On the optimization of step size in the continuation method

Article · June 2018

1 author:

Marijan Ribarič
Jožef Stefan Institute

MATHEMATICA BALKANICA
4.97 (1974) 517—521

RIBARIČ M., SELIŠKAR M. (Ljubljana, Yugoslavia)

ON OPTIMIZATION OF STEPSIZE IN THE CONTINUATION METHOD*

The purpose of this paper is to find an optimal strategy for solving functional equations by iterative methods. Our work is based on the majorant method which is a widely used tool in the theory of iterative equation-solving methods. It was introduced in the year 1949 by KANTOROVIČ and applied to prove the convergence of Newton's method in B — space (cf. KANTOROVIČ [1]). The majorant method enables us to study convergence of a given abstract iterative process by studying the appropriate difference equation, whose solution is the error sequence of the iterative process considered.

Suppose we have a functional equation in B-space, say $T(s, x) = 0$, depending on parameter $s \in [0, L]$ from whose solution at $s = 0$ we would like to derive the solution at $s = L$ by the continuation method (cf. ORTEGA and RHEINBOLDT [2]). Let us denote by $E_n, n = 0, 1, \ldots, N$, the estimated errcr at the n-th step of the continuation method (i. e. at $s = s_n$), then

(1) $$E_0 \geqslant 0,^{1)} \quad E_{n+1} = \psi(E_n, s_{n+1} - s_n, s_n), \qquad n = 0, 1, \ldots, N-1,$$

where $\psi(\ldots, \ldots, \ldots)$ is an estimate of the absolute value of the difference between the exact and approximate solution of the given functional equation at the intermediary point $s = s_n$, and N is the number of intermediary points

(2) $$0 = s_0 < s_1 \leqslant s_2 \leqslant s_N = L$$

In many practical applications it turns out that the estimated erroı function ψ is of the folowing form

(3) $$\psi(u, v, s) = [f(s)\, u + a(s)\, u^2 + g(s)\, v + b(s)\, v^2]^{\varkappa},$$

where the exponent $\varkappa > 1$ and $f(s)$, $a(s)$, $g(s)$ and $b(s)$ are nonnegative bounded functions of $s \in [0, L]$. In the case when the continuation method is based upon Newton's method we have $\varkappa = 2$ (cf. ORTEGA and RHEINBOLDT [2]). Let us denote by $S(N, L)$ the set of all possible sequences (2) of intermediary points. In general for any $\varepsilon > 0$ there are many sequences for which the corresponding estimated end-point errcr $E_N \leqslant \varepsilon$. So the following optimization problem presents itself:

For a given $\varepsilon > 0$ find the smallest integer N such that there is at least one sequence $\{s_n\} \in S(N, L)$ of intermediary points for which the corresponding estiamted end-point error $E_N \leqslant \varepsilon$; let us denote this integer as $N_{op}(\varepsilon)$, i. e.

(4) $$N_{op}(\varepsilon) = \min \{N \mid \exists \{s_n\} \in S(N, L),\ E_N \leqslant \varepsilon\},$$

---

* Presented at the 5th Balkan Mathematical Congress (Beograd, 24—30. 06 1974)

1) $E_0 = 0$ when we know the exact solution of the functional equation at $s = 0$.

In general there are many sequences of $N_{op}(\varepsilon)$ *intermediary points such that the corresponding estimated end-point errore* $E_{Nop(\varepsilon)}$ *is not greater than* ε. *Now for this integer* $N_{op}(\varepsilon)$ *we would like to find an optimal sequence of intermediary points, say* $\{s_n\}_{op} \in S(N_{op}(\varepsilon), L)$, *such that the corresponding estimated end-point error, say* $E_{op}(\varepsilon)$, *will be minimal, i. e.*

$$(5) \qquad E_{op}(\varepsilon) = N_{Nop(\varepsilon)\min} \equiv \min\{EN_{op(\varepsilon)} \mid \{s_n\} \in S(N_{op}(\varepsilon), L)\}.$$

We consider this sequence $\{s_n\}_{op}$ *as the solution to the stated optimization problem,*

The purpose of the paper is to show that we can solve the formulated optimization problem by solving the following auxiliary problem:

For a given integer N *find such a sequence* $\{s_n\}_{op} \in S(N, L)$ *that the corresponding estimated end-point error* E_N *is minimal, i. e.*

$$(6) \qquad E_N = E_{N\min}.$$

We have formulated our auxiliary problem so that for a given $\varepsilon > 0$ the solution to the optimization problem posed is at the same time the solution of the auxiliary problem for the case $N = N_{op}(\varepsilon)$. In the special case when the function ψ of (3) is a function of only the first two variables, i. e. when

$$(7) \qquad \psi(u, v) = (fu + au^2 + gv + bv^2)^{\varkappa},$$

where f, a, g, b and $\varkappa$ are non-negative constants such that

$$(8) \qquad f + a > 0, \quad g + b > 0 \quad \text{and} \quad \varkappa > 1,$$

we will show how we can solve the posed optimization problem indirectly by solving the stated auxiliary problem. To this end we will need some simple properties of the function $\psi(u, v)$, collected in the following

Lemma 1. Function $\psi(u, v)$, defined by relations (7), (8), is a continuos and strictly increasing function of two variables such that $\lim\limits_{u+v\to\infty} \psi(u, v) = \infty$.

Equation

$$(9) \qquad \psi(z, 0) = z, \qquad z > 0,$$

has a unique solution, say $E_{cr'}$ and we have

$$(10) \qquad \psi(z, 0) \begin{cases} = 0 & \text{if } z = 0, \\ \in (0, z) & \text{if } z \in (0, E_{cr}), \end{cases} \qquad \psi(z, 0) \begin{cases} = z & \text{if } z = E_{cr}, \\ \in (z, \infty) & \text{if } z \in (E_{cr}, \infty). \end{cases}$$

Furthermore, for any initial term, say $e_0 \geqslant 0$, the recursively defined sequence

(11) $e_{n+1} = \psi(e_n, 0)$, $n = 0, 1, \ldots$ has the following properties

$e_n = 0 \ \forall n = 0, 1, \ldots$ if $e_0 = 0$, $\quad e_n > e_{n+1} \ \forall n = 0, 1, \ldots,$ and $\lim\limits_{n\to\infty} e_n = 0$ if $e_0 \in (0, E_{cr})$; $\quad e_n = E_{cr} \ \forall n = 0, 1, \ldots$ if $e_0 = E_{cr}$; $\quad e_n < e_{n+1} \ \forall n = 0, 1, \ldots$

$$(12) \qquad \text{and} \quad \lim_{n\to\infty} e_n = \infty, \quad \text{if } e_0 \in (E_{cr}, \infty).$$

Proof. It follows from (7) that

$$\frac{\psi(z, 0)}{z} = z^{\varkappa-1}(f+az)^{\varkappa}, \quad z>0. \tag{13}$$

By (8) we have that (13) defines a continuous function, increasing from 0 to ∞ as z moves from 0 to ∞. Hence the equation (9) has a unique solution, say E_{cr}, such that the relations (10) are valid.

Let us prove the validity of relations (12). In the case when either $e_0=0$ or $e_0=E_{cr}$ relations (12) are an immediate consequence of (10). Now, suppose that

$$e_0 \in (0, E_{cr}). \tag{14}$$

Using relation (10) we see from (14, 11) that

$$e_n > e_{n+1} \text{ and } e_n \in (0, E_{cr}), \qquad n=0, 1, \ldots, \tag{15}$$

implying that the sequence $\{e_n\}$ is convergent, where

$$e^* = \lim_{n\to\infty} e_n \in [0, E_{cr}). \tag{16}$$

The function ψ being continuous the relations (16, 11) imply that

$$e^* = \psi(e^*, 0). \tag{17}$$

As $e^* \in [0, E_{cr})$ we conclude from (17), using relations (10) once again, that $e^*=0$. So we have proved relations (12) for the case when (14). Using similar arguments as above we can also prove relations (12) for the case when $e_0 \in (E_{cr}, \infty)$. Q. E. D.

Using the above properties of ψ, we show in the following lemma that by choosing a sufficiently large number N of intermediary points, it is possible to make the corresponding estimated end-point error E_N as small as desired; a fact which implies the existence of $N_{op}(\varepsilon)$ for any $\varepsilon>0$.

Lemma 2. For a given initial estimated error

$$E_0 \in [0, E_{cr}) \tag{18}$$

and any $\varepsilon>0$ there is an integer N and a sequence of intermediary points $\{s_n\} \in S(N, L)$ such that the corresponding estimated end-point error $E_N \leqslant \varepsilon$.

Proof. Let us choose a positive constant, say

$$e \in (E_0, E_{cr}). \tag{19}$$

Relations (10) and $e<E_{cr}$ imply that $\psi(e, 0)<e$. Thence and from properties of function ψ given in lemma 1 we infer that the equation

$$\psi(e, z)=e, \qquad z>0, \tag{20}$$

has a unique solution, say h, such that

$$0 \leqslant \psi(u, v) < e \quad \text{if} \quad (u, v) \in [0, e) \times [0, h]. \tag{21}$$

Now let us define the sequence

$$s_0=0; \qquad s_n=s_{n-1}+h, \quad (n=1, 2, \ldots, M-1); \qquad s_M=L, \tag{22}$$

where M is the largest integer such that

(23) $$S_{M-1} < L.$$

Note that

(24) $$s_n - s_{n-1} \leqslant h, \qquad n = 1, 2, \ldots, M.$$

Using the relation (21) we infer by induction from $E_0 < e$ and from (24), (1) with $n = 1, 2, \ldots, M$ that the estimated end-point error

(25) $$E_M < e.$$

Thence and from $e < E_{cr}$ we conclude that $E_M < E_{cr}$. This result together with lemma 1 shows that by making a sufficient number, say $N - M$, of additional iterations (1) at the end-point $s = L$ we can achieve an arbitrary small estimated end-point error E_N. This proves the lemma. Q. E. D.

For our special case (7), (8) we can show that the auxiliary problem has a unique solution for any integer N as witnessed by the following.

Lemma 3. For any integer N there is only one sequence $\{s_n\} \in S(N, L)$ such that by (1), (7), (8) the corresponding estimated end-point error E_N is minimal.

Proof. Taking an integer N and using recursive relation (1) we can calculate explicitly the estimated end-point error E_N for any sequence $\{s\} \in S(N, L)$ of intermediary points. The initial and the end-point s_N being the same for any sequence $\{s_n\} \in S(N, L)$, the estimated end-point error E_N depends only on intermediary points $s_1, s_2, \ldots, s_{N-1}$. We consider therefore estimated end-point error

(26) $$E_N = E_N(s_1, s_2, \ldots, s_{N-1})$$

as a function of $N-1$ variables s_n, defined on the convex compact subset

(27) $$D_N = \{(s_1, s_2, \ldots, s_{N-1}) \in R^{N-1} \mid 0 \leqslant s_1 \leqslant s_2 \leqslant \cdots \leqslant s_{N-1} \leqslant L\}.$$

The properties of the function ψ as given by (7) and (8) imply that function E_N is a strictly convex and continuous function on the convex compact subset D_N and so it has there only one minimum say $(\bar{s}_1, \bar{s}_2, \ldots, \bar{s}_{N-1})$; at this minimum the following conditions have to be satisfied:

(28) $$\frac{\partial E_N(\bar{s}_1, \bar{s}_1, \ldots, \bar{s}_{N-1})}{\partial s_n} \begin{cases} = 0 & \text{if } \bar{s}_{n-1} < \bar{s}_n < \bar{s}_{n+1}, \\ \leqslant 0 & \text{if } \bar{s}_{n-1} < \bar{s}_n = s_{n+1}, \\ \geqslant 0 & \text{if } \bar{s}_{n-1} = \bar{s}_n < \bar{s}_{n+1}. \end{cases}$$

As N is an arbitrary integer the above demostrates that the auxiliary problem has a unique solution for any integer N. Q. E. D.

Considering the posed optimization problem, the preceeding results enable us for any $\varepsilon > 0$ to determine the corresponding optimal number $N_{op}(\varepsilon)$ and the associated optimal sequence $\{s_n\}_{op}$ of $N_{op}(\varepsilon)$ intermediary points for which $E_{Nop}(\varepsilon)$ is minimal. To this end we proceed as follows. Given $\varepsilon > 0$ we define two sets

(29) $\quad A_1 \equiv \{N \mid \exists \{s_n\} \in S(N, L),\ E_N \leqslant \varepsilon\}$ and

(30) $\quad A_2 \equiv \{N \mid E_{N\,\min} \leqslant \varepsilon\}$. Thence and from (4) we see that

(31) $$N_{op}(\varepsilon) = \min A_1.$$

Using lemmas 2 and 3 we infer that

(32) $$A_1 = A_2 \neq \varnothing \quad \text{implying}$$

(33) $N_{op}(\varepsilon) = \min\{N \mid E_{N\,\min} \leqslant \varepsilon\}$. We sumarize these results as

Theorem 1. *When the estimated initial error*

(34) $$E_0 \in [0, E_{cr}),$$

then for any $\varepsilon > 0$ *there is exactly one solution of the posed optimitization problem which solves also the auxiliary problem for the integer*

(35) $$N_{op}(\varepsilon) = \min\{N \mid E_{N\,\min} \leqslant \varepsilon\}.$$

This theorem brings forth mathematically the sense in which the posed optimization problem and the the stated auxiliary problem are equivalent. This if of importance for practical applications since the solutions of the auxiliary problem can be obtained explicitly by solving the system of equations corresponding to the conditions (28).

References

[1] Kantorovič, L.: *On Newton's method. Trudy Mat. Inst. Steklov* **28,** 1949.

[2] Ortega, J. M., Rheinboldt, W. C.: *Iterative solution of nonlinear equations in several variables.* Academic Press. New York and London, 1970.

Acknowledgement. The autors thank referee for his helpfull remarks.

(Stiglo 22. 07. 1974. resp. 28. 11. 1974.).

Dr Marjan Ribarič
Mgr Marjan Seliškar
Institute J. Stefan,
University of Ljubljana, Jamova 39
61 000 Ljubljana, Jugoslavija, Telefon: 63-261

4.98 PROBLEM*

Es sei $B = \{a^2 + b^2 : a \in Z \wedge b \in Z\}$, also $B = \{0, 1, 2, 4, 5, 8, 9, 10, \ldots\}$. Nach Hooley (1973) gibt es nach Vorgabe von ganzen Zahlen $0 < h < k$ unendlichviele $n \in B$ mit $n + h \in B$ und $n + k \in B$. Man gebe eine nichttriviale untere Abschätzung für die Anzahl $A_{h,k}(x)$ solcher $n \leqslant x$. Nach der Siebmethode ist übrigens $A_{h,k}(x) \ll x(\log x)^{-3/2}$ $(x \to \infty)$. In Verallgemeinerung des obigen Problems der Drillinge in B unteruche man Vierlinge in B und stelle der oberen Abschätzung $x(\log x)^{-2}$ eine untere an die Seite.

* Dargelegt am 28. 06. 1974. auf der Sitzung über math. Probleme (5. BalkanMathematiker-Kongress, Beograd, 24—30. 06. 1974).

RIEGER, G. J.
Mathematisces Institut d. T. H.
Hannover, DBR