ON NEWTON-LIKE ITERATION FUNCTIONS: GENERAL CONVERGENCE THEOREMS AND A SPECIFIC ALGORITHM

by

K.M. Brown and J.E. Dennis, Jr.

Technical Report
No. 68-22
June 1968

Kenneth M. Brown
Department of Computer Science
Cornell University
Ithaca, New York 14850

J.E. Dennis, Jr.
Dept. of Mathematics
The University of Utah
Salt Lake City, Utah 84112

ON NEWTON-LIKE ITERATION FUNCTIONS: GENERAL CONVERGENCE THEOREMS AND A SPECIFIC ALGORITHM*

by Kenneth M. Brown
Department of Computer Science
Cornell University
Ithaca, New York 14850

and

J. E. Dennis, Jr.
Department of Mathematics
The University of Utah
Salt Lake City, Utah 84112

* Accepted for publication in Numerische Mathematik

ON NEWTON-LIKE ITERATION FUNCTIONS: GENERAL CONVERGENCE THEOREMS AND A SPECIFIC ALGORITHM

KENNETH M. BROWN[1]) and J. E. DENNIS, JR.[2])

1. Introduction.

As more and more physical phenomena are being represented as nonlinear mathematical formulations, increasing attention is being paid to the solution of systems of nonlinear equations. Newton's method [11] has long been a popular tool for the solution of such systems, but computationally simpler variants of Newton's method have been proposed recently. See for example [1], [2], [3], [4], [12], [16], and [17]. In this paper we give convergence theorems for a class of these methods. We also present an algorithm suggested by the theory and include several numerical examples.

1) Department of Computer Science; Cornell University; Ithaca, N.Y. 14850

2) Department of Mathematics; The University of Utah; Salt Lake City, Utah 84112

2. Notation and Preliminaries.

Let F be defined on Ω, $\Omega \subset E^N$ and $F : \Omega \to E^N$. Let Ω_0 be the closure of an open convex set in E^N with $\Omega_0 \subset \Omega$ and let Ω_0 be bounded. We seek solutions of $F(x) = 0$ and consider the class of iterative methods

$$(1) \qquad x^{n+1} = x^n - G^{-1}(x^n) \cdot F(x^n), \qquad n = 0, 1, \dots,$$

with $G(x^n)$ an $N \times N$ matrix and $G^{-1}(x^n) \equiv [G(x^n)]^{-1}$. Define $x \equiv (x_1, \dots, x_N)$, $F \equiv (f_1, \dots, f_N)$, $f_{ij} \equiv \partial f_i / \partial x_j$ and $J(x^n) \equiv [f_{ij}(x^n)]$, the Jacobian matrix of F at x^n. If in (1) $G(x^n) = J(x^n)$, we have the standard Newton's method. As $G(x^n)$ is allowed to vary from $J(x^n)$ in a controlled manner, one has a class of "perturbed" Newton methods whose convergence will be studied in the next section.

$N(x,r)$ will denote the open ball centered at x with radius r and $\mathrm{cl}\, N(x,r)$ will be its closure.

Finally, $F''(x)$ will denote the $N \times N \times N$ array $[f^k_{ij}]$, where $f^k_{ij} \equiv \dfrac{\partial^2 f_k}{\partial x_i \partial x_j}$. $||F''(x)||_2$ is then bounded by

$$\left(\sum_{k=1}^{N} \sum_{i=1}^{N} \sum_{j=1}^{N} |f^k_{ij}|^2 \right)^{1/2} .$$

3. Convergence of a Class of Perturbed Newton Methods.

The following is a specialization of a theorem due to Dennis [7].

THEOREM 1. Let the following conditions be satisfied:

i) For every $x \in \Omega_0$, $G^{-1}(x)$ exists and $||G^{-1}(x)|| \leq B$.

ii) For every $x \in \Omega_0$, $||F''(x)|| \leq K$.

iii) $0 < ||F(x^0)|| \leq \eta$, for some $x^0 \in \Omega$.

iv) For every n such that $x^{n+1} = x^n - G^{-1}(x^n)F(x^n)$ is defined, $||I - J(x^n)G^{-1}(x^n)|| \leq \delta < 1$.

v) $h = \frac{B^2K\eta}{1-\delta} < 2$.

vi) $N(x^0,r) \subset \Omega_0$, where $r = \frac{B\eta}{1-\alpha}$ and $\alpha = \delta + (1 - \delta)\cdot\frac{h}{2}$.

Then $F(x) = 0$ has a solution σ, $||\sigma - x^0|| < r$, to which the iteration defined by equation (1) converges and the speed of convergence is given by $||x^n - \sigma|| < \frac{B\eta\alpha^n}{1-\alpha}$.

Corollary 1. Let $C = B^2K\eta$. Then hypothesis v) of Theorem 1 allows the replacement of $h' = \frac{C}{(1-\delta)^2}$ by the smaller $h = \frac{C}{1-\delta}$ in Theorem 4 of [8], a convergence theorem for the Newton-Jacobi iteration (see (2) below).

In the next two theorems $||\cdot||$ will denote the ℓ_2 norm.

THEOREM 2. Let $\lambda(x)$ be defined as the positive square root of the smallest eigenvalue of $G(x)G(x)^*$, and let $\mu(x)$ be defined as the positive square root of the largest eigenvalue of $[J(x) - G(x)][J(x) - G(x)]^*$, where $J(x)$ denotes the Jacobian of F at x.

If

i) $F \in C^2(\Omega_o)$ and $\forall x \in \Omega_o$, $||F''(x)|| \leq K$;

ii) $\exists x^o \in \Omega_o \ni ||F(x^o)|| \leq \eta$;

iii) $\lambda(x) > 0$ for every $x \in \Omega_o$;

iv) $\frac{\mu(x)}{\lambda(x)} < 1$ for every $x \in \Omega_o$;

then $G^{-1}(x)$ exists for every $x \in \Omega_o$, and

$$||G^{-1}(x)|| \leq \frac{1}{\min_{x \in \Omega_o} \lambda(x)} \equiv B < \infty .$$

Furthermore, $||I - J(x)G^{-1}(x)|| \leq \max_{x \in \Omega_o} \frac{\mu(x)}{\lambda(x)} \equiv \delta < 1$.

Proof. Zero is not an eigenvalue of $H(x) \equiv G(x)G(x)^*$; hence $H(x)$ is invertible and $G^{-1}(x) = G(x)^* H^{-1}(x)$. Thus $||G^{-1}(x)|| = 1/\lambda(x)$. (See for example [6, p. 173].) Now $\lambda(x)$ is continuous (in x) on the compact set Ω_o and hence $\min_{x \in \Omega_o} \lambda(x)$ exists and is positive since zero is not in the range of λ ; but $||J(x) - G(x)|| = \mu(x)$, so that

$$||I - J(x)G^{-1}(x)|| = ||[G(x) - J(x)] \cdot G^{-1}(x)||$$
$$\leq \frac{\mu(x)}{\lambda(x)} \leq \max_{x \in \Omega_o} \frac{\mu(x)}{\lambda(x)} \equiv \delta$$

and $\delta < 1$, since 1 is not in the range $\frac{\mu(x)}{\lambda(x)}$.

THEOREM 3. If in addition to i) - iv) of the previous theorem, it is true that

v) $h \equiv \frac{B^2 K \eta}{1-\delta} < 2$ and $N(x^o, r) \subset \Omega_o$, where $r = \frac{B\eta}{1-\alpha}$, $\alpha = \delta + (1 - \delta)\frac{h}{2}$,

then the iteration (1) is defined for every n and converges to a point $\sigma \in \Omega_0$ with $F(\sigma) = 0$, at a rate given by $||x^n - \sigma|| \leq r\alpha^n$.

Proof. Follows directly from Theorems 1 and 2.

In the problem of minimizing a positive function $P(x)$, one often examines the zeros of the gradient function, ∇P , as candidates for yielding the minimum. If P is twice continuously differentiable, the Jacobian matrix $J_P(x)$, of ∇P at x (i.e., the Hessian of P) will be Hermitean. If $G(x)$ is chosen to be Hermitean, the foregoing theorems simplify with $\lambda(x)$ being the minimum modulus of the eigenvalues of $G(x)$ and $\mu(x)$ being the maximum modulus of the eigenvalues of $J_P(x) - G(x)$.

If $G(x)$ is a diagonal matrix with diag $G(x)$ = diag $J(x)$ then (1) becomes the Newton-Jacobi iteration

$$(2)\quad x_i^{k+1} = x_i^k - f_i(x^k) \,/\, f_{ii}(x^k) \ , \ i = 1,\dots,N \ ; \ k = 0,1,\dots ,$$

whose convergence has been studied by several authors [5], [8], and [13]. The method was originally proposed by Lieberstein [12].

Remark 1. If $J(x)$ is Hermitean and $\min_{1\leq i\leq N}|f_{ii}(x)| > \max_{1\leq i\leq N}\sum_{\substack{j=1\\ j\neq i}}^{N}|f_{ij}(x)|$,

then iii) and iv) of Theorem 2 hold for the Newton-Jacobi iteration (2).

Proof. $G(x)$ is the diagonal of $J(x)$ which, by hypothesis, has no zero diagonal elements. The eigenvalues of $G(x)$ are its nonzero elements and so $\lambda(x)$ is greater than zero and is the minimum over i of $|f_{ii}(x)|$. Now by the Gerschgorin Theorem,

$$\mu(x) \leq \max_i \sum_{\substack{j=1 \\ j\neq i}}^{N} |f_{ij}(x)| . \quad \text{Thus,}$$

$$\frac{\mu(x)}{\lambda(x)} \leq \frac{\max_i \sum_{\substack{j=1 \\ j\neq i}}^{N} |f_{ij}(x)|}{\min_i |f_{ii}(x)|} < 1 .$$

The following very nice result, which we particularize to the present setting, is due to Rheinboldt [15, p. 16].

THEOREM 4. IF: i) $||J(x) - J(y)|| \leq \gamma||x - y||$, for $x,y \in \Omega_o$.

ii) $||G(x) - G(x^o)|| \leq \eta||x - x^o||$, $x \in \Omega_o$,

iii) $||J(x) - G(x)|| \leq \delta$, $x \in \Omega_o$,

iv) $||G^{-1}(x^o)|| \leq \beta$,

v) $||G^{-1}(x^o)F(x^o)|| \leq \alpha$ and

vi) $\beta\delta < 1$, $h = \frac{\sigma\beta\gamma\alpha}{(1-\beta\delta)^2} \leq 1/2; \sigma = \max(1,\frac{\eta}{\gamma})$ and

$$\text{cl } N(x^o,t^*) \subset \Omega_o , \quad t^* = \frac{1 - \sqrt{1-2h}}{h} \cdot \frac{\alpha}{1-\beta\delta} ;$$

THEN: the sequence $\{x^n\}$ given by (1) is defined, remains in cl $N(x^o,t^*)$ and converges to $\sigma \ni F(\sigma) = 0$ and σ is unique in

$$N(x^o,t^{**}) , \quad t^{**} = \frac{1 + \sqrt{1-2h}}{h} \cdot \frac{\alpha}{1-\beta\delta} .$$

4. An Algorithm for the Solution of Nonlinear Systems of Equations.

The foregoing theorems suggest that if in (1) $G(x^n)$ is close to the Jacobian matrix at each step of the iteration, then a numerical method based on (1) will converge if the Jacobian matrix is not too ill-conditioned [6] near the root. Now as the root is approached, the successive iterates and hence the elements of the Jacobian matrix (assumed well conditioned) show less change from step to step. An algorithm is suggested, therefore, in which those elements of the Jacobian matrix which show little inclination to change are held constant during further iterations. We detail the algorithm as follows:

1. Use any method or combination of methods until a point x^n is obtained for which a normalized value of $||F(x^n)||$ is small; e.g.,

$$\frac{||F(x^n)||}{||F(x^o)||} < 1 \qquad \text{if} \quad 0 < ||F(x^o)|| \leq 1$$

or

$$||F(x^n)|| < 1 \qquad \text{if} \qquad ||F(x^o)|| > 1.$$

2. Compute and store $J(x^n)$ and form x^{n+1} using the standard Newton iteration

$$x^{n+1} = x^n - [J(x^n)]^{-1}F(x^n) \quad .$$

3. Form the elements, $f_{ij}(x^{n+1})$, of $J(x^{n+1})$ one at a time comparing them with the corresponding elements of $J(x^n)$. If $f_{ij}(x^{n+1})$ and $f_{ij}(x^n)$ agree in a relative sense to within a specified tolerance T, record the subscript pair (i,j) in a list $\mathcal{L}$. If the tolerance is not met make no such record. In any event after making the comparison, replace $f_{ij}(x^n)$ in storage by $f_{ij}(x^{n+1})$.

4. Form x^{n+2}, again using the standard Newton iteration.

5. Now form the elements, $f_{ij}(x^{n+2})$, of $J(x^{n+2})$ one at a time.

If $(i,j) \notin \mathcal{L}$, make no comparison test. If $(i,j) \in \mathcal{L}$, compare $f_{ij}(x^{n+2})$ with $f_{ij}(x^{n+1})$ using the same test given above; i.e., is

$$|f_{ij}(x^{n+2}) - f_{ij}(x^{n+1})| \leq T\,|f_{ij}(x^{n+2})| \quad ?$$

(Note, if $f_{ij}(x^{n+2}) = 0$, an absolute error test can be used.)

If the test is not met, delete the subscript pair (i,j) from the list $\mathcal{L}$. If the test <u>is</u> met, allow the subscript pair (i,j) to remain in the list $\mathcal{L}$. In either case, store $f_{ij}(x^{n+2})$ in place of $f_{ij}(x^{n+1})$.

6. Calculate x^{n+3} from the standard Newton iteration.
7. Set $k = n + 3$.
8. Form x^{k+1} from the relation

$$x^{k+1} = x^k - [G(x^k)]^{-1} F(x^k), \text{ where}$$

$G(x^k) \equiv [g_{ij}(x^k)]$ is formed as either

$$g_{ij}(x^k) = f_{ij}(x^{n+2}) \text{ (stored)}, \quad (i,j) \in \mathcal{L}, \text{ or}$$

$$g_{ij}(x^k) = f_{ij}(x^k), \quad (i,j) \notin \mathcal{L}.$$

9. Replace k by $k + 1$.
10. Repeat steps 8 and 9 until convergence occurs.

Remark 1. Obviously any implementation of this algorithm will include tests for convergence after steps 2, 4 and 6.

Remark 2. The foregoing algorithm represents a kind of middle ground between a) the stationary [11] (or periodically-stationary [4], [9]) Newton's method and b) the standard Newton's method [11].

Remark 3. In many problems it is inconvenient to give explicit analytic expressions for the partial derivatives. The algorithm presented in this section can also be implemented using differences to approximate the exact partial derivatives; e.g.,

$$f_{ij}(x^n) \approx \frac{f_i(x^n + h\cdot e_j) - f_i(x^n)}{h} ,$$

where e_j is the jth unit vector and h is a small number which may depend on i, j and n .

5. Numerical Results.

The method of section 4 was tested on a number of systems of small order with uniformly good results: convergence to the same accuracy which Newton's method yielded with fewer or at most the same number of function evaluations required. A typical example was the 2×2 system considered first by Freudenstein and Roth [10] and later by Broyden [4] and Brown and Conte [3] :

$$f_1(x_1,x_2) = -13 + x_1 + [(-x_2 + 5)x_2 - 2]x_2$$

$$f_2(x_1,x_2) = -29 + x_1 + [(x_2 + 1)x_2 - 14]x_2 .$$

Using initial values of

$$x_1^0 = 4.5 , \quad x_2^0 = 4.3 ,$$

Newton's method required 24 total evaluations (8 function evaluations and 16 partial derivative evaluations) to produce the solution (5,4) to 14 digits of accuracy. The algorithm presented herein produced 14 digits of accuracy after 20 total evaluations (8 function evaluations and 12 partial derivative evaluations).

Since we conjectured that the practical advantage of the algorithm would show when solving a large, computationally complex system, it seemed only fair to test the method on a demanding example. We chose the 24×24 system of highly nonlinear equations considered by Pack and Swan [14] in explaining a certain magneto-gasdynamic flow. These equations involve considerable computations as the following "average" (in terms of amount of computation required) equation of the system shows:

$$f_5(x) = .5(1 - x_5^2) + .016(1 - x_4^2/x_6) + .4(1 - x_7/x_6)$$

$$- (.016/\sin x_1)\ [.97814 - x_5x_4\cos(x_2 - x_3)]\sin(x_1 - .20944) .$$

Robinson [16] has solved this system numerically with his method, a generalization of the secant method to nonlinear systems.

Using a copy of the program with which Robinson solved this 24×24 system, we were able to compare the results obtained by implementing the algorithm of section 4 with his results. In the implementation, first differences were used to approximate the partial derivatives. Robinson's starting guess was used for each of the methods tested and for that starting guess, x^0 , $||F(x^0)||_\infty = \max_{1 \le i \le 24} |f_i(x^0)| = .303$.

We summarize the results in Table 1.

Table 1

Method	Total Number of Evaluations	Final Value of $\|\|F\|\|_\infty$	Computer Time Used (Relative)
Robinson	4104	9.65×10^{-9}	1.87
Newton with analytic derivatives (reported in [16])	not reported	$\sim 10^{-9}$	~3.6
Algorithm of section 4 (discretized form)	1933	6.12×10^{-10}	1

The tolerance value, T , used was 0.1 (see steps 3 and 5 of the algorithm in section 4).

6. Acknowledgement.

The authors wish to thank Dr. Louis B. Rall, Assistant Director of the Mathematics Research Center, U. S. Army, University of Wisconsin, for kindly furnishing us with copies of the Robinson program and the Pack and Swan equations.

REFERENCES

1. Barnes, J.G.P., "An Algorithm for Solving Nonlinear Equations Based on the Secant Method," Comput. J., pp. 66-72 (1965).

2. Brown, K.M., "Solution of Simultaneous Non-Linear Equations," Comm. of the ACM, Vol. 10, No. 11, pp. 728-729 (1967).

3. Brown, K.M. and S.D. Conte, "The Solution of Simultaneous Nonlinear Equations," 22nd National Conference - ACM, Washington, D.C., Thompson Book Company, pp. 111-114 (1967).

4. Broyden, C.G., "A Class of Methods for Solving Nonlinear Simultaneous Equations," Math. Comp., 19, pp. 577-593 (1965).

5. Bryan, C.A., "An Iterative Method for Solving Nonlinear Systems of Equations," Ph. D. Dissertation, Univ. of Arizona, 1963.

6. Collatz, L., Functional Analysis and Numerical Mathematics, New York, Academic Press, 1966.

7. Dennis, J.E., Jr., "On Newton-Like Methods," (to appear).

8. Dennis, J.E., Jr., "On Newton's Method and Nonlinear Simultaneous Displacements," SIAM J. Numer. Anal., Vol. 4, No. 1, pp. 103-108 (1967).

9. Dennis, J.E., Jr., "On the Kantorovich Hypothesis for Newton's Method," (to appear).

10. F. Freudenstein and B. Roth, "Numerical Solutions of Systems of Nonlinear Equations," J. Assoc. Comput. Mach., Vol. 10, pp. 550-556 (1963).

11. Kantorovich, L.V. and G.P. Akilov, Functional Analysis in Normed Spaces, New York, Pergammon Press, 1964.

12. Lieberstein, H.M., "Overrelaxation for Nonlinear Elliptic Partial Differential Equations," M.R.C. Tech. Summary Report 80, Univ. of Wisconsin, 1959.

13. Ortega, J. and W. Rheinboldt, "Monotone Iterations for Nonlinear Equations with Applications to Gauss-Seidel Methods," SIAM J. on Numer. Anal., Vol. 4, No. 2, pp. 171-190 (1967).

14. Pack, D.C. and G.W. Swan, "Magneto-gasdynamic Flow Over a Wedge," J. Fluid Mech., 25, pp. 165-178 (1966).

15. Rheinboldt, Werner C., "On a Unified Convergence Theory for a Class of Iterative Processes," University of Maryland Technical Report TR-67-46, May 1967.

16. Robinson, Stephen M., "Interpolative Solution of Systems of Nonlinear Equations," SIAM J. Numer. Anal., Vol. 3, No. 4, pp. 650-658 (1966).

17. Schechter, S., "Iteration Methods for Nonlinear Problems," Trans. Amer. Math. Soc., 104, pp. 179-189 (1962).