MATHEMATICS OF COMPUTATION
VOLUME 41, NUMBER 164
OCTOBER 1983, PAGES 527-536

Local Convergence of Difference Newton-Like Methods

By T. J. Ypma

Abstract. Using affine invariant terms, we give a local convergence analysis of difference Newton-like methods for solving the nonlinear equation $F(x) = 0$. The convergence conditions are weaker than those standardly required for methods of this class. The technique and results are valid for all currently known difference Newton-like methods which require evaluation of all components of F at the same point. Radius of convergence and rate of convergence results for particular difference Newton-like methods may easily be derived from the results reported here.

1. Introduction. Newton-like methods for solving the system of nonlinear equations

$$F(x) = 0; \qquad F: D \subseteq R^N \to R^N \tag{1.1}$$

have the form

$$x_{i+1} = x_i - A_i^{-1}F(x_i), \qquad i = 0, 1, \ldots, \tag{1.2}$$

where $x_0 \in D$ is some given starting point and $A_i \in \mathcal{L}(R^N)$ $(i = 0, 1, \ldots)$ is some approximation to the Fréchet derivative $F'(x_i)$ of F at $x_i \in D$. In this paper we examine methods of the form (1.2) in which the $A_i \in \mathcal{L}(R^N)$ are difference approximations to $F'(x_i)$ in the following sense:

Let $P, Q \in \mathcal{L}(R^N)$ be two matrices with columns $p_j, q_j \in R^N$ $(j = 1, \ldots, N)$, respectively. Let $x \in D$, and suppose p_j, q_j are such that $x + p_j$, $x + q_j \in D$ $(j = 1, \ldots, N)$. Write $R \equiv P - Q$, and suppose $R \in \mathcal{L}(R^N)$ is nonsingular. If $Y(x) \in \mathcal{L}(R^N)$ is the matrix whose columns $y_j(x)$ satisfy

$$y_j(x) \equiv F(x + p_j) - F(x + q_j), \qquad j = 1, \ldots, N, \tag{1.3}$$

then the *difference approximation* $B(x) \in \mathcal{L}(R^N)$ defined by P, $Q \in \mathcal{L}(R^N)$ to the Fréchet derivative $F'(x)$ of F at x is

$$B(x) \equiv Y(x)R^{-1}. \tag{1.4}$$

The use of difference approximations in (1.2) thus gives iterative methods of the form

$$x_{i+1} = x_i - B(x_i)^{-1}F(x_i), \qquad i = 0, 1, \ldots, \tag{1.5}$$

which we refer to as *difference Newton-like methods*. Whenever we write (1.5) we imply that for each $i = 0, 1, \ldots$ there exist $P_i, Q_i \in \mathcal{L}(R^N)$ such that $B(x_i)$ is the difference approximation to $F'(x_i)$ defined by P_i and Q_i.

Received August 20, 1982.

1980 *Mathematics Subject Classification.* Primary 65H10.

Methods of this form have been extensively studied. In both [5] and [6] many such methods are described; both references provide many theoretical results, a historical survey and lengthy bibliographies. The general method (1.5) considerably extends the methods considered in [5]; in fact it includes all currently known Newton-like methods in which the Jacobian matrix $F'(x)$ is approximated by differences between values of F, provided that all the components of F are evaluated at the same point.

It is well known [1], [2], [3], [9] that many methods of the form (1.2) are *affine invariant* in the sense that when they are applied to the transformed mapping

$$\bar{F} \equiv CF, \tag{1.6}$$

where $C \in \mathfrak{L}(R^N)$ is any nonsingular matrix, then the corresponding approximation $\bar{A}_i$ to the derivative $\bar{F}'(x_i)$ of $\bar{F}$ at x_i is such that

$$\bar{A}_i^{-1}\bar{F}(x_i) \equiv A_i^{-1}F(x_i), \qquad i = 0, 1, \ldots, \tag{1.7}$$

with the consequence that

$$x_{i+1} = x_i - A_i^{-1}F(x_i) \equiv x_i - \bar{A}_i^{-1}\bar{F}(x_i), \qquad i = 0, 1, \ldots, \tag{1.8}$$

that is, the method (1.2) produces the same sequence of iterates when used to solve $\bar{F}(x) = 0$ as it does when used to solve $F(x) = 0$. Methods with this property are important because they are likely to be insensitive with respect to scaling of the mapping F. In [3] it was pointed out that affine invariant methods should be studied in affine invariant terms, that is, terms which are unaffected by the transformation (1.6). In view of these remarks the analysis in this paper is conducted largely in affine invariant terms. By (1.3)–(1.7) it is clear that difference Newton-like methods are affine invariant if and only if both the matrices P_i, $Q_i \in \mathfrak{L}(R^N)$ which define $B(x_i)$ are invariant with respect to affine transformations of F, for $i = 0, 1, \ldots$.

In this paper we provide conditions under which difference Newton-like methods converge to a solution $x^* \in R^N$ of (1.1). We show by example that radius of convergence and rate of convergence results for particular difference Newton-like methods may easily be derived from our fundamental results.

2. Basic Theory. Throughout this paper $\|x\|$ for $x \in R^N$ will denote the Euclidean norm of x and $\|C\|$ for $C \in \mathfrak{L}(R^N)$ will denote the subordinate spectral norm of C. If $C \in \mathfrak{L}(R^N)$ has columns $c_j \in R^N$ ($j = 1, 2, \ldots N$), then the Frobenius norm of C is defined as $\|C\|_F = (\sum_{j=1}^N \|c_j\|^2)^{1/2}$, and [2]

$$\|c_j\| \leq \|C\| \leq \|C\|_F, \qquad j = 1, \ldots, N. \tag{2.1}$$

For $x \in R^N$ and $\delta > 0$

$$U(x, \delta) \equiv \{y \in R^N\colon \|x - y\| < \delta\} \tag{2.2}$$

denotes the open δ-neighborhood of x in R^N. We write e_j ($j = 1, \ldots, N$) for the columns of the $N \times N$ identity matrix.

We restrict our attention to the following class of mappings.

$$\begin{aligned} \theta \equiv \{F\colon D \subseteq R^N \to R^N &\text{ where } F \text{ is continuously Fréchet differentiable} \\ &\text{on the nonempty, open, convex set } D \text{ (which may depend on } F)\}. \end{aligned} \tag{2.3}$$

Given $x \in V \subseteq R^N$, where V is open, we define

$$\Psi(x, V) \equiv \{F \in \theta\colon V \subseteq D;\ F'(x)^{-1} \text{ exists; there exists } K > 0 \text{ such that for all } y, z \in V, \tag{2.4}$$

$$\|F'(x)^{-1}(F'(y) - F'(z))\| \leq K\|y - z\|\}, \tag{2.5}$$

and for any $\sigma > 0$ we define

$$\mathfrak{F}(\sigma) \equiv \{F \in \Psi(x^*, U(x^*, \sigma)) \text{ where } x^* \in D \text{ is such that } F(x^*) = 0\}. \tag{2.6}$$

For any $F \in \Psi(x, V)$ we define

$$\mu(F, x) \equiv \sup\left\{\frac{\|F'(x)^{-1}(F'(y) - F'(z))\|}{\|y - z\|} : y, z \in V;\ y \neq z\right\}. \tag{2.7}$$

Notice in particular that if $F \in \Psi(x, V)$, then for all $a, b \in V$

$$\|F'(x)^{-1}(F'(a) - F'(b))\| \leq \mu(F, x)\|a - b\|.$$

Using the perturbation lemma [5], it is easy to show [7]–[9] that if $F \in \Psi(x, V)$, then, for any $y \in V$ with $\|x - y\| < \mu(F, x)^{-1}$, $F'(y)$ is nonsingular and for all a, $b \in V$

$$\|F'(y)^{-1}(F'(a) - F'(b))\| \leq \frac{\mu(F, x)}{1 - \mu(F, x)\|x - y\|} \cdot \|a - b\|,$$

from which it follows that

$$F \in \Psi(y, V); \qquad \mu(F, y) \leq \frac{\mu(F, x)}{1 - \mu(F, x)\|x - y\|}. \tag{2.8}$$

We now provide bounds on the error in approximating the Jacobian matrix $F'(x)$ by a difference approximation $B(x)$. The results extend those described in [2] and [9]. We require the following definition:

If $P, Q \in \mathfrak{L}(R^N)$ are such that $R \equiv P - Q$ is nonsingular, with columns r_j $(j = 1, \ldots, N)$, then

$$h(P, Q) \equiv (\|P\|_F + \|Q\|_F)\|\operatorname{diag}(\|r_1\|, \ldots, \|r_N\|)R^{-1}\|. \tag{2.9}$$

(2.10) Theorem. *If $F \in \theta$, $x \in D$ and there exists $L \geq 0$ such that for all $y \in V$*

$$\|F'(x) - F'(y)\| \leq L\|x - y\|,$$

and if $B(x)$ is the difference approximation to $F'(x)$ defined by P, $Q \in \mathfrak{L}(R^N)$, then

$$\|B(x) - F'(x)\| \leq \tfrac{1}{2}Lh(P, Q).$$

Proof. By definition

$$\begin{aligned} B(x) - F'(x) &\equiv Y(x)R^{-1} - F'(x) = (Y(x) - F'(x)R)R^{-1} \\ &= (Y(x) - F'(x)R)C^{-1}CR^{-1}, \end{aligned}$$

where $C \equiv \operatorname{diag}(\|r_1\|, \ldots, \|r_N\|)$ is nonsingular. By a well-known mean value theorem [6, 2.2.10]

$$\begin{aligned} \|(Y(x) - F'(x)R)e_j\| &= \|F(x + Pe_j) - F(x + Qe_j) - F'(x)(Pe_j - Qe_j)\| \\ &\leq \tfrac{1}{2}L\|Pe_j - Qe_j\|(\|Pe_j\| + \|Qe_j\|) = \tfrac{1}{2}L\|r_j\|(\|p_j\| + \|q_j\|). \end{aligned}$$

It follows that

$$\begin{aligned}\|(Y(x)-F'(x)R)C^{-1}\|^2 &\leq \|(Y(x)-F'(x)R)C^{-1}\|_F^2\\ &= \sum_{j=1}^{N}\|(Y(x)-F'(x)R)C^{-1}e_j\|^2 = \sum_{j=1}^{N}\left(\frac{\|(Y(x)-F'(x)R)e_j\|}{\|r_j\|}\right)^2\\ &\leq \left(\frac{1}{2}L\right)^2\sum_{j=1}^{N}\left(\|p_j\|+\|q_j\|\right)^2.\end{aligned}$$

Combining these expressions gives

$$\begin{aligned}\|B(x)-F'(x)\| &\leq \frac{1}{2}L\left(\sum_{j=1}^{N}\left(\|p_j\|+\|q_j\|\right)^2\right)^{1/2}\|CR^{-1}\|\\ &= \frac{1}{2}L\left(\sum_{j=1}^{N}\left(\|p_j\|+\|q_j\|\right)^2\right)^{1/2}\left\|\operatorname{diag}(\|r_1\|,\ldots,\|r_N\|)R^{-1}\right\|.\end{aligned}$$

But by the Minkowski inequality

$$\left(\sum_{j=1}^{N}\left(\|p_j\|+\|q_j\|\right)^2\right)^{1/2} \leq \|P\|_F+\|Q\|_F$$

from which the result follows. □

An analogous result appears in [6, 5.3.4]. In the frequently used case that $Q \equiv 0$ Theorem (2.10) reduces to the result of [2, Theorem 3.4]. We now provide an affine invariant form of Theorem (2.10).

(2.11) THEOREM. *Let $F \in \Psi(x, V)$. If $B(x)$ is the difference approximation to $F'(x)$ defined by $P, Q \in \mathfrak{L}(R^N)$ and if*

$$h(P,Q) < 2\mu(F,x)^{-1}, \tag{2.12}$$

then $B(x)$ is nonsingular and

$$\left\|B(x)^{-1}F'(x)-I\right\| \leq \frac{\mu(F,x)h(P,Q)}{2-\mu(F,x)h(P,Q)}, \tag{2.13}$$

$$\left\|B(x)^{-1}F'(x)\right\| \leq \frac{2}{2-\mu(F,x)h(P,Q)}. \tag{2.14}$$

Proof. The proof of (2.13) is entirely analogous to that of [2, Theorem 3.5]. Then (2.14) follows from $\|B(x)^{-1}F'(x)\| \leq \|B(x)^{-1}F'(x)-I\|+1$. □

The following lemma is a useful aid to satisfying condition (2.12).

(2.15) LEMMA. *Let $F \in \Psi(x, V)$. If $y \in V$ satisfies*

$$\|x-y\|+\tfrac{1}{2}\varepsilon \leq \mu(F,x)^{-1}$$

for some $\varepsilon \in (0, 2\mu(F,x)^{-1})$, then $F \in \Psi(y, V)$ and for all $P, Q \in \mathfrak{L}(R^N)$ with $h(P,Q) \leq \varepsilon$

$$h(P,Q) \leq 2\mu(F,y)^{-1}.$$

Proof. As noted in (2.8) $F \in \Psi(y, V)$ and

$$\mu(F, y) \leqslant \frac{\mu(F, x)}{1 - \mu(F, x)\|x - y\|},$$

from which it follows that

$$h(P, Q) \leqslant \varepsilon \leqslant 2\left(\mu(F, x)^{-1} - \|x - y\|\right) \leqslant 2\mu(F, y)^{-1}. \quad \square$$

The following mean value results will also prove to be useful.

(2.16) LEMMA. *Let $F \in \Psi(x, V)$ and suppose $B(x)$ is the difference approximation to $F'(x)$ defined by $P, Q \in \mathfrak{L}(R^N)$ with $h(P, Q) < 2\mu(F, x)^{-1}$. Then for all $y, z \in V$*

$$\text{(2.17)} \qquad \left\|B(x)^{-1}(F'(y) - F'(z))\right\| \leqslant \frac{2\mu(F, x)}{2 - \mu(F, x)h(P, Q)}\|y - z\|,$$

$$\text{(2.18)} \qquad \left\|B(x)^{-1}(F(y) - F(z) - F'(z)(y - z))\right\| \leqslant \frac{\mu(F, x)\|y - z\|^2}{2 - \mu(F, x)h(P, Q)}.$$

Proof.

$$\left\|B(x)^{-1}(F'(y) - F'(z))\right\| \leqslant \left\|B(x)^{-1}F'(x)\right\| \left\|F'(x)^{-1}(F'(y) - F'(z))\right\|,$$

which, with (2.14), gives (2.17). Defining $\bar{F}(y) \equiv B(x)^{-1}F(y)$ for all $y \in V$, we obtain from (2.17)

$$\left\|\bar{F}'(y) - \bar{F}'(z)\right\| \leqslant \frac{2\mu(F, x)}{2 - \mu(F, x)h(P, Q)}\|y - z\|$$

for all y, $z \in V$. Together with a well-known mean value theorem [5, 3.2.12] this gives (2.18). $\square$

3. Local Convergence. The result underlying our local convergence analysis is the following lemma.

(3.1) LEMMA. *Let $F \in \mathfrak{F}(\sigma)$ for some $\sigma > 0$. For any $x \in U(x^*, \sigma)$ let $B(x)$ be a difference approximation to $F'(x)$ defined by some $P, Q \in \mathfrak{L}(R^N)$ with $h(P, Q) \leqslant \varepsilon$ for some $\varepsilon > 0$. If $\|x - x^*\| + \frac{1}{2}\varepsilon < \mu(F, x^*)^{-1}$, then $B(x)$ is nonsingular and*

$$\text{(3.2)} \qquad \left\|x - B(x)^{-1}F(x) - x^*\right\| \leqslant \frac{\mu(F, x)\|x - x^*\|}{2 - \mu(F, x)h(P, Q)}\left(\|x - x^*\| + h(P, Q)\right).$$

Proof. Clearly $\varepsilon < 2\mu(F, x^*)^{-1}$. Since $F \in \Psi(x^*, U(x^*, \sigma))$, by Lemma (2.15) $F \in \Psi(x, U(x^*, \sigma))$ and $h(P, Q) < 2\mu(F, x)^{-1}$. By Theorem (2.11) it follows that $B(x)$ is nonsingular, and using also Lemma (2.16)

$$\begin{aligned}
\left\|x - B(x)^{-1}F(x) - x^*\right\| &= \left\|B(x)^{-1}(F(x^*) - F(x) - B(x)(x^* - x))\right\| \\
&\leqslant \left\|B(x)^{-1}(F(x^*) - F(x) - F'(x)(x^* - x))\right\| \\
&\quad + \left\|B(x)^{-1}(F'(x) - B(x))(x^* - x)\right\| \\
&\leqslant \frac{\mu(F, x)\|x - x^*\|^2 + \mu(F, x)h(P, Q)\|x - x^*\|}{2 - \mu(F, x)h(P, Q)},
\end{aligned}$$

which gives the result. $\square$

This leads directly to the convergence theorem.

(3.3) THEOREM. *Let $F \in \mathfrak{F}(\sigma)$ for some $\sigma > 0$. For all $x_i \in U(x^*, \sigma)$ let $B(x_i)$ be a difference approximation to $F'(x_i)$ defined by $P_i, Q_i \in \mathfrak{L}(R^N)$ such that*

$$h(P_i, Q_i) \leqslant \varepsilon \tag{3.4}$$

for some $\varepsilon > 0$. Then the iterative method

$$x_{i+1} = x_i - B(x_i)^{-1} F(x_i), \qquad i = 0, 1, \ldots, \tag{3.5}$$

converges to x^* if the starting point $x_0 \in U(x^*, \sigma)$ satisfies

$$\|x_0 - x^*\| + \tfrac{2}{3}\varepsilon < \tfrac{2}{3}\mu(F, x^*)^{-1}. \tag{3.6}$$

The sequence $\{x_i\}$ then satisfies

$$\|x_{i+1} - x^*\| \leqslant \frac{\mu(F, x^*)\|x_i - x^*\|(\|x_i - x^*\| + h(P_i, Q_i))}{2(1 - \mu(F, x^*)\|x_i - x^*\|) - \mu(F, x^*)h(P_i, Q_i)}. \tag{3.7}$$

Proof. Notice $\frac{2}{3}(\mu(F, x^*)^{-1} - \varepsilon) < \mu(F, x^*)^{-1} - \frac{1}{2}\varepsilon$. Assume $\|x_i - x^*\| \leqslant \|x_0 - x^*\|$; then $x_i \in U(x^*, \sigma)$ and $\|x_i - x^*\| + \frac{2}{3}\varepsilon \leqslant \frac{2}{3}\mu(F, x^*)^{-1}$, so $\|x_i - x^*\| + \frac{1}{2}\varepsilon < \mu(F, x^*)^{-1}$. By Lemma (3.1) it follows that $B(x_i)$ is nonsingular, so the iteration (3.5) is feasible and (3.7) follows from (3.2) using

$$\mu(F, x_i) \leqslant \frac{\mu(F, x^*)}{1 - \mu(F, x^*)\|x_i - x^*\|}. \tag{3.8}$$

It follows by induction that $\{\|x_i - x^*\|\}$ decreases monotonically to zero if (3.6) holds, since in that case

$$\frac{\mu(F, x^*)(\|x_0 - x^*\| + \varepsilon)}{2(1 - \mu(F, x^*)\|x_0 - x^*\|) - \mu(F, x^*)\varepsilon} < 1. \quad \square$$

The above result is sharp.

(3.9) THEOREM. *For any $\sigma > 0$ and $\delta > 0$ there exist $F \in \mathfrak{F}(\sigma)$ and $\varepsilon > 0$, with $\sigma + \frac{2}{3}\varepsilon > \frac{2}{3}\mu(F, x^*)^{-1}$, and P: $U(x^*, \sigma) \to \mathfrak{L}(R^N)$ and Q: $U(x^*, \sigma) \to \mathfrak{L}(R^N)$ with $P(x) - Q(x)$ nonsingular and $h(P(x), Q(x)) < \varepsilon$ for all $x \in U(x^*, \sigma)$, and there exists $x_0 \in U(x^*, \sigma)$ with*

$$\|x_0 - x^*\| + \tfrac{2}{3}\varepsilon \leqslant \tfrac{2}{3}\mu(F, x^*)^{-1} + \delta,$$

such that if iteration (3.5) is used with $B(x_i)$ defined using $P(x_i), Q(x_i) \in \mathfrak{L}(R^N)$, then the method does not converge to x^*.

Proof. For $N = 1$ define F to be the pseudo-cubic

$$F(x) = \begin{cases} x + \frac{1}{2}Lx^2 & \text{if } x \leqslant 0, \\ x - \frac{1}{2}Lx^2 & \text{if } x > 0, \end{cases}$$

where $0 < L < \sigma^{-1} < \frac{3}{2}L$. With $x^* = 0$ it is easy to show that $F \in \mathfrak{F}(2L^{-1})$, so $F \in \mathfrak{F}(\sigma)$, and $\mu(F, x^*) = L$. Clearly $\sigma > \frac{2}{3}L^{-1}$. Let $\varepsilon > 0$ and $\beta > 0$ be such that $\sigma + \frac{2}{3}\varepsilon > \frac{2}{3}L^{-1}$ and $\varepsilon - \frac{2}{3}\delta < \beta < \varepsilon$, and for all $x \in U(x^*, \sigma)$, $x \neq 0$, let $Q(x) \equiv 0$ and $P(x) \equiv \beta x / |x|$. Clearly $h(P(x), Q(x)) = \beta < \varepsilon$ for all $x \neq 0$ and

$$B(x) = \begin{cases} \frac{1}{2}(2 + 2Lx - L\beta) & \text{if } x < 0, \\ \frac{1}{2}(2 - 2Lx - L\beta) & \text{if } x > 0. \end{cases}$$

It is easy to verify that $x - B(x)^{-1}F(x) = -x$ if and only if $x = \pm \frac{2}{3}(L^{-1} - \beta)$. Taking $x_0 = \frac{2}{3}(L^{-1} - \beta)$ it follows that iteration (3.5) fails to converge to x^* although

$$|x_0 - x^*| + \tfrac{2}{3}\varepsilon = \tfrac{2}{3}(L^{-1} - \beta + \varepsilon) < \tfrac{2}{3}L^{-1} + \delta. \quad \square$$

It is sometimes easier to estimate $\mu(F, x_0)$ than $\mu(F, x^*)$ [7]–[9]. Hence the following corollary is useful.

(3.10) COROLLARY. *Let $F \in \mathfrak{F}(\sigma)$ for some $\sigma > 0$. For all $x_i \in U(x^*, \sigma)$ let $B(x_i)$ be a difference approximation to $F'(x_i)$ defined by $P_i, Q_i \in \mathfrak{L}(R^N)$ such that*

$$h(P_i, Q_i) \leqslant \varepsilon \tag{3.11}$$

for some $\varepsilon > 0$. If $x_0 \in U(x^*, \sigma)$ satisfies

$$\|x_0 - x^*\| + \tfrac{4}{3}\varepsilon < \tfrac{2}{3}\mu(F, x_0)^{-1} \tag{3.12}$$

and if $F \in \Psi(x_0, U(x^*, \sigma))$, then the iteration (3.5) *is feasible and converges to x^* with convergence characterized by* (3.7).

Proof. As in (2.8)

$$\mu(F, x^*) \leqslant \frac{\mu(F, x_0)}{1 - \mu(F, x_0)\|x_0 - x^*\|};$$

with $3\mu(F, x_0)\|x_0 - x^*\| < 2 - 4\mu(F, x_0)\varepsilon$ this gives

$$\|x_0 - x^*\| \leqslant \frac{2(1 - \mu(F, x_0)\|x_0 - x^*\| - 2\mu(F, x_0)\varepsilon)}{\mu(F, x_0)} \leqslant 2\big(\mu(F, x^*)^{-1} - 2\varepsilon\big).$$

Also $\|x_0 - x^*\| < \frac{2}{3}(\mu(F, x_0)^{-1} - 2\varepsilon)$ implies

$$\frac{\mu(F, x_0)(\|x_0 - x^*\| + \varepsilon)}{2 - \mu(F, x_0)\varepsilon} < \frac{1}{3}.$$

Now x_1 generated by (3.5) is well defined since $h(P_0, Q_0) \leqslant \varepsilon < 2\mu(F, x_0)^{-1}$, so that by Theorem (2.11) $B(x_0)$ is nonsingular. It follows as in Lemma (3.1) that

$$\begin{aligned}\|x_1 - x^*\| &\leqslant \frac{\mu(F, x_0)(\|x_0 - x^*\| + \varepsilon)}{2 - \mu(F, x_0)\varepsilon}\|x_0 - x^*\| \\ &< \tfrac{1}{3}\|x_0 - x^*\| < \tfrac{2}{3}\big(\mu(F, x^*)^{-1} - 2\varepsilon\big) < \tfrac{2}{3}\big(\mu(F, x^*)^{-1} - \varepsilon\big).\end{aligned}$$

The result follows from Theorem (3.3). $\square$

The results of Theorem (3.3), Theorem (3.9) and Corollary (3.10) are analogous to those proved for Newton's method in [7] and for difference Newton-like methods with $Q \equiv 0$ in [9].

The appearance of the term ε in the radius of convergence results (3.6) and (3.12), together with the restriction $h(P, Q) \leqslant \varepsilon$, may be regarded as limiting the choice of P and Q. Conversely, if $h(P, Q)$ can be related to $\|x_0 - x^*\|$ directly, then it is possible to calculate a limit on $\|x_0 - x^*\|$ from (3.6). For example if P and Q are always selected in such a way that for some $M > 0$

$$h(P, Q) \leqslant M\|x_0 - x^*\|, \tag{3.13}$$

then condition (3.6) is satisfied if

$$\|x_0 - x^*\| < \frac{2}{(3+2M)\mu(F, x^*)}. \tag{3.14}$$

Thus for many difference Newton-like methods currently in use explicit radius of convergence results are obtainable from (3.6) and (3.12).

Notice that we have proved convergence of difference Newton-like methods requiring from P_i and Q_i only that $P_i - Q_i$ be nonsingular and $h(P_i, Q_i)$ uniformly bounded, for all P_i and Q_i used. This is weaker than the standard requirement [5], [6] that for all P_i and Q_i used, $\|P_i\|_F$ and $\|Q_i\|_F$ are sufficiently small and there exists $\Omega > 0$ such that if $R_i \equiv P_i - Q_i$ has columns r_{ij} $(j = 1,\dots,N)$, then

$$\|\operatorname{diag}(\|r_{i1}\|,\dots,\|r_{iN}\|)R_i^{-1}\| \leqslant \Omega, \qquad i = 0,1,\dots. \tag{3.15}$$

Our analysis permits (3.15) to be violated so long as $h(P_i, Q_i)$ remains bounded; this is useful in view of certain difference Newton-like methods in which $\|P_i\|_F$ and $\|Q_i\|_F$ are related to $\|x_i - x^*\|$ and may therefore decrease rapidly enough to keep $h(P_i, Q_i)$ bounded, despite violating (3.15). Moreover, our result is in line with one given in [4] in the following sense. It is well known [5], [6] that (3.15) is equivalent to the existence of $\omega > 0$ such that

$$\det\Big(R_i \operatorname{diag}\Big(\|r_{i1}\|^{-1},\dots,\|r_{iN}\|^{-1}\Big)\Big) \geqslant \omega, \qquad i = 0,1,\dots, \tag{3.16}$$

where ω and Ω are inversely proportional. In [4] it was proved that for certain multivariate secant methods convergence is obtainable even if (3.16) is violated. Our results show that for *any* difference Newton-like method condition (3.16) is not a necessary condition for local convergence.

A bound on the rate of convergence of the method (3.5) may be deduced from the inequality (3.7). It is immediately evident that if $\{h(P_i, Q_i)\}$ is bounded below by some positive number, then convergence may be no more than linear, while if $\{h(P_i, Q_i)\}$ converges to zero, then convergence is superlinear. The actual rate of convergence depends on the definitions of P_i and Q_i, which determine how rapidly $\{h(P_i, Q_i)\}$ converges to zero. We note in addition that superlinear convergence is possible even though (3.15) is violated, an observation in line with the results of [4].

4. Applications. In order to illustrate the use and value of the theoretical results we briefly discuss the application of the theory to some particular difference Newton-like methods.

(i) Given $d^i \in R^N$, $d^i \equiv (d_1^i,\dots,d_N^i)$ with $d_j^i \neq 0$ $(j = 1,\dots,N)$, define $Q_i \equiv 0$ and $P_i = (d_1^i e_1,\dots,d_N^i e_N)$. This is the standard one-point approximation [5], [6]. Clearly $h(P_i, Q_i) \equiv \|P_i\|_F = \|d^i\|$.

(ii) Given $d^i \in R^N$, $d^i \equiv (d_1^i,\dots,d_N^i)$ with $d_j^i \neq 0$ $(j = 1,\dots,N)$, define:

$$P_i = \big(d_1^i e_1,\, d_1^i e_1 + d_2^i e_2,\dots,d_1^i e_1 + \cdots + d_N^i e_N\big),$$
$$Q_i = \big(0,\, d_1^i e_1,\dots,d_1^i e_1 + \cdots + d_{N-1}^i e_{N-1}\big).$$

This is another frequently used one-point approximation [4], [5], [6]. Clearly $h(P_i, Q_i) \equiv \|P_i\|_F + \|Q_i\|_F < 2\sqrt{N}\|d^i\|$.

(iii) Given $d^i \in R^N$, $d^i \equiv (d_1^i,\dots,d_N^i)$ with $d^i \neq 0$, define $Q_i \equiv 0$ and $P_i = \|d^i\|T_i$, where $T_i \in \mathfrak{L}(R^N)$ is an orthogonal matrix calculated as follows [6]:

Define $b \in R^N$, $b = (b_1,\dots,b_N)^T$ by $b_1 \equiv \alpha(\|d^i\| + |d_1^i|)$, $b_j \equiv d_j^i$ $(j = 2,\dots,N)$, where

$$\alpha \equiv \begin{cases} +1 & \text{if } d_1^i \geqslant 0, \\ -1 & \text{if } d_1^i < 0; \end{cases}$$

set $2\beta \equiv \|d^i\|(\|d^i\| + |d_1^i|)$ and $c \equiv b/\sqrt{2\beta}$; then $T_i \equiv -\alpha(I - cc^T)$. Then $T_i = T_i^{-1} = T_i^T$ and $P_i e_1 = d^i/\|d_i\|$, hence $h(P_i, Q_i) = \|d^i\|\|T_i\|_F\|T_i\| = \sqrt{N}\|d^i\|$.

For each of the methods (i)–(iii) both the radius of convergence and the rate of convergence thus depend on $\|d^i\|$. Typically d^i is selected such that, for some $k > 0$, $\|d^i\| \leqslant k\|x_i - x_{i-1}\|$, with $x_{-1} \in R^N$ some additional point, or for some $l > 0$, $\|d^i\| \leqslant l\|F(x_i)\|$ $(i = 0, 1,\dots)$. In the first case we obtain

$$\|d^i\| \leqslant k\big(\|x_i - x^*\| + \|x_{i-1} - x^*\|\big);$$

it follows from Theorem (3.3) that the methods are locally convergent and that for some $\gamma > 0$

$$\|x_{i+1} - x^*\| \leqslant \gamma\|x_i - x^*\|\,\|x_{i-1} - x^*\|, \qquad i = 0, 1,\dots,$$

from which follows the classic result [5], [6] that the R-order of convergence is $\frac{1}{2}(1 + \sqrt{5})$. In the second case we obtain

$$\|d^i\| \leqslant l\|F'(x^*)\|\|x_i - x^*\|\left[1 + \tfrac{1}{2}\mu(F, x^*)\|x_i - x^*\|\right],$$

hence these methods too are locally convergent, and for some $\rho > 0$

$$\|x_{i+1} - x^*\| \leqslant \rho\|x_i - x^*\|^2, \qquad i = 0, 1,\dots.$$

Thus quadratic convergence occurs.

Multistep difference methods are slightly more complicated to analyze. We give a fairly general example.

(iv) Given some sequence $\{d^i\} \subseteq R^N$ of vectors such that $d^i, d^{i-1},\dots,d^{i-N+1}$ are linearly independent $(i = 0, 1,\dots)$, and given some starting points $x_0, x_{-1},\dots,x_{-N+1} \in R^N$, define P_i and Q_i by:

$$P_i e_j = x_{i-N+j} + d^{i-N+j} - x_i,$$
$$Q_i e_j = x_{i-N+j} - x_i, \qquad j = 1,\dots,N.$$

Then $R_i \equiv P_i - Q_i = (d^{i-N+1},\dots,d^i)$. Repeated applications of the Minkowski inequality produce

$$\|P_i\|_F + \|Q_i\|_F \leqslant 2\sqrt{N}\,\|x_i - x^*\| + \|R_i\|_F + 2\left(\sum_{j=1}^{N} \|x_{i-N+j} - x^*\|^2\right)^{1/2}$$

Our Theorem (3.3) therefore guarantees local convergence of such a method if the sequence $\{\|d^i\|\} \subseteq R$ is uniformly bounded and condition (3.15) holds. For particular choices of $\{d^i\}$ it may even be possible to prove local convergence without satisfying (3.15).

The choice $d^j = x_{j-1} - x_j$ in (iv) leads to the notorious N-point secant method. It is easy to prove that in this case

$$\|R_i\|_F \leqslant \left(\sum_{j=1}^{N} \|x_{i-N+j-1} - x^*\|^2 \right)^{1/2} + \left(\sum_{j=1}^{N} \|x_{i-N+j} - x^*\|^2 \right)^{1/2},$$

but, since R_i may become singular, this method is dangerous. More generally a choice of d^i with the property $\|d^i\| \leqslant k\|x_{i-1} - x_i\|$ for some $k > 0$ leads to a similar bound on $\|R_i\|_F$, so that if we assume (3.15) holds, then we obtain local convergence in the same way as before. In the latter case it follows from (3.7) and the expression for $h(P_i, Q_i)$ obtained from the bound on $\|R_i\|_F$ that there exists $\tau > 0$ such that

$$\|x_{i+1} - x^*\| \leqslant \tau\|x_i - x^*\| \, \|x_{i-N} - x^*\|, \qquad i = 0, 1, \ldots,$$

from which it follows that the R-order of convergence is the unique positive root of $z^{N+1} - z^N - 1 = 0$ [6]. This analysis covers the generalized N-point secant method, the cyclic N-point secant method and the stabilized N-point secant method discussed in [6]. For the latter two methods the condition (3.15) is automatically satisfied by the choice of the vectors d^i.

Department of Applied Mathematics
University of the Witwatersrand
Johannesburg 2001, South Africa

1. J. C. P. Bus, "Newton-like methods for solving nonlinear simultaneous equations," *Proc. Third Symp. Operations Research*, Mannheim, 1978, pp. 143–152.

2. J. C. P. Bus, *Numerical Solution of Systems of Nonlinear Equations*, Mathematical Centre Tract 122, Mathematisch Centrum, Amsterdam, 1980.

3. P. Deuflhard & G. Heindl, "Affine invariant convergence theorems for Newton's method and extensions to related methods," *SIAM J. Numer. Anal.*, v. 16, 1979, pp. 1–10.

4. J. Jankowska, "Theory of multivariate secant methods," *SIAM J. Numer. Anal.*, v. 16, 1979, pp. 547–562.

5. J. M. Ortega & W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York, 1970.

6. H. Schwetlick, *Numerische Lösung Nichtlinearer Gleichungen*, VEB-DVW, Berlin, 1979.

T. J. Ypma, "Affine invariant convergence results for Newton's method," *BIT*, v. 22, 1982, pp. 108–118.

T. J. Ypma, "Following paths through turning points," *BIT*, v. 22, 1982, pp. 368–383.

9. T. J. Ypma, *Numerical Solution of Systems of Nonlinear Algebraic Equations*, D. Phil. thesis, Oxford, 1982.