J. Austral. Math. Soc. **20** (Series B) (1978), 157–164

A NOTE ON CONTINUATION METHODS FOR THE SOLUTION OF NONLINEAR EQUATIONS

JAMES P. ABBOTT and RICHARD P. BRENT

(Received 12 September 1977)

Abstract

In this note we present a variable order continuation method for the solution of nonlinear equations when only a poor estimate of a solution is known. The method changes continuously from one which improves the global convergence characteristics to one which attains rapid convergence to a solution and proves to be more efficient than methods previously presented in [2].

1. Introduction

In a previous paper [2] we described some methods for the solution of a system of nonlinear equations when only a poor initial estimate of a zero is known. In this note we continue this work and present an adaptive method which has proved to be more efficient than those described in [2]. We have attempted to be brief and refer the interested reader to [1] for a full description of the details.

The general approach goes back to the last century but see, in particular, [3], [10], [11] and their references for the major work. Suppose we wish to find a zero x^* of the function f: $D \subset R^n \to R^n$. We embed this problem in a family of problems of the form

$$H(x(t), t) = 0, \tag{1.1}$$

where $t \in [0, \tau)$, for some $\tau > 0$. (τ may be infinite but we do not specifically distinguish this case.) The embedding is chosen so that, for $t = 0$, the solution $x(t)$ of (1.1) is known to be x_0, that is, $x(0) = x_0$, and $x(\tau)$ is the required solution x^*. We assume subsequently that $x(t)$ exists for all $t \in [0, \tau)$ and refer, for sufficient conditions, to Rheinboldt [12] for the general case and [2] and its references for particular choices of $H(x, t)$. The problem is now one of the following the solution trajectory from $x(0) = x_0$ to $x(\tau) = x^*$. By application of the chain-rule, it follows

that the solution of (1.1) satisfies the initial value problem

$$\dot{x}(t) = -\partial_x H(x,t)^{-1}\,\partial_t H(x,t), \quad x(0) = x_0. \tag{1.2}$$

The choice of H in [2] was

$$H(x,t) = f(x) - e^{-t} f(x_0), \tag{1.3}$$

whence (1.2) becomes

$$\dot{x}(t) = -J(x)^{-1} f(x), \quad x(0) = x_0, \tag{1.4}$$

where $J(x)$ is the Jacobian of f at x. In this case, under general conditions on f and x_0, $\lim_{t\to\infty} x(t) = x^*$. In [2] we developed single and multistep methods for integrating (1.4). These methods have rapid final convergence to x^* and improve on the efficiency of some of the methods previously suggested. However, the methods of [2] make no use of the fact that the solution of (1.4) also satisfies (1.1) and in this note we describe a method which uses this relation to advantage.

2. An adaptive Newton method

As in [2] we do not wish to follow $x(t)$ with high accuracy since following $x(t)$ is only a means to finding x^*. For efficiency we require a low order method for integrating (1.4) which is stable in the sense that an error over one step of the integration will not adversely affect subsequent steps. This is the case if a method has a positive order, in the sense of Henrici [8].

DEFINITION [8]. If the initial value problem

$$\dot{x}(t) = g(x,t), \quad x(0) = x_0, \tag{2.1}$$

is solved by the iterative process

$$x_{i+1} = G(x_i, t_i, h_i), \quad i = 0, 1, \ldots, \tag{2.2}$$

where x_i is an approximation to $x(t_i)$ and $h_i = t_{i+1} - t_i$, then method (2.2) has order r for (2.1) if

$$G(x,t,h) = z(t+h) + O(h^{r+1}),$$

where $z(u)$ is the solution of

$$\dot{z}(u) = g(z,u), \quad z(t) = x.$$

(By $a = b + O(\delta)$ we mean that there is a constant K, independent of δ, such that $\|a-b\| \leqslant K|\delta|$ for all sufficiently small δ.)

If $x_i \neq x(t_i)$ then, at the $(i+1)$st stage, a method with positive order tries to follow the solution of

$$\dot{z}(t) = g(z,t), \quad z(t_i) = x_i,$$

which is different from $x(t)$. Now, when $g(x,t)$ is given by

$$g(x,t) = -J(x)^{-1}f(x),$$

then, under general conditions on f and x_i, $z(t)$ *still converges to* x^*. This is because (1.4) is stable, in the Liapunov sense, at x^* (see [2] for details). Therefore, the fact that $x_i \neq x(t_i)$ is not necessarily detrimental.

Consider (2.2) with $G(x,t,h)$ given by

$$G(x,t,h) = p_m(x,t,h), \tag{2.3a}$$

where

$$p_0(x,t,h) = x \tag{2.3b}$$

and

$$p_{j+1} = p_j - \partial_x H(p_j, t+h)^{-1} H(p_j, t+h), \quad j = 0, 1, \ldots, m-1. \tag{2.3c}$$

(For brevity we shall often omit the arguments x, t and h from $p_j(x,t,h)$.) This is an obvious method which has been suggested many times for other choices of H (see [2], [11] and the references therein) and is simply m steps of Newton's method for finding x_{i+1} by solving $H(z, t_{i+1}) = 0$, using x_i as initial guess. With $H(x,t)$ given by (1.3), (2.3c) becomes

$$p_{j+1} = p_j - J(p_j)^{-1}(f(p_j) - e^{-t-h}f(x_0)).$$

The resulting method does not have a positive order (although it does satisfy a different order condition [1]) but we can modify it to suit our needs. We note, using (1.1) and (1.3), that $x(t)$ satisfies

$$f(x(t+h)) = e^{-t-h}f(x_0) = e^{-h}f(x(t)).$$

This suggests the iterative process given by

$$x_{i+1} = G(x_i, h_i) \tag{2.4}$$

where

$$G(x,h) = p_m(x,h), \tag{2.5a}$$

$$p_0(x,h) = x \tag{2.5b}$$

and

$$p_{j+1} = p_j - J(p_j)^{-1}(f(p_j) - e^{-h}f(x)), \quad j = 0, 1, \ldots, m-1. \tag{2.5c}$$

This method has order $2^m - 1$ for (1.4) as we now show.

THEOREM 2.1. *Suppose that* f: $D \subset R^n \to R^n$ *has a derivative* $J(x)$ *which is Lipschitz continuous in a convex neighbourhood* S *of a point* $x \in \mathrm{Int}(D)$. *Assume also that* $J(x)^{-1}$ *exists and is bounded on* S. *Then the method given by* (2.4) *and* (2.5) *has order* $2^m - 1$ *for* (1.4).

PROOF. With the given conditions the Implicit Function Theorem ensures the existence of a unique continuous solution, $z(h) \in S$, of

$$f(z(h)) - e^{-h} f(x) = 0, \quad z(0) = x,$$

and therefore of

$$\dot{z}(h) = -J(z)^{-1} f(z), \quad z(0) = x, \tag{2.6}$$

and $h \in (-\delta, \delta)$ for some $\delta > 0$. We require to show that $\| z(h) - G(x,h) \| = O(h^{2^m})$ and, noting that

$$\| z(h) - G(x,h) \| = \| z(h) - p_m(x,h) \|,$$

we derive the required result by induction. Define α_j, for $j = 0, 1, \ldots, m$ by

$$\alpha_j = \| z(h) - p_j \|.$$

Also, for any $w, y \in S$, define $u(w,y)$ by

$$f(w) = f(y) + J(y)(w-y) + u(w,y). \tag{2.7}$$

Then, with the given conditions and from [11, Theorem 3.2.5], it follows that there exist constants K_1 and K_2 such that

$$\| J(w)^{-1} \| \leqslant K_1, \quad \| u(w,y) \| \leqslant K_2 \| w - y \|^2, \tag{2.8}$$

for all $w, y \in S$.

Now from (2.6), $\| \dot{z}(h) \|$ is bounded for $h \in (-\delta, \delta)$, and it follows from [11, Theorem 3.2.3] that

$$\alpha_0 = \| z(h) - x \| = \| z(h) - z(0) \| = O(h).$$

So, for small enough h, p_0 is contained in an open sphere $S(h) \subset S$, centred at $z(h)$. We assume that, for some j, $p_j \in S(h)$ and that $\alpha_j = O(h^{2^j})$ and note that these are true for $j = 0$. Now

$$\alpha_{j+1} = \| z(h) - p_{j+1} \|$$

and from (2.5c) we have

$$\alpha_{j+1} = \| z(h) - p_j + J(p_j)^{-1} (f(p_j) - e^{-h} f(x)) \|.$$

But $f(z(h)) = e^{-h} f(x)$ and so

$$\alpha_{j+1} = \| z(h) - p_j + J(p_j)^{-1} (f(p_j) - f(z(h))) \|$$

which, from (2.7), gives

$$\alpha_{j+1} = \| J(p_j)^{-1} u(z(h), p_j) \|.$$

Now, from (2.8), we have

$$\alpha_{j+1} \leqslant K_1 K_2 \alpha_j^2$$

and so, for small enough h, $p_{j+1} \in S(h)$. Also $\alpha_{j+1} = O(h^{2^{j+1}})$ and the result follows by induction. This completes the proof.

This result indicates the order of accuracy of the method in following the solution of (1.4) but now we consider the method as an iterative scheme for finding x^*. The process (2.4) is of the type discussed in [1] and [2]. For brevity here we assume that the step size, h_i, varies at each iteration until, for some i_0, $h_i = h^*$ for each $i > i_0$. This assumption is not necessary, however it is what we do in practice. Now we can use the results of [11, section 10.1] which show that (2.4) converges linearly to x^* if $0 < \rho(\partial_x G(x^*, h^*)) < 1$, where $\rho(\cdot)$ denotes the spectral radius, and (2.4)+(2.5) can converge at least quadratically to x^* only if $\partial_x G(x^*, h^*) = 0$. Some simple algebra shows that, for (2.5),

$$\partial_x G(x^*, h) = e^{-h} I$$

and so $\rho(\partial_x G(x^*, h)) = e^{-h} > 0$, for all h. Thus, local convergence to x^* is linear for all $h > 0$. This is unsatisfactory and so we modify (2.5c) to

$$p_{j+1} = p_j - J(p_j)^{-1}(f(p_j) - \phi(h) f(x)), \quad j = 0, 1, \ldots, m-1, \tag{2.9}$$

where ϕ: $R \to R$, is a function which is an approximation to e^{-h}. We first give a result on the order of this method for following the solution of (1.4).

THEOREM 2.2. *Suppose* ϕ: $R \to R$ *is continuous and* $\phi(h) = e^{-h} + O(h^{k+1})$, $k \geqslant 0$. *Then, under the conditions of Theorem* 2.1, *the method given by* (2.4), (2.5a, b) *and* (2.9) *has order* $\min(2^m - 1, k)$ *for* (1.4).

We omit the proof since it is similar to that of Theorem 2.1 and can be found in [1].

Now, from (2.9),

$$\partial_x G(x^*, h) = \phi(h) I$$

so we can achieve quadratic convergence to x^* if ϕ and h^* are chosen so that $\phi(h^*) = 0$. In order to maintain a certain order, r say, for following the solution of (1.4), a suitable choice for $\phi(h)$ is

$$\phi(h) = \sum_{j=0}^{k} \frac{(-h)^j}{j!}$$

where $k = r$ or $r+1$ is chosen to be odd, for then $\phi(h)$ satisfies the conditions of Theorem 2.2 and has a unique positive root. A practical algorithm is therefore to allow the step-sizes, h_i, to increase subject to suitable step length tests, and finally to hold each h_i fixed at h^*, where $\phi(h^*) = 0$. If k is large enough, the order of the method will be $2^m - 1$ and the sequence $\{x_i\}$ will converge rapidly to x^*. The method changes in a continuous way from one which follows the solution trajectory $x(t)$ accurately to one which converges rapidly to x^*. The final method, when $h = h^*$, is simply Newton's method.

3. Practical results

We tested the above method and compared the results with those described in [2]. For a fair comparison we employed identical criteria for step-size changing as used for the method RK3 in [2]. The methods in [2] were fixed order whereas a major advantage of the above method is the ease with which we can vary its order, simply by changing m. As might be expected, we found that varying the order at each iteration improved efficiency and it is for this implementation that the results are given. To compare with the order 3 methods in [2] we chose m, at each iteration, so that the order was at most 3 but stopped iterations as soon as the relevant test [2] indicated that the step-size would be doubled at the subsequent step.

TABLE 3.1
Method using (2.9)

Problem	1	2	3	4	5	6	7	8
Jacobian evaluations	10	15	6	26	15	8	27	29
Function evaluations	11	18	7	28	16	9	31	33
Equivalent function evaluations	31	48	19	80	61	57	112	120

The method was tested on the 8 problems in [2] and Table 3.1 gives the effort required to reduce each component of f to less than 10^{-6}. The first line gives the number of Jacobian evaluations, the second gives the number of function evaluations and the third gives the number of equivalent function evaluations, counting each Jacobian evaluation as n function evaluations (except for problems 7 and 8 where the Jacobian is tridiagonal and its evaluation is counted as 3 function evaluations). We note that these results represent a significant improvement over the results in [2], and therefore upon those in, for example, [4], [5], [6], [9], [10] which do not make special use of the characteristics of the solution of (1.2).

We also note that to evaluate $J(p_j)$ in (2.9) for each j may be inefficient. In particular, if $J(x)$ is evaluated only once per step and (2.9) replaced by

$$p_{j+1} = p_j - J(x)^{-1}[f(p_j) - \phi(h)f(x)], \quad j = 0, 1, \ldots, m-1, \tag{3.1}$$

then the resulting method has order m for (1.4). (This result is extended and proved in [1].) For example, to achieve order 3, this method requires 3 function and 1 Jacobian evaluation whereas with (2.9) the method requires 2 function and 2 Jacobian evaluations. So, for small h at least, using (3.1) is theoretically more efficient than using (2.9). Table 3.2 gives the results of the method with (3.1) replacing (2.9). The table shows that this method is generally the most efficient. The only notable exceptions were when the solution of (1.4) ran close to a region where $J(x)$ was singular, as in problem 4, and so $J(x)^{-1}$ was changing rapidly.

It seems reasonable to monitor $\mathrm{Det}(J(x))$, which can be done at little cost since $J(x)$ is factorized into triangular factors, and evaluate $J(x)$ more often only when $\mathrm{Det}(J(x))$ becomes small.

TABLE 3.2
Method using (3.1)

Problem	1	2	3	4	5	6	7	8
Jacobian evaluations	8	8	6	36	9	7	14	15
Function evaluations	11	29	7	134	20	9	61	65
Equivalent function evaluations	27	45	19	206	56	51	113	110

Finally, we note that this work is extended in [1] where tests are also performed on other continuation methods, for example, Branin's method [7] and the methods which result from choosing $H(x,t)$ as

$$H(x,t) = f(x) - (1-t)f(x_0).$$

(See [11].) The results indicate that the new methods are more reliable and efficient. In particular, for the new methods the step control is easier and they have a greater chance of success in more difficult problems.

REFERENCES

[1] J. P. Abbott, "Numerical continuation methods for nonlinear equations and bifurcation problems" (Ph.D. Thesis, Computer Centre, Australian National University, 1977).
[2] J. P. Abbott and R. P. Brent, "Fast local convergence with single and multistep methods for nonlinear equations", *J. Austral. Math. Soc.* 19 (Series B) (1975), 173–199.
[3] J. H. Avila, "The feasibility of continuation methods for nonlinear equations", *SIAM J Numer. Anal* 11 (1974), 102–122.
[4] L. Bittner, "Einige kontinuierliche Analogien von Iterationsverfahren", in *Functional-analysis, Approximationstheorie Numerische Mathematik*, ISNM 7 (Birkhauser Verlag, Basel, 1967), pp. 114–135.
[5] P. T. Boggs, "The solution of nonlinear systems of equations by A-stable integration techniques", *SIAM J. Numer. Anal.* 8 (1971), 767–785.
[6] W. E. Bosarge, "Iterative continuation and the solution of nonlinear two point boundary value problems", *Numer. Math.* 17 (1971), 268–283.
[7] F. H. Branin Jr., "Widely convergent method for finding multiple solutions of simultaneous nonlinear equations", *IBM J. Res. Develop.* 16 (1972), 504–522.
[8] P. Henrici, *Discrete variable methods for ordinary differential equations* (John Wiley, New York, 1962).
[9] W. Kizner, "A numerical method for finding solutions of nonlinear equations", *SIAM J. Appl. Math.* 12 (1964), 424–428.
[10] H. Kleinmichel, "Stetige Analoga und Iterationsverfahren für nichtlineare Gleichungen in Banachräumen", *Math. Nachr.* 37 (1968), 313–344.

[11] J. Ortega and W. Rheinboldt, *Iterative solution of nonlinear equations in several variables* (Academic Press, New York, 1970).
[12] W. C. Rheinboldt, "Local mapping relations and global implicit function theorems", *Trans. Amer. Math. Soc.* 138 (1969), 183–198.

Computer Centre
Australian National University
Canberra 2600
Australia