[4] W. Kahan, *A survey of error analysis*, IFIP 1971, I, pp. 200–206.
[5] J. Stoer, *Einführung in die numerische Mathematik I*, Springer Verlag, 1972.
[6] V. V. Voevodin, *Rounding errors and stability* (in Russian), MGU, Moscow 1969.
[7] J. H. Wilkinson, *Rounding errors in algebraic processes*, London 1963.

Presented to the Semester
Mathematical Models and Numerical Methods
(February 3–June 14, 1975)

BANACH CENTER PUBLICATIONS
VOLUME 3

MULTIVARIATE SECANT METHOD

JANINA JANKOWSKA

Institute of Informatics, University of Warsaw, Warsaw, Poland

We consider the problem of solving a system of nonlinear equations

(1) $$f(x) = 0$$

for $f\colon D \to C^n$, where C^n denotes the n-dimensional complex space and D is an open and convex set in C^n. We assume that f satisfies the following two conditions,

(2)
(i) there exists a simple zero $\alpha = \alpha(f) \in D$;
(ii) $f'(x)$ is a Lipschitz function in D.

We solve (1) by the multivariate secant method—shortly the MS-method—defined as follows. Let $x_i, \ldots, x_{i-n} \in D$ be approximations of α. If the matrices

$$X_i = [\delta x_{i-n}, \ldots, \delta x_{i-1}], \quad F_i = [\delta f_{i-n}, \ldots, \delta f_{i-1}],$$

where $\delta x_j = x_{j+1} - x_j$, $\delta f_j = f_{j+1} - f_j$, $f_j = f(x_j)$, are nonsingular then the next approximation of α in the MS-method is given by the formula:

(3) $$z_i = \varphi(x_i; f) = x_i - X_i \cdot F_i^{-1} \cdot f_i.$$

We can put $x_{i+1} = z_i$ or define x_{i+1} otherwise.

The problems of our interest are,

(i) the convergence and the character of convergence of the MS-method,
(ii) the numerical stability of a chosen algorithm of the MS-method.

For the first problem we got the following result. Let us define

$$d_i = \left|\det\left[\frac{\delta x_{i-n}}{\|\delta x_{i-n}\|}, \ldots, \frac{\delta x_{i-1}}{\|\delta x_{i-1}\|}\right]\right| \quad (d_i \leqslant 1),$$

(4) $$\mathfrak{M}(c, \xi) = \{(x_n, x_{n-1}, \ldots, x_0)\colon x_j \in C^n, d_n \geqslant c\|x_n - x_0\|^{\xi}\},$$

where $\xi \in [0, 1)$, $c \in (0, 1]$.

Then we have

THEOREM 1. *Let*

(i) *f satisfy* (2),

(ii) $x_0, \ldots, x_n \in D$: $\|x_n-\alpha\| \leqslant \|x_j-\alpha\| \leqslant \|x_0-\alpha\|$, $j = 1, \ldots, n-1$.

If $\forall i \geqslant n$, $(x_i, \ldots, x_{i-n}) \in \mathfrak{M}(c, \xi)$ *for fixed* $\xi \in [0, 1)$ *and* $c \in (0, 1]$, *then for sufficiently small* $\|x_0-\alpha\|$ *we have*:

(i) *the sequence* $\{x_i\}_{i=0}^{\infty}$, *where* $x_{i+1} = \varphi(x_i; f)$, *is well-defined and satisfies*

$$\|x_{i+1}-\alpha\| < \|x_i-\alpha\| \quad \text{and} \quad \lim_{i\to\infty} \varphi(x_i, f) = \alpha,$$

(ii) $\|x_{i+1}-\alpha\| \leqslant \dfrac{K_1(f)}{c} \|x_{i-n}-\alpha\|^{1-\xi}\|x_i-\alpha\|$, *where* K_1 *depends only on* f.

One can prove that the inequality (ii) is sharp. Hence, if all succesive points $(x_i, \ldots, x_{i-n})$ are in a good position, i.e. if they belong to $\mathfrak{M}(c, 0)$, then the order of convergence p_n of the MS-method is equal to the unique positive zero of the polynomial $t^{n+1}-t^n-(1-\xi)$.

The order $p_n(\xi)$ is a decreasing function of $\xi \in [0, 1)$, $\max p_n(\xi) = p_n(0)$. The results due to Bittner [2], Barnes [1] and probably others (see Ortega–Rheinboldt), involve the case $\xi = 0$.

We know that the assumption of good position of the points $(x_i, \ldots, x_{i-n})$ is necessary for any iteration which uses the same information on f, see Woźniakowski [7]. Furthermore, it is possible to prove that the MS-method makes the optimal use of the information on f with respect to $\mathfrak{M}(c, \xi)$ and the order $p_n(\xi)$ is as high as possible.

From this theorem it follows that any algorithm of the MS-method should involve a certain control of the d_i values. If d_i is too small, one has to redefine the points $(x_i, \ldots, x_{i-n})$ to ensure that the new d_i is sufficiently large. For instance, if $x_{i-j} = x_{i-j+1}+e_j\|f_i\|$, e_j the jth axis unit vector, $j = 1, 2, \ldots, n$, then $d_i = 1$.

As regards the second problem, our result concerns the case of $\xi = 0$ and the following algorithm of calculation of the z_i from (3) in t-digit, floating point arithmetic fl, is proposed.

A-algorithm

z_i: $F_i \cdot z_i = f_i$;

Here it is assumed that we use a numerically well-behaved algorithm for the solution of the linear system satisfying the following condition. If w_i is the computed solution in fl-arithmetic then there exists a matrix E_i such that

$$(F_i+E_i)\cdot w_i = f_i$$

and for every column E_j^i of E_i

$$\|E_j^i\| \leqslant 2^{-t}K_E \cdot \|\delta f_j\|.$$

$p_i := X_i \cdot z_i$;

$x_{i+1} := x_i - p_i$;

We also assume that f depends on a so-called data-vector $d \in C^m$: $f(x) = f(x; d)$, and the computed value of f in fl-arithmetic satisfies

(5) $$\mathrm{fl}(f(x; d)) = [I-\Delta f(x; d)] \cdot f(x+\Delta x; d+\Delta d), \quad \forall x \in D,$$

where

$$\|\Delta f(x; d)\| \leqslant K_f \cdot 2^{-t}, \quad \|\Delta x\| \leqslant K_x \cdot 2^{-t} \cdot \|x\|, \quad \|\Delta d\| \leqslant K_d \cdot 2^{-t} \cdot \|d\|$$

and the nonnegative constants K_f, K_x, K_d do not depend on x, d or t.

(Condition (5) means that the algorithm used for the evaluation of $f(x)$ is well-behaved. See Kiełbasiński [5] and Woźniakowski [8].)

We consider a system of nonlinear equations

(6) $$f(x; d) = 0,$$

where the data-vector d belongs to a close neighbourhood S of d_α.

We assume that for every $d \in S$, the system (6) has a simple zero $x(d) \in D$ and $x(d_\alpha) = \alpha$. One can prove that for a sufficiently regular function f the condition number for the equation (6) is given by the formula

$$\mathrm{cond}(f; d) = \|[f_x'(\alpha; d_\alpha)]^{-1} f_d'(\alpha; d_\alpha)\| \frac{\|d_\alpha\|}{\|\alpha\|},$$

which means that

$$\frac{\|\alpha - x(d)\|}{\|\alpha\|} \leqslant \mathrm{cond}(f; d_\alpha) \frac{\|d_\alpha - d\|}{\|d_\alpha\|}, \quad d \in S.$$

The next theorem explains the properties of the numerically computed sequence $\{x_i\}$.

THEOREM 2. *Let*

(i) *f have the above properties,*

(ii) $x_0, \ldots, x_n \in D$: $\|x_n-\alpha\| \leqslant \|x_j-\alpha\| \leqslant \|x_0-\alpha\|$ *for* $j = 1, \ldots, n-1$,

(iii) $x_{i+1} = \mathrm{fl}(\varphi(x_i; f))$—*computed by the A-algorithm.*

If $\forall i \geqslant n$

1. $(x_i, \ldots, x_{i-n}) \in \mathfrak{M}(c, 0)$,

2. $\min\limits_{0\leqslant j\leqslant n-1} \|x_{i-j}-x_{i-j+1}\| \geqslant 2^{-t} \dfrac{K_2}{c}$ cond $(f; a)$, $K_2 = K_2(n, K_x, K_d, K_f) > 0$,

then for sufficiently small $\|x_0-\alpha\|$ *we have*:

(i) $\forall i \geqslant n$ $\quad \|x_i-\alpha\| \leqslant \|x_0-\alpha\|$,

(ii) $\limsup\limits_{i\to\infty} \dfrac{\|x_i-\alpha\|}{\|\alpha\|} \leqslant 2^{-t} \dfrac{K_3}{c}$ cond $(f; d)$, *where* $K_3 = K_3(n, K_x, K_d, K_f, K_E)$ *does not depend on* 2^{-t}.

The inequality (ii) means numerical stability of the considered algorithm (cf. [8]).

Note that the good choice of the constant c is very important from the practical point of view. For a small value of c the point $(x_i, \ldots, x_{i-n})$ belongs to $\mathfrak{M}(c, 0)$ with large probability. But $\|x_i-\alpha\|$ is directly proportional to $1/c$ and for small c

it has a bad estimate. The problem of the optimal choice of c is, to our best knowledge, still open.

For a detailed discussion and the proofs of the presented and other theorems see [3] and [4].

References

[1] J. G. Barnes, *An algorithm for solving nonlinear equations based on the secant method*, Computer Journ. 8 (1965).

[2] L. Bittner, *Eine Verallgemeinerung des Sekantenverfahrens zur naherungsweisen Bererchnung der Nullstellen eines nichtlinearen Gleichungssystem s.*, Wiss. Z. Techn. Univ. Dresden 9, 1959/60.

[3] J. Jankowska, *Multivariate secant method*, Ph.D. thesis, University of Warsaw, 1975.

[4] —, *The theory of the multivariate secant method*, pending for publication in SIAM Journ. Numer. Anal.

[5] A. Kiełbasiński, *Basic concepts of rounding error analysis in numerical methods of linear algebra*, Matematyka Stosowana 4 (1975), pp. 5–27.

[6] J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equation in several variables*, Academic Press 1970.

[7] H. Woźniakowski, *Generalized information and maximal order of iteration for operator equations*, SIAM Journ. Numer. Anal. 12, No. 1, pp. 121–135.

[8] —, *Numerical stability for solving nonlinear equations*, Computer Science Department report, Carnegie-Mellon University, Pittsburgh, Pa., 1975.

Presented to the Semester
Mathematical Models and Numerical Methods
(February 3–June 14, 1975)

BANACH CENTER PUBLICATIONS
VOLUME 3

МЕТОДЫ ПЕРЕНОСА ДЛЯ СИСТЕМ ЛИНЕЙНЫХ АЛГЕБРАИЧЕСКИХ УРАВНЕНИЙ С ПОЯСНЫМИ МАТРИЦАМИ

ЛЮБОР МАЛИНА

Университет им. Коменского, Институт Прикладной Математики и Вычислительной Техники, Братислава, Чехословакия

1. Введение

Пусть заданная краевая задача

(1.1) $$y''(x) = f(x) \quad \text{для} \quad x \in (a, c),$$

(1.2) $$y(a) = \alpha, \quad y(c) = \beta,$$

где $y''(x)$ значит вторую производную от функции y в точке x. Численно можно эту краевую задачу решить, например, заменой производной на конечную разность. Значит, пусть $x_i = a+ih$, где $h>0$, $i = 0(1)N$ и $x_N = c$. Тогда

$$y''(x) \approx \frac{y(x_i+h)-2y(x_i)+y(x_i-h)}{h^2}$$

и уравнение (1.1) заменяем в точке $x = x_i$ уравнением

(1.3) $$y_{i-1}-2y_i+y_{i+1} = h^2 f_i, \quad i = 1(1)N-1,$$

(1.4) $$y_0 = \alpha, \quad y_N = \beta,$$

где $y_i \approx y(x_i)$.

Т.е., вместо задачи (1.1)–(1.2) мы решаем систему линейных алгебраических уравнений (1.3)–(1.4), которую формально запишем

(1.5) $$Ly = f,$$

где L есть тридиагональная матрица, $y = [y_0, \ldots, y_N]^{\mathrm{T}}$ есть вектор незнакомых, размерности $N+1$ и f есть вектор правых частей, размерности $N+1$. Систему (1.5) можно решать, например, процессом элиминации Гаусса. Если мы хорошо отдаем себе отчет в нем, то прямой ход процесса элиминации обозначает, что мы постепенно оформляем тридиагональную матрицу L на бидиагональную верхнюю троеугольную матрицу D. Диагональные элементы этой ма-