the interval operation in the formation of x_{i+1}. This results in the following algorithm.

Global secant method (valid if (15) holds):
Put $\tilde{x}_0 := \underline{x}_0 := \underline{x}$, $\tilde{x}_1 := \bar{x}_1 := \bar{x}$.
If $f(\tilde{x}_0)f(\tilde{x}_1) > 0$ then stop with message "no zero exists".
For $i = 1, 2, 3, \ldots$ do:
  If $f(\tilde{x}_i) \geqq 0$ then $(p_i := \underline{m}, q_i := \bar{m})$ else $(p_i := \bar{m}, q_i := \underline{m})$;
  $\underline{x}_{i+1} := \max(\underline{x}_i, \tilde{x}_i - f(\tilde{x}_i)/p_i)$;
  $\bar{x}_{i+1} := \min(\bar{x}_i, \tilde{x}_i - f(\tilde{x}_i)/q_i)$.
If $(\underline{x}_{i+1} = \underline{x}_i, \bar{x}_{i+1} = \bar{x}_i)$ or $(\underline{x}_{i+1} \geqq \bar{x}_{i+1})$ then stop with message "zero approximation $\frac{1}{2}(\underline{x}_i + \bar{x}_i)$";

$$r_i := \bar{x}_i - \frac{\bar{x}_{i-1} - \bar{x}_i}{f(\bar{x}_{i-1}) - f(\bar{x}_i)} \cdot f(\bar{x}_i);$$

$$\bar{x}_{i+1} := \begin{cases} \underline{x}_{i+1} & \text{if } r_i < \underline{x}_{i+1}, \\ \bar{x}_{i+1} & \text{if } r_i > \bar{x}_{i+1}, \\ r_i \text{ otherwise.} \end{cases}$$

The algorithm produces a monotonically increasing sequence $\underline{x}_i$ and a monotonically decreasing sequence $\bar{x}_i$ both converging to the zero x^* and enclosing it; but in finite precision arithmetic the enclosing property can be lost due to rounding errors.

On the other hand, as a referee remarked, floating point arithmetic with controllable rounding is called for by the IEEE standard for floating-point arithmetic. If this capability is available, narrow upper and lower bounds for $f(\bar{x}_i)$ can be computed, and the global secant method can hence be performed in such a way that the computed $\underline{x}_{i+1}$ does not exceed the true $\underline{x}_{i+1}$, and the computed $\bar{x}_{i+1}$ is at least as large as the true $\bar{x}_{i+1}$. Then inclusion remains guaranteed, and the stopping criterion can be replaced by

If $(\underline{x}_{i+1} = \underline{x}_i, \bar{x}_{i+1} = \bar{x}_i)$ then stop with message "zero in $[\underline{x}_i, \bar{x}_i]$ (limit accuracy)".

REFERENCES

1. N. S. Asaithamba, S. Zuhe and R. E. Moore, *On computing the range of values*, Computing 28 (1982), 225–237.
2. G. Alefeld and J. Herzberger, *Einführung in die Intervallrechnung*. Bibl. Institut, Mannheim – Wien – Zürich 1974.
3. R. P. Brent, *An algorithm with guaranteed convergence for finding a zero of a function*, Computer J. 14 (1972), 422–425.
4. H. Cornelius, *Ableitungsfreie Iterationsverfahren zur global-konvergenten Nullstelleneinschließung*, Beitr. Numer. Math. 11 (1983), 7–19.
5. J. Herzberger, *Global konvergente Interpolationsmethoden zur Nullstelleneinschließung*, Beitr. Numer. Math. 7 (1979), 65–74.
6. R. F. King, *An improved Pegasus method for root finding*, BIT 13 (1973), 423–427.
7. R. E. Moore, *Methods and applications of interval analysis*. SIAM, Philadelphia 1979.
8. A. Neumaier, *Interval iteration for zeros of systems of equations*, submitted to BIT.
9. J. Traub, *Iterative Methods for Solution of Equations*, Prentice-Hall, Englewood Cliffs, N.J., 1964.