SIAM J. NUMER. ANAL.
Vol. 20, No. 2, April 1983

0036-1429/83/2002-0006 $01.25/0

CONVERGENCE RATES FOR NEWTON'S METHOD AT SINGULAR POINTS*

D. W. DECKER,† H. B. KELLER‡ AND C. T. KELLEY§

Abstract. If Newton's method is employed to find a root of a map from a Banach space into itself and the derivative is singular at that root, the convergence of the Newton iterates to the root is linear rather than quadratic. In this paper we give a detailed analysis of the linear convergence rates for several types of singular problems. For some of these problems we describe modifications of Newton's method which will restore quadratic convergence.

1. Introduction.

Suppose F is a Fréchet differentiable mapping between two Banach spaces E_1 and E_2 and suppose $x^* \in E_1$ is a solution of

$$F(x^*)=0. \tag{1.1}$$

This root is called isolated if and only if the derivative at the solution, denoted $F'(x^*)$, is invertible. For isolated solutions, modest additional continuity conditions on $F'(x)$ insure that Newton's method

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

will converge to x^* provided $\|x_0-x^*\|$ is sufficiently small [12]. Indeed if $F'(x)$ is Lipschitz continuous then the convergence is quadratic:

$$\|x_{i+1}-x^*\|\leqq K\|x_i-x^*\|^2, \qquad i=0,1,\cdots. \tag{1.3}$$

This result is essentially the well-known Kantorovich theorem [9].

A root x^* at which $F'(x^*)$ fails to be invertible is variously denoted non-isolated [10], multiple [13], or singular [2]–[4], [6]–[8], [14]–[15], with the latter term employed here. A study of the convergence properties of Newton's method at singular points was first initiated by Rall [13]. It is well known that at branch and limit points of nonlinear functional equations the first Fréchet derivative is singular and an interest in the computation of such solution points [5], [11], [16], [17] has provided much of the motivation for the more recent attention directed toward singular Newton's method problems [1]–[4], [6]–[8], [13]–[15].

The types of convergence behavior previously known for these singular problems may be roughly described as follows. The Banach space E_1 is written as a direct, $N\oplus X$, sum of the null space N of $F'(x^*)$ and an appropriate complementing space X. Then, when the iterates converge, the component of the iterate error in N converges linearly, while the X-component of the error converges with asymptotic order 2 as in (1.3).

The added difficulty of convergence proofs for such singular problems is essentially twofold. First, there may be a family of codimension-one manifolds through x^* on

* Received by the editors February 22, 1982, and in revised form May 25, 1982.

† Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650. The research of this author was supported by the National Science Foundation under grant MCS-81-04254.

‡ Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650. The research of this author was supported by the Army Research Office under Contract DAAG 29-78-C-0011 and by the Department of Energy under Contract EX-76-S-03-0767 Project Agreement No. 12.

§ Department of Mathematics, North Carolina State University, Raleigh, North Carolina 27650. The research of this author was supported by the National Science Foundation under grant MCS-7902659A01.

which $F'(x)$ is singular. Hence initial guesses must be chosen from some region about x^* in which invertibility of $F'(x)$ is assured. Second, one must show that subsequent iterates are well defined, that is, they remain in some region of invertibility.

The results of §§ 2 and 3 do not involve Newton's method and as such may be of independent interest. In § 2 we assume the singularity of $F'(x^*)$ is that of a Fredholm operator of index zero, and then characterize the invertibility of $F'(x)$ in terms of a related finite dimensional operator ("Shür complement") acting on the null space N. Where invertible, an expression for $F'(x)^{-1}$ is determined. The technique employed to determine invertibility was developed in the context of bifurcation theory [10] and has also been employed in the study of Newton's method in $\mathbb{R}^n$ [7]. In § 3 we study the problem of determining appropriate regions in which the operator $F'(x)$ is invertible. In [7], a more general formula for regions of invertibility for a class of finite dimensional problems was presented. It is likely that this analysis would carry over to our setting. However, the regions we employ are simple to construct and are adequate for an analysis of convergence rates. The results clearly indicate the dependence of invertibility regions on the detailed structure of $F'(x)$. Section 4 describes Newton's method and derives a useful expression relating the error of successive iterates. In § 5 we consider the two simplest singular Newton problems. They may be roughly considered to be generalizations of the simple and higher order zero, scalar problems. Although convergence, with overall linear rate $\frac{1}{2}$, for the first problem was known previously [14], it is included here to demonstrate the greater power of the approach derived in §§ 2–4. For the pth order zero problem, $p \geqq 2$, we sharpen, under weaker hypotheses, a result of [15] by demonstrating X-component error convergence of asymptotic order $p+1$. In addition a theorem guaranteeing convergence with initial guesses chosen from a region less demanding of the structure of $F'(x)$ is presented. In § 3 it is shown that problems with more restrictive regions of invertibility of $F'(x)$ than those described in § 5 are possible. The demand that all Newton iterates remain in such regions is hence more difficult to satisfy and such problems are studied in § 6. These new results indicate N-component convergence with linear rate $p/(p+1)$ and X-component convergence of order q for integers $p, q \geqq 2$, dependent on the structure of $F'(x)$. In § 7 we address the problem of accelerating the overall convergence rate of the two problems considered in § 5. We generalize from $\mathbb{R}^n$ to a general Banach space a modified scheme of [6], with overall convergence of asymptotic order $\sqrt[3]{2}$. A new scheme leading to convergence of order $\sqrt{2}$ for the higher order zero problem is then presented.

In § 8 we present simple algebraic examples exhibiting the convergence conclusions of the results detailed in the preceding sections.

2. $F'(x)^{-1}$ near singular points. We consider a nonlinear mapping F, between two Banach spaces, E_1 and E_2. The results of this and the following sections will place clear differentiability requirements on F, which in each case will be assumed. We shall not require $F(x^*)=0$ here, but rather shall suppose that at the location $x^* \in E_1$, the Fréchet derivative of F, denoted $F'(x^*)$, is a Fredholm operator of index zero. That is, there exists a finite dimensional subspace $N_1 \subseteq E_1$ and a closed subspace $X_2 \subseteq E_2$ such that

$$\mathcal{N}(F'(x^*)) = N_1, \quad \mathcal{R}(F'(x^*)) = X_2, \quad \operatorname{codim}(X_2) = \dim(N_1). \tag{2.1}$$

We choose complementing subspaces X_1, N_2 such that

$$E_1 = N_1 \oplus X_1, \qquad E_2 = N_2 \oplus X_2 \tag{2.2}$$

and define for $i=1,2$ the projections P_{N_i} onto N_i parallel to X_i and $P_{X_i}=I-P_{N_i}$. We define the following operators

$$\begin{aligned} A(x) &\equiv P_{X_2}F'(x)P_{X_1}, \\ B(x) &\equiv P_{X_2}F'(x)P_{N_1}, \\ C(x) &\equiv P_{N_2}F'(x)P_{X_1}, \\ D(x) &\equiv P_{N_2}F'(x)P_{N_1}. \end{aligned} \tag{2.3}$$

and hence

$$F'(x)=A(x)+B(x)+C(x)+D(x). \tag{2.4}$$

Letting

$$A(x^*)=P_{X_2}F'(x^*)P_{X_1}\equiv \hat{F}, \tag{2.5}$$

we see $\hat{F}$ has an inverse when viewed as an operator from X_1 into X_2. For x near x^* it is possible to characterize the invertibility of $F'(x)$, and further, to derive an expression for $F'(x)^{-1}$ when it exists. This we do in the following theorem.

THEOREM 2.6. *For ρ sufficiently small and $x\in S_\rho\equiv\{x\in E_1 \mid \|x-x^*\|\leqq\rho\}$, $F'(x)$ is nonsingular if and only if*

$$\tilde{D}(x)\equiv D(x)-C(x)A^{-1}(x)B(x) \tag{2.7}$$

is nonsingular (viewed as a mapping from N_1 into N_2). Further, for ρ sufficiently small and $\tilde{D}(x)$ invertible we have

$$\begin{aligned} F'(x)^{-1}=P_{X_1}[A(x)^{-1}+A^{-1}(x)B(x)\tilde{D}(x)^{-1}C(x)A^{-1}(x)]P_{X_2} \\ -P_{X_1}A(x)^{-1}B(x)\tilde{D}(x)^{-1}P_{N_2} \\ -P_{N_1}\tilde{D}(x)^{-1}C(x)A(x)^{-1}P_{X_2}+P_{N_1}\tilde{D}(x)^{-1}P_{N_2}. \end{aligned} \tag{2.8}$$

Proof. We consider the operator $M: X_1\oplus N_1\to X_2\oplus N_2$ defined by

$$M(x)\equiv\begin{pmatrix} A(x) & B(x) \\ C(x) & D(x)\end{pmatrix}. \tag{2.9}$$

From (2.4) it is clear that $F'(x)$ is invertible if and only if $M(x)$ is invertible. Since $A(x^*): X_1\to X_2$ is invertible, so is $A(x)$ for ρ sufficiently small. But when $A(x)$ is invertible, $M(x)$ is invertible if and only if the Schür complement $\tilde{D}(x)$ is invertible. To compute $F'(x)^{-1}$ assume ρ is sufficiently small and $\tilde{D}(x)$ is invertible. Define the operator $\hat{M}(x): X_2\oplus N_2\to X_1\oplus N_1$ as

$$\hat{M}(x)\equiv\begin{pmatrix} \hat{A}(x) & \hat{B}(x) \\ \hat{C}(x) & \hat{D}(x)\end{pmatrix} \tag{2.10}$$

where

$$\begin{aligned} \hat{A}(x) &= P_{X_1}[A(x)^{-1}+A(x)^{-1}B(x)\tilde{D}(x)^{-1}C(x)A(x)^{-1}]P_{X_2}, \\ \hat{B}(x) &= -P_{X_1}A(x)^{-1}B(x)\tilde{D}(x)^{-1}P_{N_2}, \\ \hat{C}(x) &= -P_{N_1}\tilde{D}(x)^{-1}C(x)A(x)^{-1}P_{X_2}, \\ \hat{D}(x) &= P_{N_1}\tilde{D}(x)^{-1}P_{N_2}. \end{aligned} \tag{2.11}$$

Then one may directly verify $\hat{M}(x)=M(x)^{-1}$. From this we have

$$\begin{aligned}\begin{pmatrix}P_{X_1}\\P_{N_1}\end{pmatrix}&=\begin{pmatrix}\hat{A}(x)&\hat{B}(x)\\\hat{C}(x)&\hat{D}(x)\end{pmatrix}\begin{pmatrix}A(x)&B(x)\\C(x)&D(x)\end{pmatrix}\begin{pmatrix}P_{X_1}\\P_{N_1}\end{pmatrix}\\&=\begin{pmatrix}\hat{A}(x)&\hat{B}(x)\\\hat{C}(x)&\hat{D}(x)\end{pmatrix}\begin{pmatrix}P_{X_2}F'(x)\\P_{N_2}F'(x)\end{pmatrix}.\end{aligned}\tag{2.12}$$

Adding the components on each side of (2.12) we have

$$I=(\hat{A}(x)+\hat{B}(x)+\hat{C}(x)+\hat{D}(x))F'(x)$$

which from the definitions (2.11) yields the conclusion (2.8). □

This result now shifts the question of invertibility of $F'(x)$ to that of $\tilde{D}(x)$.

3. Regions of invertibility of $F'(x)$. As we must now introduce more notation we first take the simplifying step of setting $E_1=E_2$, eliminating the subscripts on projections and subspaces. At any stage the method of their reinstatement should be clear.

We set $\tilde{x}=x-x^*$ and define the candidates for regions of invertibility as

$$W(\rho,\theta,m)=\{x\in E\,|\,0<\|\tilde{x}\|\leqq\rho,\ \|P_X\tilde{x}\|\leqq\theta\|P_N\tilde{x}\|^m\}.\tag{3.1}$$

Here $m\geqq 1$ and θ will generally be required to be small. As will be seen later, the slowest convergence behavior for Newton's method is to be expected in the null space directions, motivating the choice of regions "weighted" toward N. The approach will be to choose the smallest m still guaranteeing invertibility.

We define $\beta_k(x)$ as any element of E or any operator on E whose norm is at least $O(\|x-x^*\|^k)$. At times this order symbol is used in the following extended sense. In writing, say, $F_1(x)=F_2(x)+\beta_k(x)$, it may occur that $F_2(x)=o(\|\tilde{x}\|^k)$ in which case the previous equation merely states $\|F_1(x)\|$ is at least order k.

Now for some $p\geqq 1$, and $n\geqq p$ we may write

$$F'(x)=F'(x^*)+\sum_{k=p}^{n}\frac{1}{k!}F^{(k+1)}(x^*)(\tilde{x}^k,\cdot)+\beta_{n+1}(x)\tag{3.2}$$

where $F^{(k+1)}(x^*)\,(\cdot,\cdot,\cdots,\cdot)$ is a multilinear $k+1$ form and $\tilde{x}^k$ indicates the first k arguments are all $\tilde{x}$.

Recalling (2.4), we have, for some integers a, b, c, d

$$\begin{aligned}A(x)&=P_XF'(x^*)P_X+\sum_{k=a}^{n}\frac{1}{k!}P_XF^{(k+1)}(x^*)(\tilde{x}^k,P_X\cdot)+\beta_{n+1}(x)\\&\equiv\hat{F}+\sum_{k=a}^{n}A_k(x)+\beta_{n+1}(x),\\B(x)&=\sum_{k=b}^{n}\frac{1}{k!}P_XF^{(k+1)}(x^*)(\tilde{x}^k,P_N\cdot)+\beta_{n+1}(x)\\&\equiv\sum_{k=b}^{n}B_k(x)+\beta_{n+1}(x),\\C(x)&=\sum_{k=c}^{n}\frac{1}{k!}P_NF^{(k+1)}(x^*)(\tilde{x}^k,P_X\cdot)+\beta_{n+1}(x)\\&\equiv\sum_{k=c}^{n}C_k(x)+\beta_{n+1}(x),\end{aligned}\tag{3.3}$$

$$D(x)=\sum_{k=d}^{n}\frac{1}{k!}P_NF^{(k+1)}(x^*)(\tilde{x}^k, P_N\cdot)+\beta_{n+1}(x)$$

$$\equiv\sum_{k=d}^{n}D_k(x)+\beta_{n+1}(x).$$

Thus, $p=\min(a, b, c, d)$ and we choose $n\geqq\max(a, b, c, d)$. We further define

$$\bar{A}_k(x)\equiv\frac{1}{k!}P_XF^{(k+1)}(x^*)((P_N\tilde{x})^k, P_X\cdot), \tag{3.4}$$

etc., and assume $\bar{a}$, $\bar{b}$, $\bar{c}$, $\bar{d}$ to be the smallest integers such that for some $P_N\tilde{x}\neq 0$:

$$\begin{aligned}\bar{A}_{\bar{a}}(x)&=\frac{1}{\bar{a}!}P_XF^{(\bar{a}+1)}(x^*)((P_N\tilde{x})^{\bar{a}}, P_X\cdot)\neq 0,\\ \bar{B}_{\bar{b}}(x)&=\frac{1}{\bar{b}!}P_XF^{(\bar{b}+1)}(x^*)((P_N\tilde{x})^{\bar{b}}, P_N\cdot)\neq 0,\\ \bar{C}_{\bar{c}}(x)&=\frac{1}{\bar{c}!}P_NF^{(\bar{c}+1)}(x^*)((P_N\tilde{x})^{\bar{c}}, P_X\cdot)\neq 0,\\ \bar{D}_{\bar{d}}(x)&=\frac{1}{\bar{d}!}P_NF^{(\bar{d}+1)}(x^*)((P_N\tilde{x})^{\bar{d}}, P_N\cdot)\neq 0.\end{aligned} \tag{3.5}$$

Note that $\bar{a}\geqq a$, $\bar{b}\geqq b$, $\bar{c}\geqq c$ and $\bar{d}\geqq d$. Now for any k, $m\geqq 1$ and for $x\in W(\rho, \theta, m)$ we note, recalling (3.1):

$$F^{(k+1)}(x^*)(\tilde{x}^k, \cdot)=F^{(k+1)}(x^*)((P_N\tilde{x})^k, \cdot)+\theta\beta_{m+k-1}(x). \tag{3.6}$$

With this in mind we see that

$$\begin{aligned}A(x)&=\hat{F}+\bar{A}_{\bar{a}}(x)+\beta_{\bar{a}+1}(x)+\theta\beta_{a+m-1}(x),\\ B(x)&=\bar{B}_{\bar{b}}(x)+\beta_{\bar{b}+1}(x)+\theta\beta_{b+m-1}(x),\\ C(x)&=\bar{C}_{\bar{c}}(x)+\beta_{\bar{c}+1}(x)+\theta\beta_{c+m-1}(x),\\ D(x)&=\bar{D}_{\bar{d}}(x)+\beta_{\bar{d}+1}(x)+\theta\beta_{d+m-1}(x).\end{aligned} \tag{3.7}$$

It should be noted that m is not yet determined, and the choice of m to be made shortly may force the last order symbol to be the dominant term in one or more of (3.7).

Assuming ρ so small that $A^{-1}(x)$ exists we have

$$A^{-1}(x)=\hat{F}^{-1}+\beta_{\bar{a}}(x)+\theta\beta_{a+m-1}(x)=\hat{F}^{-1}+\beta_1(x). \tag{3.8}$$

Now (3.7–8) in (2.7) yields

$$\begin{aligned}\tilde{D}(x)=\bar{D}_{\bar{d}}(x)&-\bar{C}_{\bar{c}}(x)\hat{F}^{-1}\bar{B}_{\bar{b}}(x)\\ &+\theta(\beta_{d+m-1}(x)+\beta_{\bar{b}+c+m-1}(x)+\beta_{b+\bar{c}+m-1}(x)+\theta\beta_{b+c+2m-2}(x))\\ &+\beta_{\bar{b}+\bar{c}+1}(x)+\beta_{\bar{d}+1}(x).\end{aligned} \tag{3.9}$$

The nonsingularity of the dominant term of (3.9) will be shown to be sufficient to guarantee the invertibility of $\tilde{D}(x)$. We isolate the term we shall later require to be dominant by defining the operator $H(x)$ as:

$$H(x)=\begin{cases}\bar{D}_{\bar{d}}(x) & \text{if } \bar{d}<\bar{b}+\bar{c},\\ \bar{D}_{\bar{d}}(x)-\bar{C}_{\bar{c}}(x)\hat{F}^{-1}\bar{B}_{\bar{b}}(x) & \text{if } \bar{d}=\bar{b}+\bar{c},\\ -\bar{C}_{\bar{c}}(x)\hat{F}^{-1}\bar{B}_{\bar{b}}(x) & \text{if } \bar{d}>\bar{b}+\bar{c}.\end{cases} \tag{3.10}$$

We use this to state

LEMMA 3.11. *Let $\rho>0$ be such that $A(x)$ is invertible for $x\in S_\rho$. Assume that for all x with $P_N\tilde{x}\neq 0$ the operator $H(x)$ is invertible as an operator on N. Then there exist $\bar\rho>0$, $\bar\theta>0$ such that $F'(x)$ is invertible for all $x\in W(\bar\rho,\bar\theta,m)$ where $m=1+q$ and q is the smallest nonnegative integer satisfying*

$$\min(d+q,\bar b+c+q,b+\bar c+q,b+c+2q)\geqq\min(\bar d,\bar b+\bar c). \tag{3.12}$$

Proof. First, by assumption $A(x)^{-1}$ exists so $\tilde D(x)$ is well defined. From the definition (3.10) of $H(x)$ we see that it has order exactly $\min(\bar d,\bar b+\bar c)$ and that further

$$\tilde D(x)=H(x)+\theta\beta_r(x)+\beta_{\bar b+\bar c+1}(x)+\beta_{\bar d+1}(x), \tag{3.13}$$

where

$$r=\min(d+m-1,\bar b+c+m-1,b+\bar c+m-1,b+c+2(m-1)). \tag{3.14}$$

Let $q=m-1$ be the minimum integer satisfying (3.12). If the inequality holds, select any $\bar\theta>0$ and then choose $\bar\rho$ so small that the invertibility of $\tilde D(x)$ for $x\in W(\bar\rho,\bar\theta,m)$ follows from that of $H(x)$ for all x with $P_N\tilde x\neq 0$. If the equality holds, select a ρ, and then choose $\bar\theta$ so small that $H(x)+\theta\beta_1(x)$ is invertible in $W(\rho,\bar\theta,m)$. Now if necessary, shrink ρ to $\bar\rho$ in order that the final two terms of (3.13) cannot prevent invertibility of $\tilde D(x)$ for all $x\in W(\bar\rho,\bar\theta,m)$. □

The values of a, b, c, etc., are determined by the problem and in each case the appropriate m may be chosen through (3.12). $W(\bar\rho,\bar\theta,m)$ is then a valid region from which to select at least an initial Newton iterate. In this context, problems allowing the largest regions of invertibility are of interest, and sufficient conditions for this may be stated as

COROLLARY 3.15. *Under the assumptions of Lemma 3.11, $F'(x)$ is invertible in a region of type $W(\bar\rho,\bar\theta,1)$ provided:*

(3.16)
(i) $d=\bar d$ *when* $d<b+c$;
(ii) $d=\bar d$ *or* $b=\bar b$ *and* $c=\bar c$ *when* $d=b+c$;
(iii) $b=\bar b$ *and* $c=\bar c$ *when* $d>b+c$.

Proof. The relation (3.12) is satisfied for $q=0$ if and only if $\min(\bar d,\bar b+\bar c)=\min(d,b+c)$. But $\bar d\geqq d$, $\bar b+\bar c\geqq b+c$ and hence if $d<b+c$ then $d<\bar b+\bar c$ and thus $d=\bar d$. The other cases are similar. □

It is often possible to get an explicit formula for m as we show in

COROLLARY 3.17. *If the hypotheses of Lemma 3.11 hold and $\bar d>\bar b+\bar c$, then $F'(x)$ is invertible in region of type $W(\rho,\theta,m)$ where*

$$m=1+\bar b+\bar c-\min(d,\bar b+c,b+\bar c). \tag{3.18}$$

Proof. In (3.12) the assumption $\bar d>\bar b+\bar c$ forces $\bar b+c+q\geqq\bar b+c$ and $b+\bar c+q\geqq\bar b+\bar c$, and hence by adding these last two inequalities $b+c+2q\geqq\bar b+\bar c$. Thus (3.12) becomes

$$\min(d+q,\bar b+c+q,b+\bar c+q)\geqq\bar b+\bar c \tag{3.19}$$

and we see the minimum choice of q will achieve equality, resulting in (3.18). □

We note that the type of problem covered by case (3.16i) was considered by Reddien [14] and Decker and Kelley [2] when $d=1$ and by Reddien [15] and Griewank

[7] for $d>1$ with $b, c \geqq d$. The situation $d=b=\bar{b}=c=\bar{c}=1$ but $\bar{d}=2$, yielding a choice of $m=2$ in (3.18) was detailed by Decker and Kelley in [3].

Although Lemma 3.11 will be used most frequently it is of advantage, when $\dim(N)>1$, to state a result giving invertibility of $F'(x)$ in a smaller region about x^*, but under weaker assumptions. In this we make use of and extend the approach of Reddien [15].

We consider a one-dimensional subspace $M \subseteq N$ and let $N=M \oplus M_0$ for some M_0. Define the projection P_M onto M parallel to $X \oplus M_0$. We define new candidates for regions of invertibility as

$$W(\rho,\theta,\eta,m)=\{x\in E \mid 0<\|\tilde{x}\|\leqq\rho, \|P_X\tilde{x}\|\leqq\theta\|P_N\tilde{x}\|^m, \|(P_N-P_M)\tilde{x}\|\leqq\eta\|P_N\tilde{x}\|\}. \tag{3.20}$$

We next modify the operator $H(x)$ of (3.10) by replacing $P_N\tilde{x}$ by $P_M\tilde{x}$ in (3.5). We shall call this new operator $\check{H}(x)$. That is, for example, in the first case of (3.10)

$$\check{H}(x)\equiv\check{D}_{\bar{d}}(x)\equiv\frac{1}{\bar{d}!}P_N F^{(\bar{d}+1)}(x^*)((P_M\tilde{x})^{\bar{d}}, P_N\,\cdot). \tag{3.21}$$

Using this we may state

LEMMA 3.22. *Assume there exists an M such that for all $\tilde{x}$ with $P_M\tilde{x}\neq 0$, $\check{H}(x)$ is nonsingular as a map on N. Then the conclusions of Lemma 3.11 and Corollaries 3.15, 3.17 remain valid with $W(\bar{\rho},\bar{\theta},m)$ replaced by $W(\bar{\rho},\bar{\theta},\bar{\eta},m)$ and $\bar{\eta}$ sufficiently small.*

Proof. We see that for $x\in W(\rho,\theta,\eta,m)$

$$H(x)=\check{H}(x)+\eta\beta_s(x) \tag{3.23}$$

where $s=\min(\bar{d},\bar{b}+\bar{c})$. But the order of $\check{H}(x)$, by assumption, is exactly s for $x\in W(\rho,\theta,\eta,m)$ and so invertibility of $H(x)$ in this set follows from that of $\check{H}(x)$ for $\bar{\eta}$ sufficiently small. But now the choice of m, $\bar{\rho}$, $\bar{\theta}$ will be made in exactly the same fashion as in the proof of Lemma 3.11 to ensure that the invertibility of $\tilde{D}(x)$ is guaranteed by the nonsingularity of $H(x)$. □

Finally we note that our study of the invertibility of $\tilde{D}(x)$ is incomplete in at least the following respect. The operators $H(x)$ and $\check{H}(x)$ attempt to focus on the dominant term of $\tilde{D}(x)$. Suppose that the third alternative in (3.10) arises but the composite operator $\bar{C}_{\bar{c}}(x)\hat{F}^{-1}\bar{B}_{\bar{b}}(x)\equiv 0$. It may then be that the dominant order term lies in one of the order symbols of (3.9). The procedure for the expansion of $\tilde{D}(x)$ is straightforward, however, and a direct inspection in such a case will readily identify the correct leading term.

4. Newton's method. In this section we determine a formula for successive Newton iterates that will be widely employed in subsequent work. When $F'(x)$ is invertible, the next Newton iterate is defined as

$$y=x-F'(x)^{-1}F(x). \tag{4.1}$$

Hence with $\tilde{y}\equiv y-x^*$, $\tilde{x}\equiv x-x^*$:

$$\tilde{y}=\tilde{x}-F'(x)^{-1}F(x). \tag{4.2}$$

Now, for any $n\geqq p$:

$$F(x)=\hat{F}\tilde{x}+\sum_{j=p}^{n}\frac{1}{(j+1)!}F^{(j+1)}(x^*)(\tilde{x}^{j+1})+\beta_{n+2}(x). \tag{4.3}$$

Recalling the definitions (3.3):

$$F(x)=\hat{F}\tilde{x}+\sum_{j=p}^{n}\frac{1}{j+1}[A_j(x)+B_j(x)+C_j(x)+D_j(x)]\tilde{x}+\beta_{n+2}(x). \tag{4.4}$$

But

$$F'(x)=\hat{F}+\sum_{j=p}^{n}\frac{1}{j!}F^{(j+1)}(x^*)(\tilde{x}^j,\cdot)+\beta_{n+1}(x) \tag{4.5}$$

and hence with (3.3) again:

$$\hat{F}\tilde{x}=F'(x)\tilde{x}-\sum_{j=p}^{n}[A_j(x)+B_j(x)+C_j(x)+D_j(x)]\tilde{x}+\beta_{n+2}(x). \tag{4.6}$$

Using (4.6) in (4.4) we get

$$F(x)=F'(x)\tilde{x}-\sum_{j=p}^{n}\frac{j}{j+1}[A_j(x)+B_j(x)+C_j(x)+D_j(x)]\tilde{x}+\beta_{n+2}(x). \tag{4.7}$$

Using this in (4.2) we have finally:

$$\tilde{y}=F'(x)^{-1}\left\{\sum_{j=p}^{n}\frac{j}{j+1}[A_j(x)+B_j(x)+C_j(x)+D_j(x)]\tilde{x}+\beta_{n+2}(x)\right\}. \tag{4.8}$$

This formula, coupled with the expression (2.8) for $F'(x)^{-1}$ will provide the basis for the convergence analysis of the Newton iterates.

Before proceeding, we define a new order symbol $\gamma_p^q(x)$ which will be quite useful. This symbol will represent any term with order at least $O(\|\tilde{x}\|^p)$, but further, the operator is such that $P_X\gamma_p^q(x)=\beta_{p+q}(x)$. For example,

$$P_NF''(x^*)(\tilde{x}^2)+P_XF'''(x^*)(\tilde{x}^3)$$

could be represented as $\gamma_2^1(x)$. With this definition we have the rule

$$\gamma_p^q(x)\gamma_r^s(x)=\gamma_{p+r}^q(x)$$

and in particular

$$\gamma_p^q(x)\cdot\beta_r(x)=\gamma_{p+r}^q(x).$$

Note that these new order symbols do not commute while the β-symbols do.

5. The cases $c\geqq d=\bar{d}\geqq 1$. We begin this section with a result from [2] to indicate the simplicity and improved efficiency of the method of proof when the results of §§ 2–4 are employed.

THEOREM 5.1. *Assume, for all $P_N\tilde{x}\neq 0$, that $\bar{D}_1(x)$ is nonsingular as a map on N. Then, for ρ and θ sufficiently small, $F'(x_0)^{-1}$ exists for all $x_0\in W(\rho,\theta,1)$; further, all subsequent Newton iterates remain in this set and converge to x^* with rate determined by*

$$\|P_X(x_i-x^*)\|\leqq K_1\|x_{i-1}-x^*\|^2,\quad \textit{some } K_1>0, \tag{5.2}$$

and

$$\lim_{i\to\infty}\frac{\|P_N(x_i-x^*)\|}{\|P_N(x_{i-1}-x^*)\|}=\frac{1}{2},\qquad i=1,2,\cdots. \tag{5.3}$$

Proof. For this case $d=\bar{d}=1$, b, $c\geqq 1$ and hence, by Corollary 3.15, $F'(x)^{-1}$ exists in $W(\rho,\theta,1)$ for sufficiently small ρ and θ. In this region, from (2.7) and (3.3),

$\tilde{D}(x)=D_1(x)+\beta_2(x)$ and $D_1(x)^{-1}$ exists as a map on N. But then $\tilde{D}(x)^{-1}=D_1(x)^{-1}+\beta_0(x)$ and so from (2.8), $F'(x)^{-1}=P_ND_1(x)^{-1}P_N+\beta_0(x)=\gamma_{-1}^1(x)$. But now from (4.8) with $p=1$:

$$\tilde{x}_1=\tfrac{1}{2}[P_ND_1(x_0)^{-1}P_N+\beta_0(x_0)] \times\{[A_1(x_0)+B_1(x_0)+C_1(x_0)+D_1(x_0)]\tilde{x}_0+\beta_3(x_0)\}. \tag{5.4}$$

However we recall that: $P_NA_1(x_0)=P_NB_1(x_0)=0$. Further $C_1(x_0)\tilde{x}_0=\theta\beta_2(x_0)$ and so now $F'(x_0)^{-1}C_1(x_0)\tilde{x}_0=\theta\gamma_1^1(x_0)$. Thus we have from these observations in (5.4):

$$\tilde{x}_1=\tfrac{1}{2}P_N\tilde{x}_0+\theta\gamma_1^1(x_0)+\beta_2(x_0). \tag{5.5}$$

This is the crucial relation and equations of similar form will arise in the proofs of many of the results that are to follow. These equations, in each case, directly provide the convergence rate conclusions. We shall proceed in detail for this situation only as the other derivations follow in a closely analogous manner.

First, we have from (5.5) for some $K_0, K_1>0$:

$$(\tfrac{1}{2}-K_0\theta)\|P_N\tilde{x}_0\|\leqq\|P_N\tilde{x}_1\|\leqq(\tfrac{1}{2}+K_0\theta)\|P_N\tilde{x}_0\|, \tag{5.6}$$

$$\|P_X\tilde{x}_1\|\leqq K_1\|(x_0-x^*)\|^2. \tag{5.7}$$

If we define the sequences $\rho_i=\|x_i-x^*\|$, $\theta_0=\theta$, and $\theta_i=K_1(1+\theta_{i-1})\times(\tfrac{1}{2}-K_0\theta_{i-1})^{-1}\rho_{i-1}$ for $i\geqq1$, then since $x_0\in W(\rho_0,\theta_0,1)$ equations (5.6)–(5.7) yield

$$\begin{aligned}\|P_X\tilde{x}_1\|&\leqq K_1\rho_0(1+\theta_0)\|P_N\tilde{x}_0\|\leqq K_1\rho_0(1+\theta_0)(\tfrac{1}{2}-K_0\theta_0)^{-1}\|P_N\tilde{x}_1\|,\\ \rho_1&\leqq\|P_N\tilde{x}_1\|+\|P_X\tilde{x}_1\|\leqq((\tfrac{1}{2}+K_0\theta_0)(1-\theta_0)^{-1}+K_1\rho_0)\rho_0.\end{aligned} \tag{5.8}$$

Thus $\|P_X\tilde{x}_1\|\leqq\theta_1\|P_N\tilde{x}_1\|$ and $x_1\in W(\rho_1,\theta_1,1)$. Given some $r\in(\tfrac{1}{2},1)$ we see from (5.8) and the definition of θ_1 that we may choose ρ_0, θ_0 so small that $\rho_1<r\rho_0$, $\theta_1<r\theta_0$. Then since $W(\rho_1,\theta_1,1)\subseteq W(\rho_0,\theta_0,1)$ we may replace x_0, x_1, θ_0 by x_1, x_2, θ_1 in (5.6)–(5.8) and hence $\rho_2<r\rho_1$, $\theta_2<r\theta_1$. In this fashion we see $\lim_{i\to\infty}\rho_i=\lim_{i\to\infty}\theta_i=0$. Thus $x_i\to x^*$, (5.7) becomes (5.2) and letting $i\to\infty$ in

$$(\tfrac{1}{2}-K_0\theta_{i-1})\|P_N\tilde{x}_{i-1}\|\leqq\|P_N\tilde{x}_i\|\leqq(\tfrac{1}{2}+K_0\theta_{i-1})\|P_N\tilde{x}_{i-1}\|$$

provides the conclusion (5.3). □

We note that a more careful analysis of the $\beta_0(x)$ term in (5.4) will, for $i\geqq2$, allow the conclusion $\|P_X\tilde{x}_i\|=\beta_3(x_{i-1})$ [4], [6]. We omit this analysis for simplicity.

A general result of this type is:

THEOREM 5.9. *Assume $d=\bar{d}\leqq c$ and $\tilde{D}_d$ is nonsingular for all $P_N\tilde{x}\neq0$. Then for ρ and θ sufficiently small, $F'(x_0)^{-1}$ exists for all $x_0\in W(\rho,\theta,1)$, all subsequent iterates remain in this set and $x_i\to x^*$ with rate determined by*

$$\|P_X(x_i-x^*)\|\leqq K_1\|x_{i-1}-x^*\|^{p+1} \quad \textit{some } K_1>0 \tag{5.10}$$

and

$$\lim_{i\to\infty}\frac{\|P_N(x_i-x^*)\|}{\|P_N(x_{i-1}-x^*)\|}=\frac{d}{d+1}. \tag{5.11}$$

Proof. In this case $\bar{d}<\bar{b}+\bar{c}$ and Corollary 3.15 guarantees that $F'(x)^{-1}$ exists in $W(\rho,\theta,1)$ for ρ and θ sufficiently small. Since

$$\tilde{D}(x)=D(x)+\beta_{2d}(x)=D_d(x)+\beta_{d+1}(x) \tag{5.12}$$

and $D_d(x)^{-1}$ exists in $W(\rho, \theta, 1)$, we have, in this region

$$\tilde{D}(x)^{-1} = D_d(x)^{-1} + \beta_{-d+1}(x). \tag{5.13}$$

But then an examination of (2.8) shows the first three terms of $F'(x_0)^{-1}$ to be $P_X\beta_{b-d}(x)P_N + \beta_0(x)$. Hence

$$F'(x)^{-1} = P_N D_d(x)^{-1} P_N + P_X\beta_{b-d}(x)P_N + \gamma^{d-1}_{-d+1}(x). \tag{5.14}$$

For our present situation, (4.8) and (5.14) imply

$$\begin{aligned} \tilde{x}_1 = {} & [P_N D_d(x)^{-1} P_N + P_X\beta_{b-d}(x)P_N + \gamma^{d-1}_{-d+1}(x)] \\ & \times\left\{\frac{p}{p+1}[A_p(x_0) + B_p(x_0)]\tilde{x}_0 + P_X\beta_{p+2}(x_0)\right. \\ & \left. + \frac{d}{d+1}[D_d(x_0) + C_d(x_0)]\tilde{x}_0 + P_N\beta_{d+2}(x_0)\right\}. \end{aligned} \tag{5.15}$$

As $F'(x_0)^{-1}P_X = \beta_0(x_0)$,

$$\begin{aligned} \tilde{x}_1 = {} & \frac{d}{d+1}[P_N D_d(x_0)^{-1} P_N + P_X\beta_{b-d}(x_0)P_N] \\ & \times[(C_d(x_0) + D_d(x_0))\tilde{x}_0 + \beta_{d+2}(x_0)] + \beta_{p+1}(x_0) \end{aligned} \tag{5.16}$$

and hence

$$\tilde{x}_1 = \frac{d}{d+1} P_N\tilde{x}_1 + \theta\gamma^p_1(x_0) + \beta_{p+1}(x_0). \tag{5.17}$$

From this we have

$$P_N\tilde{x}_1 = \frac{d}{d+1} P_N\tilde{x}_0 + \theta\beta_1(x_0), \tag{5.18}$$

$$P_X\tilde{x}_1 = \beta_{p+1}(x_0). \tag{5.19}$$

The conclusions of the theorem now follow from (5.18)–(5.19) in precisely the same fashion detailed in the previous proof. □

A similar result was derived in Reddien [15] but under the additional restrictions that dim $(N) = 1$, and that for all $x \in E$, $\phi \in N$, and some $c > 0$

$$\|F^{(p+1)}(x^*)(\tilde{x}^p, \phi)\| \geqq c\|\phi\|\,\|\tilde{x}\|^p. \tag{5.20}$$

In addition, the convergence rate obtained corresponding to (5.10) had $(p+1)$ reduced to 2. For finite dimensions Griewank [7] obtained a similar result.

The assumption $\bar{D}_d$ invertible on N for all $P_N\tilde{x} \neq 0$ is quite strong and, if $d = 1$, implies that dim $N = 1$ or 2. This assumption may be weakened [15] by employing Lemma 3.22 rather than Lemma 3.11. We have

THEOREM 5.21. *Assume there exists a one-dimensional subspace $M \subseteq N$ such that $\check{D}_d(x)$ is nonsingular on N for all $P_M\tilde{x} \neq 0$. Then for $\bar{\rho}$, $\bar{\theta}$, $\bar{\eta}$ sufficiently small $F'(x)^{-1}$ exists for all $x \in W(\bar{\rho}, \bar{\theta}, \bar{\eta}, 1)$. Further, there exists $\hat{\eta} \leqq \bar{\eta}$ such that for any $x_0 \in W(\bar{\rho}, \bar{\theta}, \hat{\eta}, 1)$ all subsequent iterates stay in $W(\bar{\rho}, \bar{\theta}, \bar{\eta}, 1)$ and $x_i \to x^*$ with rate determined by (5.10)–(5.11).*

Proof. The added difficulty in the proof of this result is to conclude all iterates remain in $W(\bar{\rho}, \bar{\theta}, \bar{\eta}, 1)$. This analysis is straightforward, but lengthy, and hence is omitted. □

We conclude this section with some remarks on the geometric isolation of the root x^* [10]. From Theorems 5.1 and 5.9, all initial iterates in $W(\rho, \theta, 1)$ yield Newton iterates converging to x^*, and since $x^* \notin W(\rho, \theta, 1)$ there is no root in this region. Now suppose there is a root $y^* \neq x^*$ with $\|\tilde{y}^*\| \leqq \rho$. Then we have for some $p \geqq 1$, and with $\tilde{y}^* \equiv y^* - x^*$:

$$0 = F(y^*) = F'(x^*)\tilde{y}^* + \frac{1}{(p+1)!} F^{(p+1)}(x^*)(\tilde{y}^{*p+1}) + \beta_{p+2}(y^*). \tag{5.22}$$

Hence

$$P_X\tilde{y}^* = \frac{1}{(p+1)!} \hat{F}^{-1} F^{(p+1)}(x^*)(\tilde{y}^{*p+1}) + \beta_{p+2}(y^*) \tag{5.23}$$

and from this we see $P_X\tilde{y}^* = \beta_{p+1}(y^*)$ where $p+1 \geqq 2$. But since $P_N\tilde{y}^* + P_X\tilde{y}^* = \beta_1(y^*)$ we find that for any given θ, $\|P_X\tilde{y}^*\| \leqq \theta \|P_N\tilde{y}^*\|$ for ρ chosen sufficiently small, and hence $y^* \in W(\rho, \theta, 1)$. This contradiction implies there are no other roots of F in a sufficiently small sphere about x^*.

The same argument may be applied to the case in which the initial iterates are chosen from $W(\rho, \theta, m)$. The requirement for geometric isolation of x^* then becomes $p+1 > m$, the satisfaction of which may be directly verified for all of the results which follow. Hence the only case in which geometric isolation is not assured is that described by Theorem 5.21.

6. Higher order regions of invertibility. In accordance with Lemma 3.11 a choice of $m > 1$ is often required to guarantee a region of invertibility. However, in applying Newton's method, the demand for all iterates to remain in such regions becomes more difficult to satisfy. A problem of this type, with $p = d = 1$ but $\bar{d} = p + 1$, leading to a choice of $m = 2$, was studied in [3]. We now extend this result to problems with $p > 1$.

THEOREM 6.1. *Assume $p = d \geqq 2$, $\bar{d} = p + 1$, and that $\bar{D}_{p+1}(x)^{-1}$ exists as a map on N for all $\tilde{x}$ with $P_N\tilde{x} \neq 0$. Then for ρ, θ sufficiently small $F'(x_0)^{-1}$ exists for all $x_0 \in W(\rho, \theta, 2)$, all subsequent Newton iterates remain in this set, and $x_i \to x^*$ with rate determined by*

$$\|P_X(x_i - x^*)\| \leqq K_1 \|x_{i-1} - x^*\|^{p+1}, \quad \textit{some } K_1 > 0 \tag{6.2}$$

and

$$\lim_{i\to\infty} \frac{\|P_N(x_i - x^*)\|}{\|P_N(x_{i-1} - x^*)\|} = \frac{p+1}{p+2}. \tag{6.3}$$

Proof. Here $b \geqq p$, $c \geqq p$ and hence $H(x) = \bar{D}_{p+1}(x)$ in (3.10). Since $H(x)$ is then invertible for all $P_N\tilde{x} \neq 0$, an application of Lemma 3.11 results in $q = 1$ as the solution of (3.12), and $F'(x)$ invertible in $W(\rho, \theta, 2)$ for ρ, θ sufficiently small. To compute $F'(x)^{-1}$ we note

$$\tilde{D}(x) = D_p(x) + D_{p+1}(x) - \beta_{p+2}(x) = D_{p+1}(x) + \theta\beta_{p+1}(x) \tag{6.4}$$

since $\bar{D}_p(x) \equiv 0$ and $x \in W(\rho, \theta, 2)$. Thus

$$\tilde{D}(x)^{-1} = D_{p+1}(x)^{-1} + \theta\beta_{-p+1}(x). \tag{6.5}$$

Now for this case (2.8) yields

$$\begin{aligned} F'(x)^{-1} = {} & P_X\hat{F}^{-1}P_X - P_X\hat{F}^{-1}B_p(x)\tilde{D}(x)^{-1}P_N \\ & - P_N\tilde{D}(x)^{-1}C_p(x)\hat{F}^{-1}P_X + P_N\tilde{D}(x)^{-1}P_N + \beta_1(x). \end{aligned} \tag{6.6}$$

From this we conclude

$$P_XF'(x)^{-1}P_X=\beta_0(x),\ P_NF'(x)^{-1}P_X=\beta_{-1}(x), \tag{6.7}$$

and hence

$$F'(x)^{-1}P_X=\gamma_{-1}^1(x). \tag{6.8}$$

In addition

$$P_NF'(x)^{-1}P_N=\beta_{-p-1}(x),\ P_XF'(x)^{-1}P_N=\beta_{-1}(x), \tag{6.9}$$

and hence

$$F'(x)^{-1}P_N=\gamma_{-p-1}^p(x). \tag{6.10}$$

Turning to Newton's method and denoting $x_0=x$, $x_1=y$, (4.8) yields

$$\begin{aligned}\tilde{y}=F'(x)^{-1}\Big\{&\frac{p}{p+1}[A_p(x)+B_p(x)+C_p(x)+D_p(x)]\tilde{x}\\&+\frac{p+1}{p+2}[A_{p+1}(x)+B_{p+1}(x)+C_{p+1}(x)+D_{p+1}(x)]\tilde{x}+\beta_{p+3}(x)\Big\}.\end{aligned} \tag{6.11}$$

Using (3.3), (6.8) and (6.10) we can rewrite this as

$$\begin{aligned}\tilde{y}=&\frac{p+1}{p+2}F'(x)^{-1}D_{p+1}(x)\tilde{x}+\gamma_{-1}^1(x)(A_p(x)+B_p(x)+A_{p+1}(x)+B_{p+1}(x))\tilde{x}\\&+\gamma_{-p-1}^p(x)(C_p(x)+C_{p+1}(x)+D_p(x))\tilde{x}+\gamma_{-p-1}^p(x)\beta_{p+3}(x).\end{aligned} \tag{6.12}$$

We shall now make use of the fact that for $x\in W(\rho,\theta,2)$:

$$\begin{aligned}&A_p(x)\tilde{x}=\theta\beta_{p+2}(x), &&C_p(x)\tilde{x}=\theta\beta_{p+2}(x),\\&A_{p+1}(x)\tilde{x}=\theta\beta_{p+3}(x), &&C_{p+1}(x)\tilde{x}=\theta\beta_{p+3}(x),\\&B_{p+1}(x)\tilde{x}=\beta_{p+2}(x), &&D_{p+1}(x)\tilde{x}=\beta_{p+2}(x),\\&&B_p(x)\tilde{x}=\beta_{p+1}(x),\end{aligned} \tag{6.13}$$

while

$$D_p(x)\tilde{x}=\theta\beta_{p+2}(x) \tag{6.14}$$

since $\bar{D}_p(x)\equiv 0$. The expressions (6.13)–(6.14) used in (6.12) result in

$$\tilde{y}=\frac{p+1}{p+2}F'(x)^{-1}D_{p+1}(x)\tilde{x}+\theta\gamma_1^p(x)+\gamma_p^1(x). \tag{6.15}$$

But from (6.5) and (6.9) we sharpen (6.10) to

$$P_NF'(x)^{-1}P_N=D_{p+1}(x)^{-1}+\theta\beta_{-p-1}(x), \tag{6.16}$$

$$P_XF'(x)^{-1}P_N=\beta_{-1}(x). \tag{6.17}$$

These two relations allow one to deduce from (6.15) that

$$P_N\tilde{y}=\frac{p+1}{p+2}P_N\tilde{x}+\theta\beta_1(x), \tag{6.18}$$

$$P_X\tilde{y}=\beta_{p+1}(x). \tag{6.19}$$

Equations (6.18)–(6.19) now, as in the proof of Theorem 5.1, yield (6.2)–(6.3) after noting that $p+1\geqq 3$ and hence the iterates remain in $W(\rho,\theta,2)$. □

We conclude this section with the statement of a result that proves convergence for a class of problems in which invertibility regions with $m > 2$ are allowed.

THEOREM 6.20. *Assume $p = d \geqq 2$ and $p < \bar{d} < \bar{b} + p$. Further assume $\bar{D}_{\bar{d}}(x)^{-1}$ exists as a map on N for all $\tilde{x}$ with $P_N\tilde{x} \neq 0$. Then for ρ, θ sufficiently small $F'(x_0)^{-1}$ exists for all $x_0 \in W(\rho, \theta, m)$ with $m = 1 + \bar{d} - d$, all subsequent Newton iterates remain in this set, and $x_i \to x^*$ with rate determined by*

$$\lim_{i \to \infty} \frac{\|P_N(x_i - x^*)\|}{\|P_N(x_{i-1} - x^*)\|} = \frac{\bar{d}}{\bar{d} + 1}, \tag{6.21}$$

$$\|P_X(x_i - x^*)\| \leqq K_1 \|x_{i-1} - x^*\|^{r+1}, \quad \textit{some } K_1 > 0, \tag{6.22}$$

where

$$r = \begin{cases} \min(\bar{b}, \bar{d} - d + a) & \textit{if } \bar{d} - d \geqq \bar{b} - b, \\ \bar{d} - d + \min(a, b) & \textit{if } \bar{b} - b \geqq \bar{d} - d. \end{cases} \tag{6.23}$$

The proof is similar to that of the previous result but considerably more complicated and is omitted in consideration of its length. □

We note that $r \geqq p$ and point out two special cases. The choice $\bar{b} = b = p$ and $\bar{d} = p + 1$ retrieves Theorem 6.1. The choice $\bar{b} = b = p$ but $2p > \bar{d} - d \geqq 2$ gives order $p + 1$ convergence in (6.22) and invariance of the set $W(\rho, \theta, 1 + \bar{d} - d)$.

7. Acceleration of convergence. It is the rapid quadratic convergence, in the nonsingular case, that makes Newton's method attractive for a broad class of problems. The major distinguishing feature of singular Newton's method problems is their overall linear convergence rate, and hence schemes that could return quadratic convergence for such problems are of strong interest. This question has a well known answer in the scalar case [13]. Suppose

$$f: \mathbb{R} \to \mathbb{R}, \qquad f(x^*) = 0 \tag{7.1}$$

and for some $p \geqq 1$

$$f^{(j+1)}(x^*) = 0, \quad j = 0, \cdots, p-1, \qquad f^{(p+1)}(x^*) \neq 0. \tag{7.2}$$

Then regular Newton's method converges linearly with rate $p/p+1$, while the scheme

$$x_{i+1} = x_i - (1+p) f'(x_i)^{-1} f(x_i), \qquad i = 0, 1, \cdots \tag{7.3}$$

exhibits quadratic convergence to x^*. The major difficulty in extending this result to a general Banach space is in guaranteeing the iterates remain in an appropriate region of invertibility of $F'(x)$. In this section we prove two convergence acceleration results, and as was the case in § 4, the results are different for $d \geqq 1$ and $d > 1$ with the latter being stronger.

THEOREM 7.4. *Assume that $d = \bar{d} \leqq c$, $b \geqq \min(2, d)$, and that for all $P_N\tilde{x} \neq 0$, $\bar{D}_d(x)$ is invertible as a map on N, and that for all $x \in W(\rho, \theta, 1)$*

$$\|\bar{D}_{d+1}(x) P_N \tilde{x}\| \geqq K_0 \|P_N \tilde{x}\|^{d+2}, \quad \textit{some } K_0 > 0. \tag{7.5}$$

Then for ρ and θ sufficiently small $F'(x_0)^{-1}$ exists for all $x_0 \in W(\rho, \theta, 1)$, and the iterates defined by

$$\begin{aligned} y_i &= x_i - F'(x_i)^{-1} F(x_i), \\ z_i &= y_i - F'(y_i)^{-1} F(y_i), \qquad i = 0, 1, \cdots, \\ x_{i+1} &= z_i - (d+1) F'(z_i)^{-1} F(z_i), \end{aligned} \tag{7.6}$$

all remain in $W(\rho, \theta, 1)$, *and* $x_i \to x^*$ *with rate determined by*

$$\|x_i - x^*\| \leqq K_1 \|x_{i-1} - x^*\|^2, \quad \text{some } K_1 > 0, \quad i = 1, 2, \cdots. \tag{7.7}$$

Proof. Here $d = \bar{d} < b + c$ and hence Corollary 3.15 guarantees $F'(x)$ invertible for ρ and θ sufficiently small and $x \in W(\rho, \theta, 1)$. In this region

$$\tilde{D}(x) = D_d(x) + D_{d+1}(x) - C_d(x)\hat{F}^{-1}B_b(x) + \beta_{d+2}(x) \tag{7.8}$$

and $D_d(x)$ is invertible so:

$$\tilde{D}(x) = D_d(x)[P_N + D_d(x)^{-1}D_{d+1}(x) - D_d(x)^{-1}C_d(x)\hat{F}^{-1}B_b(x) + \beta_2(x)]. \tag{7.9}$$

One may directly verify by the Banach Lemma that

$$\begin{aligned}&(P_N + D_d(x)^{-1}D_{d+1}(x) - D_d(x)^{-1}C_d(x)\hat{F}^{-1}B_b(x) + \beta_2(x))^{-1}\\ &\quad = P_N - D_d(x)^{-1}D_{d+1}(x) + D_d(x)^{-1}C_d(x)\hat{F}^{-1}B_b(x) + \beta_2(x),\end{aligned} \tag{7.10}$$

and hence

$$\begin{aligned}\tilde{D}(x)^{-1} &= D_d(x)^{-1} - D_d(x)^{-1}D_{d+1}(x)D_d(x)^{-1}\\ &\quad + D_d(x)^{-1}C_d(x)\hat{F}^{-1}B_b(x)D_d(x)^{-1} + \beta_{-d+2}(x).\end{aligned} \tag{7.11}$$

In this case the expression for $F'(x)^{-1}$ given by (2.8) becomes

$$\begin{aligned}F'(x)^{-1} &= P_X\hat{F}^{-1}P_X - P_X\hat{F}^{-1}B_b(x)D_d(x)^{-1}P_N - P_ND_d(x)^{-1}C_d(x)\hat{F}^{-1}P_X\\ &\quad + P_N(D_d^{-1}(x) - D_d(x)^{-1}(D_{d+1}(x) - C_d(x)\hat{F}^{-1}B_b(x))D_d(x)^{-1})\\ &\quad + P_N\beta_{-d+2}(x)P_N + P_X\beta_{-d+b+1}(x)P_N + \beta_1(x).\end{aligned} \tag{7.12}$$

For $x_0 \in W(\rho, \theta, 1)$ with ρ and θ sufficiently small, $y_0 \in W(\rho, \theta, 1)$ and hence we may construct z_0 in (7.6). We now proceed to show, for some $c_1 > 0$ that

$$\|P_X\tilde{z}_0\| \leqq c_1\|P_N\tilde{y}_0\| \cdot \|P_X\tilde{y}_0\| + \beta_3(y_0). \tag{7.13}$$

This inequality, coupled with $P_N\tilde{y}_0 = \beta_1(x_0)$ and $P_X\tilde{y}_0 = \beta_2(x_0)$ (from Theorem 5.1), will allow the crucial estimate $P_X\tilde{z}_0 = \beta_3(x_0)$. First, from (4.8) we have, as $F'(y_0)^{-1}P_X = \beta_0(y_0)$, $P_XF'(y_0)^{-1}P_N = \beta_{-d+b}(y_0)$,

$$\tilde{z}_0 = F'(y_0)^{-1}\left\{\frac{1}{2}A_1(y_0) + \frac{b}{b+1}B_b(y_0) + \frac{d}{d+1}(C_d(y_0) + D_d(y_0))\right\}\tilde{y}_0 + \gamma_2^1(y_0). \tag{7.14}$$

Hence

$$\begin{aligned}P_X\tilde{z}_0 &= \{P_X\hat{F}^{-1}P_X - P_X\hat{F}^{-1}B_b(y_0)D_d(y_0)^{-1}P_N + P_X\beta_{-d+b+1}(y_0)P_N\}\\ &\quad \times\left\{\frac{1}{2}A_1(y_0) + \frac{b}{b+1}B_b(y_0) + \frac{d}{d+1}(C_d(y_0) + D_d(y_0))\right\}\tilde{y}_0 + \beta_3(y_0).\end{aligned} \tag{7.15}$$

Therefore

$$\begin{aligned}P_X\tilde{z}_0 &= \frac{1}{2}P_X\hat{F}^{-1}P_XA_1(y_0)\tilde{y}_0 + \left(\frac{b}{b+1} - \frac{d}{d+1}\right)P_X\hat{F}^{-1}P_XB_b(y_0)\tilde{y}_0\\ &\quad - \frac{d}{d+1}P_X\hat{F}^{-1}B_b(y_0)D_d^{-1}(y_0)C_d(y_0)\tilde{y}_0 + \beta_3(y_0).\end{aligned} \tag{7.16}$$

Now the second term on the right-hand side of (7.16) vanishes if $b = 1$ and is $\beta_3(y_0)$ otherwise, and hence recalling the definitions (2.3) we may conclude (7.13). From

Theorem 5.1, $P_N\tilde{z}_0=\beta_1(x_0)$, and hence $z_0\in W(\rho,\theta,1)$ and it is possible to compute x_1. Now defining

$$w_0=z_0-F'(z_0)^{-1}F(z_0), \tag{7.17}$$

an argument identical to that providing (7.13) results in

$$\|P_X\tilde{w}_0\|\leqq c_0\|P_N\tilde{z}_0\|\,\|P_X\tilde{z}_0\|+\beta_3(z_0) \tag{7.18}$$

and hence $P_X\tilde{w}_0=\beta_3(x_0)$. But from (7.6), $\tilde{x}_1=(d+1)\tilde{w}_0-d\tilde{z}_0$ thus

$$P_X\tilde{x}_1=(d+1)P_X\tilde{w}_0-dP_X\tilde{z}_0=\beta_3(x_0). \tag{7.19}$$

All that remains in the proof is to show, for some $c_1, c_2>0$ that

$$c_1\|P_N\tilde{x}_0\|^2\leqq\|P_N\tilde{x}_1\|\leqq c_2\|P_Nx_0\|^2 \tag{7.20}$$

since estimates of the form (7.19)–(7.20) guarantee $x_1\in W(\rho,\theta,1)$ and provide (7.7). To show (7.20), we see from above that $P_N\tilde{x}_1=(d+1)P_N\tilde{w}_0-dP_N\tilde{z}_0$, and that (4.8) in this case becomes

$$\begin{aligned}P_N\tilde{x}_1=(d+1)P_NF'(z_0)^{-1}\Big\{\Big[\frac{1}{2}A_1(z_0)+\frac{b}{b+1}B_b(z_0)+\sum_{j=d}^{d+1}\frac{j}{j+1}(C_j(z_0)+D_j(z_0))\Big]\tilde{z}_0\\ +P_X\beta_3(z_0)+\beta_{d+3}(z_0)\Big\}-dP_N\tilde{z}_0.\end{aligned} \tag{7.21}$$

To estimate the various terms in (7.21) we note from (7.12) that

$$P_NF'(z_0)^{-1}P_N=\beta_{-d}(z_0),\ P_NF'(z_0)^{-1}P_X=\beta_0(z_0) \tag{7.22}$$

and further that

$$\begin{aligned}A_1(z_0)\tilde{z}_0&=A_1(z_0)P_X\tilde{z}_0=\beta_4(x_0),\\ C_d(z_0)\tilde{z}_0&=C_d(z_0)P_X\tilde{z}_0=\beta_{d+3}(x_0).\end{aligned} \tag{7.23}$$

Thus

$$\begin{aligned}P_N\tilde{x}_1=(d+1)\{D_d(z_0)^{-1}-D_d(z_0)^{-1}(D_{d+1}(z_0)-C_d(z_0)\hat{F}^{-1}B_b(z_0))D_d(z_0)^{-1}\\ -P_ND_d(z_0)^{-1}C_d(z_0)\hat{F}^{-1}P_X\}\\ \times\Big\{\Big[\frac{b}{b+1}B_b(z_0)+\frac{d}{d+1}D_d(z_0)+\frac{d+1}{d+2}D_{d+1}(z_0)\Big]\tilde{z}_0\Big\}-dP_N\tilde{z}_0+\beta_3(x_0).\end{aligned} \tag{7.24}$$

Simplification of (7.24) results in

$$\begin{aligned}P_N\tilde{x}_1&=\frac{1}{d+2}D_d(z_0)^{-1}D_{d+1}(z_0)\tilde{z}_0\\ &+\Big(\frac{d}{d+1}-\frac{b}{b+1}\Big)D_d^{-1}(z_0)C_d(z_0)\hat{F}^{-1}B_b(z_0)\tilde{z}_0+\beta_3(x_0).\end{aligned} \tag{7.25}$$

Once again the second term on the right-hand side of (7.25) vanishes if $b=1$ and is $\beta_3(z_0)=\beta_3(x_0)$ otherwise. The assumption (7.5) on $\bar{D}_{d+1}(x)$ now provides the estimate (7.20) and concludes the proof. □

In [7] Griewank proves quadratic convergence for the scheme described by (7.6) for $d=\bar{d}=1$. His analysis restricted to $\mathbb{R}^n$ and the method of proof involves expansion of the determinant function.

If $b\geqq 2$ and $c\geqq d\geqq 2$, minor modifications of the above analysis will show that an acceleration step may be performed after calculation of y_i. The result is

THEOREM 7.27. *Assume $b \geqq 2$, $c \geqq d = \bar{d} \geqq 2$ and (7.5) holds. Then for ρ and θ sufficiently small the iterates given by*

$$(7.28)\qquad \begin{aligned} y_i &= x_i - F'(x_i)^{-1}F(x_i), \\ x_{i+1} &= y_i - (d+1)F'(y_i)^{-1}F(y_i) \end{aligned}$$

remain in $W(\rho, \theta, 1)$, and $x_i \to x^*$ with rate given by

$$(7.30)\qquad \begin{aligned} \|P_N(x_i - x^*)\| &\leqq K_1\|x_{i-1} - x^*\|^2, \\ \|P_X(x_i - x^*)\| &\leqq K_2\|x_{i-1} - x^*\|^{p+1}, \end{aligned} \qquad i = 1, 2, \cdots,$$

for some $K_1, K_2 > 0$.

8. Examples. In this section we present examples of the use of Newton's method that illustrate the results of the previous sections. These examples were designed to be as simple as possible and yet demonstrate the convergence rate conclusions of the appropriate theorem. They are all algebraic equations and, except for the last, are two-dimensional problems.

For Examples 1–6 we use:

$$(8.1)\qquad \mathbf{F}: \mathbb{R}^2 \to \mathbb{R}^2, \qquad \mathbf{F}(\mathbf{x}) \equiv \begin{pmatrix} f_1(x, y) \\ f_2(x, y) \end{pmatrix} \equiv \begin{pmatrix} x + h_1(x, y) \\ h_2(x, y) \end{pmatrix}.$$

We shall require that:

$$(8.2)\qquad \begin{aligned} &\mathbf{x}^* = (0, 0)^T, \\ &\mathcal{N}(F'(x^*)) = \operatorname{span}(\phi), \qquad \phi = (0, 1)^T, \\ &\mathcal{R}(F'(\mathbf{x}^*)) = \operatorname{span}(\psi), \qquad \psi = (1, 0)^T. \end{aligned}$$

From this we must have:

$$(8.3)\qquad P_N = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}, \qquad P_X = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

and we need only choose $h_i(x, y)$, $i = 1, 2$ to be at least quadratic. With this structure invertibility requires *initial guesses near the y-axis.*

Example 1 (Theorem 5.1). Here we require $P_N F''(\mathbf{x}^*)(\phi, \phi) \neq 0$, which is satisfied provided $\partial^2 f_2/\partial y^2 \neq 0$ at $(0, 0)$. We take

$$(8.4)\qquad \mathbf{F}(x, y) = \begin{pmatrix} x + y^2 \\ \frac{3}{2}xy + y^2 + y^3 \end{pmatrix}$$

with the initial guess $x_0 = .1$, $y_0 = 1$. Theorem 5.1 implies

$$(8.5)\qquad \lim_{i \to \infty} R_i = \frac{y_{i+1}}{y_i} = \frac{1}{2}.$$

The Newton iterates for this example are given in Table 1.

Example 2 (Theorem 5.9). We take $p = d = \bar{d} = 2$ and to verify $P_N F'''(\mathbf{x}^*)(\phi, \phi, \phi) \neq 0$ we need only check $\partial^3 f_2/\partial y^3 \neq 0$ at $\mathbf{x}^*$. The mapping chosen is

$$(8.6)\qquad \mathbf{F}(x, y) = \begin{pmatrix} x + y^3 \\ xy^2 + y^3 + y^4 \end{pmatrix}.$$

The initial guess is $x_0 = .05$, $y_0 = .5$ and here $R_i \to \frac{2}{3}$ by Theorem 5.9. The results are

TABLE 1.

i	x_i	y_i	R_i
1	−0.5349D 00	0.7674D 00	0.7674D 00
2	0.1715D 00	0.2720D 00	0.3544D 00
3	−0.3058D−01	0.1922D 00	0.7067D 00
4	0.9743D−03	0.9357D−01	0.4868D 00
5	−0.4742D−03	0.4932D−01	0.5271D 00
10	−0.1609D−07	0.1603D−02	0.5008D 00
20	−0.1541D−16	0.1568D−05	0.5000D 00

given in Table 2. We next consider the acceleration schemes for the two previous examples.

TABLE 2.

i	x_i	y_i	R_i
1	−0.1966D−01	0.3596D 00	0.7191D 00
2	−0.2951D−02	0.2473D 00	0.6878D 00
3	−0.8937D−03	0.1697D 00	0.6864D 00
4	−0.2174D−03	0.1157D 00	0.6815D 00
5	−0.5067D−04	0.7838D−01	0.6776D 00
10	−0.2107D−07	0.1066D−01	0.6684D 00
20	−0.2008D−14	0.1857D−03	0.6667D 00

Example 3 (Theorem 7.4). The mapping is that of (8.4) and we choose the same initial guess $\mathbf{x}_0=(.1, 1)^T$. We list in Table 3 the quadratically convergent iterates $\mathbf{x}_i=(x_i, y_i)$ obtained by using (7.6). The intermediate quantities $\mathbf{y}_i$ and $\mathbf{z}_i$ are not given here.

TABLE 3.

i	x_i	y_i	$\|x_i\|+\|y_i\|$
1	−0.2327D 00	0.1124D 00	0.3451D 00
2	−0.2136D−02	0.5868D−02	0.8004D−02
3	0.1995D−07	0.4147D−06	0.4347D−06
4	0.2742D−20	0.5756D−14	0.5756D−14
5	0.8933D−44	0.1430D−29	0.1430D−29

Example 4 (Theorem 7.26). The mapping (8.6) of Example 2 is employed and we note $P_N F''''(\mathbf{x}^*)(\phi, \phi, \phi, \phi)\neq 0$ as $\partial^4 f_2/\partial y^4\neq 0$ at $\mathbf{x}^*$. The scheme (7.28) is used and the results $\mathbf{x}_i=(x_i, y_i)$ are given in Table 4. Note the rapid convergence of the

TABLE 4.

i	x_i	y_i	$\|x_i\|+\|y_i\|$
1	0.5670D 00	0.2551D−02	0.5695D 00
2	0.3292D−07	0.2141D−05	0.2174D−05
3	0.1991D−18	0.6857D−12	0.6858D−12
4	−0.4372D−49	0.6979D−25	0.6979D−25

x_i-component; indeed it even seems better than the cubic bound indicated in (7.29). (The initial guess was chosen $(x_0, y_0) = (.1, 1)$.) For higher order regions of invertibility we have Examples 5 and 6.

Example 5 (Theorem 6.1). The mapping chosen is

$$\mathbf{F}(x, y) = \begin{pmatrix} x+y^3 \\ x^2y+y^4 \end{pmatrix}. \tag{8.7}$$

This has $p = d = 2$ as $P_N F'''(\mathbf{x}^*)(\psi, \psi, \phi) \neq 0$ but $\bar{d} = 3$ since $P_N F'''(\mathbf{x}^*)(\phi, \phi, \phi) = 0$ while $P_N F''''(\mathbf{x}^*)(\phi, \phi, \phi, \phi) \neq 0$. From Theorem 6.1 the ratio of null space errors will approach $\frac{3}{4}$ and this behavior is present in Table 5. In addition, the cubic bound (6.2) for the X component error is seen to be sharp for this example. The initial iterate was $(x_0, y_0) = (.1, 1)$ and $\|\mathbf{x}\| \equiv |x| + |y|$.

TABLE 5.

i	x_i	y_i	R_i	$\lvert x_i\rvert/\lVert x_{i-1}\rVert^3$
1	−0.3050D 00	0.7683D 00	0.7683D 00	0.2291D 00
2	−0.1368D 00	0.5895D 00	0.7672D 00	0.1106D 00
3	−0.5694D−01	0.4476D 00	0.7593D 00	0.1486D 00
4	−0.2345D−01	0.3374D 00	0.7538D 00	0.1826D 00
5	−0.9774D−02	0.2536D 00	0.7515D 00	0.2080D 00
10	−0.1296D−03	0.6025D−01	0.7500D 00	0.2472D 00
20	−0.2315D−07	0.3393D−02	0.7500D 00	0.2500D 00
30	−0.4134D−11	0.1911D−03	0.7500D 00	0.2500D 00

Example 6 (Theorem 6.20). As a particular case of this result we pick $p = d = 2$, while $\bar{d} = 4$ and $\bar{b} = 3$. This is satisfied by the choice

$$\mathbf{F}(x, y) = \begin{pmatrix} x+y^4 \\ x^2y+y^5 \end{pmatrix}. \tag{8.8}$$

The initial guess was chosen to be $(x_0, y_0) = (.01, 1)$. In this case Theorem 6.20

TABLE 6.

i	x_i	y_i	R_i	$\lvert x_i\rvert/\lVert x_{i-1}\rVert^4$
1	−0.2033D 00	0.8008D 00	0.8008D 00	0.1954D 00
2	−0.9187D−01	0.6453D 00	0.8058D 00	0.9035D−01
3	−0.3729D−01	0.5187D 00	0.8037D 00	0.1262D 00
4	−0.1492D−01	0.4157D 00	0.8015D 00	0.1561D 00
5	−0.6045D−02	0.3328D 00	0.8006D 00	0.1757D 00
10	−0.6923D−04	0.1091D 00	0.8000D 00	0.1990D 00
20	−0.9201D−08	0.1172D−01	0.8000D 00	0.2000D 00
30	−0.1223D−11	0.1258D−02	0.8000D 00	0.2000D 00

concludes $R_i \to \frac{4}{5}$. For this example $r = \bar{b} = 3$ and Table 6 indicates a quartic bound given by (6.22).

Acknowledgments. The authors would like to thank Professor A. O. Griewank of Southern Methodist University and the referee for some very useful comments on the original version of this paper.

BIBLIOGRAPHY

[1] R. C. CAVANAUGH, *Difference equations and iterative processes*, Thesis, Computer Sci. Dept., University of Maryland, College Park, 1970.

[2] D. W. DECKER AND C. T. KELLEY, *Newton's method at singular points* I, this Journal, 17 (1980), pp. 66–70.

[3] ———, *Newton's method at singular points* II, this Journal, 17, (1980), pp. 465–471.

[4] ———, *Convergence acceleration for Newton's method at singular points*, this Journal, 19 (1982), pp. 219–229.

[5] D. W. DECKER AND H. B. KELLER, *Path following near bifurcation*, Comm. Pure. Appl. Math., 34 (1981), pp. 149–175.

[6] A. O. GRIEWANK, *Analysis and modification of Newton's method at singularities*, thesis, Australian National University, 1980.

[7] ———, *Starlike domains of convergence for Newton's method at singularities*, Numer. Math., 35 (1980), pp. 95–111.

[8] A. GRIEWANK AND M. R. OSBORNE, *Newton's method for singular problems when the dimension of the null space is* >1, this Journal, 18 (1981), pp. 145–149.

[9] L. V. KANTOROVICH AND G. P. AKILOV, *Functional Analysis in Normed Spaces*, Pergamon, New York, 1964.

[10] H. B. KELLER, *Geometrically isolated nonisolated solutions and their approximation*, this Journal, 18 (1981) pp. 822–838.

[11] ———, *Numerical solution of bifurcation and nonlinear eigenvalue problems*, Applications of Bifurcation Theory, P. H. Rabinowitz, ed., Academic Press, New York, 1977, pp. 359–384.

[12] ———, *Newton's method under mild differentiability conditions*, J. Comp. Sys. Sci., 4 (1970), pp. 15–28.

[13] L. B. RALL, *Convergence of the Newton process to multiple solutions*, Numer. Math., 9 (1966), pp. 23–37.

[14] G. W. REDDIEN, *On Newton's method for singular problems*, this Journal, 15 (1978), pp. 993–996.

[15] ———, *Newton's method and high order singularities*, Comput. Math. Appl., 5 (1980), pp. 79–86.

[16] R. SEYDEL, *Numerical computation of branch points in ordinary differential equations*, Numer. Math., 32 (1979), pp. 51–68.

[17] R. B. SIMPSON, *A method for the numerical determination of bifurcation states of nonlinear systems of equations*, this Journal 12 (1975), pp. 439–451.