SIAM J. NUMER. ANAL.
Vol. 22, No. 4, August 1985

009

LEAST-CHANGE SPARSE SECANT UPDATE METHODS WITH INACCURATE SECANT CONDITIONS*

J. E. DENNIS, JR.† AND HOMER F. WALKER‡

Abstract. We investigate the role of the secant or quasi-Newton condition in the sparse Broyden or Schubert update method for solving systems of nonlinear equations whose Jacobians are either sparse, or can be approximated acceptably by conveniently sparse matrices. We develop a theory on perturbations to the secant equation that will still allow a proof of local q-linear convergence. To illustrate the theory, we show how to generalize the standard secant condition to the case when the function difference is contaminated by noise.

Key Words. quasi-Newton methods, local convergence, sparse nonlinear equations, bounded deterioration, least-change secant methods, Schubert's method, Broyden's method

1. Introduction. In earlier work (Dennis and Walker (1983)), we addressed the effects of inaccuracy on the performance of quasi-Newton methods for solving nonlinear algebraic equations. There, the emphasis was on generality: The methods considered were general bounded-deterioration quasi-Newton methods, and we took into account inaccuracy from all sources, including not only finite-precision computer arithmetic but also the differences between an ideal problem, its mathematical model, and the computer implementation of the model. The objective was to determine rates of improvement and limiting accuracies that can be obtained near solutions in the presence of such general inaccuracy.

Here, our interest is again in inaccuracy in quasi-Newton methods; however, the focus is more specific than before. The methods considered are least-change secant update methods (see Dennis and Schnabel (1979)), in particular, methods employing the sparse secant updates of Broyden (1971) and Schubert (1970), which include of course the usual Broyden (1965) update. The inaccuracy with which we are concerned is that which residues in the secant conditions by which updates are determined, and we explicitly exclude inaccuracy arising from other sources from consideration here. Our objective is a local convergence analysis which extends that of Dennis and Walker (1981) in the sparse secant update case by relaxing the requirements on secant conditions in the theorems of that paper. The extension we obtain gives an interesting "box" of secant conditions that all lead to locally q-linearly convergent methods. Our analysis cannot distinguish any difference between the radii of local convergence of these methods. However, it does indicate differences in convergence speed, and we will argue for a generalization of the usual secant condition on that basis.

The problem we study is the following:

Problem. Given $F:\Omega \subseteq \mathbf{R}^n \to \mathbf{R}^n$, find $x_* \in \Omega$ satisfying

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

* Received by the editors December 13, 1983, and in revised form May 29, 1984.

† Mathematical Sciences Department, Rice University, Houston, Texas 77251. The research of this author was sponsored by the Department of Energy under grant DE-AS05-82ER13016, the U.S. Army Research Office under grant DAAG-79-C-0124, and the National Science Foundation under grant MCS81-16779. This work was supported in part by the International Business Machine Corporation, Palo Alto Scientific Center, Palo Alto, CA.

‡ Department of Mathematics, University of Houston, Houston, Texas 77004. The research of this author was sponsored by the Department of Energy under grant DE-AS05-82ER13016.

By a quasi-Newton method for approximating a solution of (1.1), we mean any iterative method of the general form

$$x_{k+1} = x_k - B_k^{-1} F(x_k), \tag{1.2}$$

in which B_k is regarded as an approximation of the Jacobian matrix $F'(x_k)$. Since our concern here is with local behavior, we consider only iterations in which the full step $-B_k^{-1}F(x_k)$ is appropriate. Each B_k is assumed to have the form

$$B_k = A_k + C(x_k), \tag{1.3}$$

in which $C(x_k)$ is a "computed part" of $F'(x_k)$ determined by a function $C:\Omega \subseteq \mathbf{R}^n \to \mathbf{R}^{n\times n}$ and A_k is an "approximated part" of $F'(x_k)$ maintained by updating.

Since the methods with which we are concerned here are modeled after Newton's method, we assume the following throughout the sequel:

The standard hypothesis on F and C. Let F be differentiable in an open convex neighborhood Ω of a point $x_* \in \mathbf{R}^n$ for which $F(x_*) = 0$ and $F'(x_*)$ is nonsingular. Let $\gamma \geqq 0$, $\gamma_C \geqq 0$, and $p \in (0, 1]$ be such that for $x \in \Omega$,

$$|F'(x) - F'(x_*)| \leqq \gamma |x - x_*|^p$$

and

$$|C(x) - C(x_*)| \leqq \gamma_C |x - x_*|^p,$$

where $|\cdot|$ denotes a norm on $\mathbf{R}^n$ and its subordinate operator norm on $\mathbf{R}^{n\times n.}$

The part of this standard hypothesis which applies to F is sufficient to insure that sequences of iterates produced by Newton's method converge locally to x_*, with q-order $(1+p)$. See, for example, Dennis and Schnabel (1983) or Ortega and Rheinboldt (1970).

A very successful way of maintaining the A_k's in (1.3) is through least-change secant updates (see Dennis and Schnabel (1979) and Dennis and Walker (1981)). An extensive review of these updates is not in order here; but let us recall that if one is given an inner-product norm $\|\cdot\|$ on $\mathbf{R}^{n\times n}$ and an affine subspace $\mathbf{A} \subseteq \mathbf{R}^{n\times n}$ which reflects the structure of $[F'(x) - C(x)]$, then A_{k+1} is uniquely determined as a least-change secant update of A_k in $\mathbf{A}$ by a choice of vectors $s_k \neq 0$ and y_k in $\mathbf{R}^n$. Furthermore, one has that

$$A_{k+1} s_k = y_k, \tag{1.4}$$

if it is at all possible for this equation to be satisfied by a matrix in A. The purpose of updating A_k is to incorporate in A_{k+1} currently available information about $[F'(x_{k+1}) - C(x_{k+1})]$, and so one usually chooses $s_k = x_{k+1} - x_k$ and $y_k \approx [F'(x_{k+1}) - C(x_{k+1})]s_k$. In fact, in most classical applications, $C(x) \equiv 0$ and one takes $s_k = x_{k+1} - x_k$ and $y_k = F(x_{k+1}) - F(x_k)$. In this case, (1.4) is satisfied if $F'(x) \in \mathbf{A}$ for all $x \in \mathbf{R}^n$.

In view of the usual choice of s_k and y_k, we call (1.4) a *secant condition* on A_{k+1}. In the literature it is also at times called the *quasi-Newton equation.* From here on, we assume that $s_k = x_{k+1} - x_k$; and so if $\mathbf{A}$ and a suitable norm on $\mathbf{R}^{n\times n}$ are given, then a choice of y_k alone determines A_{k+1} as a least-change secant update of A_k in $\mathbf{A}$. We shall regard choosing y_k as equivalent to specifying a secant condition (1.4) on A_{k+1}, even though it may not always be possible for (1.4) to be satisfied by a matrix in $\mathbf{A}$. For convenience, we shall also occasionally refer to y_k itself as a secant condition.

In Dennis and Walker (1981), we established very general requirements on $\mathbf{A}$ and the secant conditions (1.4) which guarantee that the resulting least-change secant updates lead to a quasi-Newton method with local q-linear asymptotic convergence which is optimal for a sequence $\{A_k\}$ in $\mathbf{A}$. In classical cases in which $[F'(x_*)-C(x_*)]\in\mathbf{A}$, these requirements give the standard local and q-superlinear convergence results. A major feature of the theorems of Dennis and Walker (1981) is that they allow one to relax the condition that $[F'(x_*)-C(x_*)]\in\mathbf{A}$. Our requirements are phrased in terms of a choice rule for determining allowable y_k's for each k. Such a choice rule is given by a function $\chi:\mathbf{R}^n\times\mathbf{R}^n\rightarrow 2^{\mathbf{R}^n}$, for each k, one chooses any $y_k\in\chi(x_k, x_{k+1})$, provided x_k and x_{k+1} are in Ω. The utility of prescribing a choice rule in this way is evident when one considers applications in which a multiplicity of workable y_k's naturally present themselves (see the discussion in Dennis and Walker (1981, p. 960)). In classical cases, the manner of determining an appropriate choice rule is clear, and it is very easy to verify that our requirements are satisfied, at least ideally.

In this report, our concern is with secant conditions which are *inaccurate* in the sense that they cannot be associated with a choice rule that satisfies the requirements of Dennis and Walker (1981). In the following, we offer an extended local convergence analysis by relaxing the requirements that secant conditions must satisfy in order to lead to least-change secant update methods which have desirable local convergence properties. Our analysis is restricted to the case in which the updates are sparse secant updates; the investigation of other cases if left to future work. We are motivated by the fact that in spite of the generality of the published requirements, there are circumstances in which secant conditions satisfying them must be regarded as unobservable in practice. We cite two such circumstances below which we feel are particularly important.

The first circumstance is that in which the computed values of F or C contain inaccuracy. In this case, any choice of y_k which is determined by computed function values will also exhibit inaccuracy, and so it seems essentially certain that any practical implementation of a choice rule which specifies y_k in this way will fail to satisfy the conditions of Dennis and Walker (1981), even though the "ideal" choice rule which uses accurate function values may be completely satisfactory. This case is important in practice; see Barrera and Dennis (1979) and Moré, Garbow and Hillstrom (1980). The analysis given here provides reassurance that in most classical applications, the traditional secant conditions should usually be usable in practice, even though they are determined by inaccurate computed function values. This reassurance is in the form of a simple generalization $\bar{y}_k$ of the traditional secant condition y_k.

The second circumstance which we cite is that in which it might be advantageous to impose some sort of special structure on each A_k, even though this structure is not fully reflected in F' and C. This is to say that it might be desirable to require that each A_k lie in some $\mathbf{A}$, despite the fact that $[F'(x)-C(x)]$ fails to belong to $\mathbf{A}$ for some $x\in\Omega$. For example, $\mathbf{A}$ might be a subspace of $\mathbf{R}^{n\times n}$ consisting of matrices having a particularly appealing pattern of sparsity, and it might happen that for each $x\in\Omega$, the part of $[F'(x)-C(x)]$ which lies outside of this pattern of sparsity is so small that one is tempted to discard it and require that $A_k\in\mathbf{A}$ for each k. In this case, one should choose $y_k\approx P_{\mathbf{A}}[F'(x_{k+1})-C(x_{k+1})]s_k$, where $P_{\mathbf{A}}$ denotes the orthogonal projection onto $\mathbf{A}$, and even though there may exist a choice rule for y_k satisfying the conditions of Dennis and Walker (1981) (p. 964, formula (3.13)), it may be impossible or impractical to implement it. The analysis given here should help to determine whether natural and readily available secant conditions can be safely used for updating in this case. If these secant conditions are not suitable, then it should suggest modifications of them which are safe to use and can be easily obtained. We are preparing some interesting applications of this analysis.

For prespective, let us recall that the local convergence analysis of least-change secant update methods in Dennis and Walker (1981) proceeds in the traditional two-stage form. In the first stage, one shows that the updates of interest exhibit a phenomenon known as bounded deterioration; it follows that methods employing these updates enjoy local q-linear convergence. In the second stage, one looks more closely at q-linearly convergent sequences of iterates produced by these methods and shows that this convergence is optimal in that it is asymptotically as fast as that of the ideal stationary iteration of the form (1.2) in which $B_k = C(x_*) + P_{\mathbf{A}}[F'(x_*) - C(x_*)]$ for each k. The properties of the choice rule for determining secant conditions are fundamental to the analysis at each stage.

The local convergence analysis given in the following is in a similar two-stage form. The first stage is developed in § 2. There it is assumed that a choice rule satisfying the conditions of Dennis and Walker (1981) is given, and it is first shown that one can make certain generous enlargements of sets of secant conditions specified by this choice rule and still retain bounded deterioration and the associated local q-linear convergence. Then this initial result is used to obtain a corollary in a form more useful in the applications toward which this work is directed. The second stage is developed in § 3, in which it is shown that q-linearly convergent iteration sequences exhibit asymptotic speeds of convergence which can be regarded as optimal when one considers the inaccuracy in secant conditions used to determine updates. Indeed, the results of § 3 make visible the extent to which inaccuracy in secant conditions degrades the asymptotic speed of convergence of an iteration sequence and show that this asymptotic speed can be made arbitrarily close to (or even equal to) that of the ideal stationary iteration provided it is feasible to exercise sufficient control over inaccuracy in secant conditions. In § 4, we conclude with an illustration of how one might make use of the results of §§ 2 and 3.

2. Secant conditions and local linear convergence. For the remainder of this report, we consider methods of the form (1.2) for solving (1.1) in which the A_k's in (1.3) are maintained through secant updates into a sparse matrix subspace $\mathbf{Z} \subseteq \mathbf{R}^{n \times n}$. In this section, our interest is in secant conditions which determine these updates and in the q-linear convergence properties of the methods which result.

We begin by establishing some notation and reviewing the basic properties of the sparse second update. Throughout the sequel, $|\cdot|$ denotes both the $\mathbf{l}_2$ norm on $\mathbf{R}^n$ and the induced operator norm on $\mathbf{R}^{n \times n}$, and $\|\cdot\|$ denotes the Frobenius norm on $\mathbf{R}^{n \times n}$. If $\mathbf{A}$ is a subspace or affine subspace of $\mathbf{R}^{n \times n}$, then $P_{\mathbf{A}}$ denotes the projection onto $\mathbf{A}$ which is orthogonal with respect to the Frobenius norm and $P_{\mathbf{A}}^{\perp} = I - P_{\mathbf{A}}$ denotes the projection orthogonal to $P_{\mathbf{A}}$. Members of the standard unit basis of $\mathbf{R}^n$ are denoted by $\varepsilon_1, \cdots, \varepsilon_n$, i.e., for $i = 1, \cdots, n$, $\varepsilon_i = (\delta_{1i}, \delta_{2i}, \cdots, \delta_{ni})^T$, where $\delta_{ii} = 1$ and $\delta_{ji} = 0$ for $j \neq i$. For $i = 1, \cdots, n$, we use P_i to denote the $\mathbf{l}_2$ projection that sends a row (column) vector into the nearest row (column) vector having the sparsity of the ith row of an element of $\mathbf{Z}$, i.e., $P_i v = \varepsilon_i^T P_{\mathbf{Z}}(\varepsilon_i v)$ for a row vector v.

With this notation, we can write down the formula for a sparse secant update as follows: If $A \in \mathbf{Z}$ and if $s, y \in \mathbf{R}^n$ with $s \neq 0$, then the sparse secant update A_+ of A in $\mathbf{Z}$ with respect to s and y is given by

$$A_+ = A + \sum_{i=1}^{n} (s^T P_i s)^+ \varepsilon_i^T (y - As) \varepsilon_i P_i s^T, \tag{2.1}$$

where for any real c,

$$c^+ = \begin{cases} c^{-1} & \text{if } c \neq 0, \\ 0 & \text{if } c = 0. \end{cases}$$

If A is not in $\mathbf{Z}$, then it is only necessary to alter (2.1) by replacing A with $P_{\mathbf{Z}}A$.

Reid (1973), Marwil (1979), and Dennis and Schnabel (1979) have shown that A_+ given by (2.1) is the least-change secant update of A in $\mathbf{Z}$ with respect to s, y, and the Frobenius norm $\|\cdot\|$. This is to say that A_+ uniquely solves

$$\min_{\bar{A}\in\mathbf{M}(\mathbf{Z},\mathbf{Q}(y,s))} \|\bar{A}-A\|,$$

where $\mathbf{Q}(y,s)=\{M\in\mathbf{R}^{n\times n}: Ms=y\}$ is the affine subspace of generalized quotients of y and s and $\mathbf{M}(\mathbf{Z},\mathbf{Q}(y,s))$ is the affine subspace of elements of $\mathbf{Z}$ for which the distance to $\mathbf{Q}(y,s)$ in the norm $\|\cdot\|$ is minimal. Note that one has $A_+=P_{\mathbf{M}}A$, where $\mathbf{M}=\mathbf{M}(\mathbf{Z},\mathbf{Q}(y,s))$. If $\mathbf{Z}\cap\mathbf{Q}(y,s)\neq\varnothing$, then $\mathbf{M}(\mathbf{Z},\mathbf{Q}(y,s))=\mathbf{Z}\cap\mathbf{Q}(y,s)$ and $A_+=P_{\mathbf{Z}\cap\mathbf{Q}(y,s)}A$.

Now suppose that $\chi:\Omega\times\Omega\to 2^{\mathbf{R}^n}$ is a choice rule for determining secant conditions.

DEFINITION 2.1. The choice rule χ and the secant conditions determined by it are *accurate* if χ has the property with $\mathbf{Z}$ that there exists an $\alpha\geqq 0$ such that for any $x, x_+\in\Omega$ and any $y\in\chi(x,x_+)$, one has

$$\|P^{\perp}_{\mathbf{Z}\cap\mathbf{Q}(0,s)}(G-A_*)\|\leqq\alpha\sigma(x,x_+)^p \tag{2.2}$$

for every $G\in\mathbf{M}(\mathbf{Z},\mathbf{Q}(y,s))$, where $s=x_+-x$, $\sigma(x,x_+)=\max\{|x-x_*|,|x_+-x_*|\}$, and

$$A_*=P_{\mathbf{Z}}[F'(x_*)-C(x_*)]. \tag{2.3}$$

Dennis and Walker (1981, Thm. 3.1) show that the sparse secant updates associated with an accurate choice rule exhibit bounded deterioration and therefore yield methods with desirable local q-linear convergence properties. We want to extend that theorem in the sparse secant update case by enlarging the sets of secant conditions given by an accurate χ without losing bounded deterioration of the updates determined by them. Toward this end, let us define for $v, w\in\mathbf{R}^n$

$$\mathbf{B}(v,w)=\{Tv+(I-T)w: T=\operatorname{diag}(t_1,\cdots,t_n),\ t_i\in[-1,1],\ i=1,\cdots,n\}. \tag{2.4}$$

One sees that $\mathbf{B}(v,w)$ is just a "box" centered at w which has v as a vertex and sides parallel to the coordinate axes. For a set $S\subseteq\mathbf{R}^n$, we also define

$$\mathbf{B}_{\cup}(v,S)=\bigcup_{w\in S}\mathbf{B}(v,w). \tag{2.5}$$

Our extension of Dennis and Walker (1981, Thm. 3.1) in the sparse secant update case is the theorem below.

THEOREM 2.2. *Let the standard hypothesis hold and let* $\mathbf{Z}$ *have the properties that* A_* *given by* (2.3) *and* $B_*=A_*+C(x_*)$ *are such that* B_* *is invertible and there exists a* r_* *for which*

$$|I-B_*^{-1}F'(x_*)|\leqq r_*<1.$$

Also assume that the choice rule χ *is accurate in the sense of Definition* 2.1. *Under these hypotheses, if* $r\in(r_*,1)$, *then there are positive constants* ε_r, δ_r *such that for* $x_0\in\Omega$ *and* $A_0\in\mathbf{Z}$ *satisfying* $|x_0-x_*|<\varepsilon_r$ *and* $|A_0-A_*|<\delta_r$, *the sequence* $\{x_k\}$ *is well defined by* $B_0=A_0+C(x_0)$ *and*

$$\begin{gathered}x_{k+1}=x_k-B_k^{-1}F(x_k),\\ s_k=x_{k+1}-x_k,\ y_k\in\mathbf{B}_{\cup}(A_ks_k,\chi(x_k,x_{k+1})),\\ B_{k+1}\in\{(A_k)_++C(x_{k+1}),(A_k)_++C(x_k)\},\end{gathered} \tag{2.6}$$

and satisfies $|x_{k+1}-x_*| \leqq r|x_k - x_*|$ *for* $k=0, 1, 2, \cdots$, *where* $(A_k)_+$ *is the sparse secant update of* A_k *with respect to* s_k *and* y_k *given by* (2.1). *Furthermore,* $\{\|B_k\|\}$ *and* $\{\|B_k^{-1}\|\}$ *are uniformly bounded.*

Proof. We will only outline the proof because it is a straightforward application of Dennis and Walker (1981, Thm. A2.1). In order to satisfy the hypothesis of that theorem, we define the update function U in a neighborhood $N=N_1\times N_2$ of (x_*, B_*). From the above hypotheses, one sees that there exist neighborhoods N_1 of x_* and N_2 of B_* such that $N_1 \subseteq \Omega$, N_2 contains only nonsingular matrices, and $x_+ \equiv x - B^{-1}F(x) \in \Omega$ for any $(x, B)\in N = N_1 \times N_2$. Now we make our formal definition.

DEFINITION 2.3. For $(x, B)\in N$ and $x_+\in\Omega$, set $s=x_+-x$, $A=P_{\mathbf{Z}}[B-C(x)]$ and

$$U(x, B)=\{A_+ + C(x_+),\ A_+ + C(x)\colon y\in \mathbf{B}_{\mathrm{U}}(As, \chi(x, x_+))\}, \tag{2.7}$$

where A_+ is the sparse secant update of A with respect to s and y given by (2.1). Any $y\in\mathbf{B}_{\mathrm{U}}(As, \chi(x, x_+))$ will be called *admissible* (with respect to χ, $\mathbf{Z}$, and $\|\cdot\|$).

Lemma 2.4 below shows that U so defined has the bounded deterioration property required by Dennis and Walker (1981, Thm. A2.1), and so the theorem follows from that result.

LEMMA 2.4. *Let the standard hypothesis hold, and suppose that the choice rule* χ *is accurate in the sense of Definition* 2.1. *Let* U *be defined on* N *by Definition* 2.3. *Then the bounded deterioration inequality*

$$\|B_+ - B_*\| \leqq \|B-B_*\| + 2(\alpha+2\beta\gamma_C)\sigma(x, x_+)^p \tag{2.8}$$

holds for $(x, B)\in N$ *and* $B_+\in U(x, B)$, *where* α *is the constant of inequality* (2.2) *and* β *is a constant, e.g.* $\sqrt{n}$, *for which* $\|\cdot\|\leqq\beta|\cdot|$.

Proof. We only prove that (2.8) holds for $B_+=A_+ + C(x_+)$, since the proof in the other case is slightly simpler. For convenience, set $C(x)=C$, $C(x_+)=C_+$, $C(x_*)=C_*$, and $\mathbf{Q}(0, s)=\mathbf{N}$.

The choice of $y\in\mathbf{B}_{\mathrm{U}}(As, \chi(x, x_+))$ which determines B_+ can be written as $y=TAs+(I-T)y^\chi$ for $y^\chi\in\chi(x, x_+)$ and $T=\operatorname{diag}(t_1, \cdots, t_n)$ with $|t_i|\leqq 1$ for $i=1, \cdots, n$. Let $G^\chi\in\mathbf{M}(\mathbf{Z}, \mathbf{Q}(y^\chi, s))$ and set $G=TA+(I-T)G^\chi$. We first need to show that $G\in\mathbf{M}(\mathbf{Z}, \mathbf{Q}(y, s))$. Our device for doing this is to show that $G_+=G$, where $G_+=P_{\mathbf{M}}G$ is the sparse secant update of G into $\mathbf{M}=\mathbf{M}(\mathbf{Z}, \mathbf{Q}(y, s))$. Since $G^\chi\in\mathbf{M}(\mathbf{Z}, \mathbf{Q}(y^\chi, s))$,

$$0=\sum_{i=1}^{n}(s^TP_is)^+\varepsilon_i^T(y^\chi-G^\chi s)\varepsilon_i(P_is)^T;$$

and multiplying both sides by the diagonal matrix $(I-T)$ gives

$$\begin{aligned}
0&=\sum_{i=1}^{n}(s^TP_is)^+\varepsilon_i^T(I-T)(y^\chi-G^\chi s)\varepsilon_i(P_is)^T\\
&=\sum_{i=1}^{n}(s^TP_is)^+\varepsilon_i^T[TAs-TAs+(I-T)(y^\chi-G^\chi s)]\varepsilon_i(P_is)^T\\
&=\sum_{i=1}^{n}(s^TP_is)^+\varepsilon_i^T(y-Gs)\varepsilon_i(P_is)^T=G_+-G.
\end{aligned}$$

Now we use (2.23) of Dennis and Walker (1981) with $\mathscr{S}=\mathbf{Z}$ to get

$$\begin{aligned}
B_+-B_*=A_++C_+-B_*&=P_{\mathbf{Z}\cap\mathbf{N}}(A+C_+-B_*)+P_{\mathbf{Z}\cap\mathbf{N}}^{\perp}(G+C_+-B_*)\\
&=P_{\mathbf{Z}\cap\mathbf{N}}(A+C_+-B_*)+P_{\mathbf{Z}\cap\mathbf{N}}^{\perp}[T(A+C_+-B_*)+(I-T)(G^\chi+C_+-B_*)]\\
&=P_{\mathbf{Z}\cap\mathbf{N}}(A+C_+-B_*)+TP_{\mathbf{Z}\cap\mathbf{N}}^{\perp}(A+C_+-B_*)\\
&\quad+(I-T)P_{\mathbf{Z}\cap\mathbf{N}}^{\perp}(G^\chi+C_+-B_*).
\end{aligned}$$

The last equality is implied by the following: left-multiplication by the diagonal matrix T commutes with the projections onto $\mathbf{Z}$ and $\mathbf{N}$, and $P_{\mathbf{Z}\cap\mathbf{N}} = \lim_{j\to\infty}(P_{\mathbf{Z}}P_{\mathbf{N}})^j$; consequently, left-multiplication by T commutes with $P_{\mathbf{Z}\cap\mathbf{N}}$ and $P^{\perp}_{\mathbf{Z}\cap\mathbf{N}}$. It is this commutativity which makes the "box" the enlargement of χ appropriate to $\mathbf{Z}$ and the Frobenius norm.

To obtain (2.8) from this equation, first note that (2.2) gives

$$\begin{aligned}\|(I-T)P^{\perp}_{\mathbf{Z}\cap\mathbf{N}}(G^{\chi}+C_+-B_*)\| &\leqq 2\|P^{\perp}_{\mathbf{Z}\cap\mathbf{N}}(G^{\chi}-A_*+C_+-C_*)\|\\ &\leqq 2[\|P^{\perp}_{\mathbf{Z}\cap\mathbf{N}}(G^{\chi}-A_*)\|+\beta\gamma_C\sigma(x,x_+)^p]\\ &\leqq 2(\alpha+\beta\gamma_C)\sigma(x,x_+)^p.\end{aligned}$$

To bound the other two terms, use the Pythagorean theorem and the form of T to obtain

$$\begin{aligned}&\|P_{\mathbf{Z}\cap\mathbf{N}}(A+C_+-B_*)+TP^{\perp}_{\mathbf{Z}\cap\mathbf{N}}(A+C_+-B_*)\|^2\\ &\quad\leqq \|P_{\mathbf{Z}\cap\mathbf{N}}(A+C_+-B_*)\|^2+\|TP^{\perp}_{\mathbf{Z}\cap\mathbf{N}}(A+C_+-B_*)\|^2\\ &\quad\leqq \|P_{\mathbf{Z}\cap\mathbf{N}}(A+C_+-B_*)\|^2+\|P^{\perp}_{\mathbf{Z}\cap\mathbf{N}}(A+C_+-B_*)\|^2\\ &\quad= \|A+C_+-B_*\|^2.\end{aligned}$$

Consequently,

$$\begin{aligned}\|B_+-B_*\| &\leqq \|A+C_+-B_*\pm C\|+2(\alpha+\beta\gamma_C)\sigma(x,x_+)^p\\ &\leqq \|B-B_*\|+2(\alpha+2\beta\gamma_C)\sigma(x,x_+)^p;\end{aligned}$$

and the lemma is proved.

The secant conditions admitted in iteration (2.6) are determined explicitly by sets of accurate secant conditions specified by an accurate choice rule. This report is directed principally toward applications in which accurate secant conditions are unobservable, at least in practice; and it may not be clear how to apply Theorem 2.2 in such applications. Consequently, we now give a local q-linear convergence result in which the role of accurate choice in determining admissible secant conditions is more implicit. This result is more directly useful than Theorem 2.2 in many applications of the sort which we have in mind.

We work from the premise that often when accurate secant conditions cannot actually be observed, one can at least determine sets which contain them. Indeed, one might often be able to observe inaccurate secant conditions together with bounds on the inaccuracy in them and through these bounds determine sets in which accurate secant conditions lie. As an illustration, consider the classical case with $C(x)\equiv 0$ in which $y_k^{\chi}=F(x_{k+1})-F(x_k)$ is an accurate secant condition for each k, and suppose that these accurate secant conditions y_k^{χ} are not observable in practice because of inaccuracy in the computed values of F. If one can bound the inaccuracy in computed F-values in some way, then one should be able to specify sets about the inaccurate observed y_k^{χ}-values which contain the true y_k^{χ}-values. To be more specific, let $\tilde{F}(x)=F(x)+N(x)$ be the computed value of $F(x)$ for each $x\in\Omega$. Then for each k, the observed value of y_k^{χ} is

$$\tilde{y}_k=y_k^{\chi}+N(x_{k+1})-N(x_k).$$

If one assumes, as in Barrera and Dennis (1979), that $|N(x_{k+1})-N(x_k)|\leqq\eta_k$ and, hence, $|y_k^{\chi}-\tilde{y}_k|\leqq\eta_k$ for each k, then each set

$$Y_k=\{y\in\mathbf{R}^n: |y-\tilde{y}_k|\leqq\eta_k\} \tag{2.9}$$

contains an accurate secant condition

Sets such as Y_k above are central to our local convergence result below, for they allow the determination of secant conditions which are admissible in iteration (2.6). As a device for determining admissible secant conditions from such sets, let us define for $v \in \mathbf{R}^n$ and compact $S \subseteq \mathbf{R}^n$ a vector $\hat{y} = \hat{y}(v, S)$ as follows: For $i = 1, \cdots, n$, set $l_i = \inf\{\varepsilon_i^T s: s \in S\}$ and $u_i = \sup\{\varepsilon_i^T s: s \in S\}$; then define $\hat{y}$ by

$$\varepsilon_i^T \hat{y} = \begin{cases} l_i & \text{if } \varepsilon_i^T v \leqq l_i, \\ \varepsilon_i^T v & \text{if } l_i \leqq \varepsilon_i^T v \leqq u_i, \\ u_i & \text{if } u_i \leqq \varepsilon_i^T v, \end{cases} \tag{2.10}$$

for $i = 1, \cdots, n$. The vector $\hat{y}$ has the geometrically appealing property that in the $\mathbf{l}_2$ norm, it is the vector closest to v in the smallest box containing S and having sides parallel to the coordinate axes. (See Fig. 2.1 below.) It is reasonable to believe that $\hat{y}$ can be easily determined for most sets S and in particular when S is a set such as Y_k in (2.9) which is naturally prescribed as one containing accurate secant conditions. In our local convergence result, it turns out that admissible secant conditions are found in sets of the form $\mathbf{B}(v, \hat{y})$ given by (2.4) with $w = \hat{y}$. We shall say more about the role of vectors such as $\hat{y}$ in determining admissible secant conditions following the statement of our local convergence result.

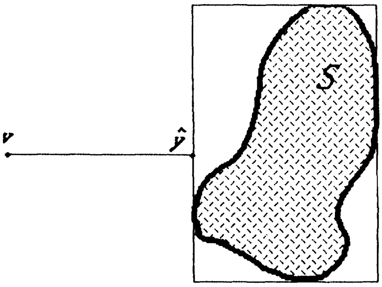

FIG. 2.1

As an aid in showing that sets of the form $\mathbf{B}(v, \hat{y})$ contain admissible secant conditions, let us also define for $v \in \mathbf{R}^n$ and $S \subseteq \mathbf{R}^n$

$$\boldsymbol{B}_\cap(v, S) = \bigcap_{w \in S} \mathbf{B}(v, w), \tag{2.11}$$

where $\mathbf{B}(v, w)$ is given by (2.4). We need the following lemma.

LEMMA 2.5. *For any $v \in \mathbf{R}^n$ and compact $S \subseteq \mathbf{R}^n$, one has*

$$\mathbf{B}(v, \hat{y}) = \mathbf{B}_\cap(v, S) \tag{2.12}$$

where $\hat{y} = \hat{y}(v, S)$ is given by (2.10), *$\mathbf{B}(v, \hat{y})$ is given by* (2.4) *with $w = \hat{y}$, and $\mathbf{B}_\cap(v, S)$ is given by* (2.11).

Proof. To show that $\mathbf{B}(v, \hat{y}) \subseteq \mathbf{B}_\cap(v, S)$, let $y = Tv + (I - T)\hat{y}$ for $T = \operatorname{diag}(t_1, \cdots, t_n)$ with $-1 \leqq t_i \leqq 1$ for $i = 1, \cdots, n$. Let $w \in S$ and note that one can write $\hat{y} = \hat{T}v + (I - \hat{T})w$ for $\hat{T} = \operatorname{diag}(\hat{t}_1, \cdots, \hat{t}_n)$ with $0 \leqq \hat{t}_i \leqq 1$ for $i = 1, \cdots, n$. Indeed, for $i = 1, \cdots, n$, one has that $l_i \leqq \varepsilon_i^T w \leqq u_i$, and so the appropriate $\hat{t}_i$ is given by

$$\hat{t}_i = \begin{cases} \dfrac{\varepsilon_i^T w - l_i}{\varepsilon_i^T (w - v)} & \text{if } \varepsilon_i^T \hat{y} = l_i, \\ 1 & \text{if } \varepsilon_i^T \hat{y} = \varepsilon_i^T v \\ \dfrac{u_i - \varepsilon_i^T w}{\varepsilon_i^T (v - w)} & \text{if } \varepsilon_i^T \hat{y} = u_i. \end{cases}$$

Now one has

$$y = Tv + (I - T)[\hat{T}v + (I - \hat{T})w] = \bar{T}v + (I - \bar{T})w,$$

where $\bar{T} = T + (I - T)\hat{T}$; and it is easily seen that $\bar{T} = \operatorname{diag}(\bar{t}_1, \cdots, \bar{t}_n)$ with $-1 \leqq \bar{t}_i \leqq 1$ for $i = 1, \cdots, n$. Thus $y \in \mathbf{B}(v, w)$, and one concludes that $\mathbf{B}(v, \hat{y}) \subseteq \mathbf{B}_{\cap}(v, S)$.

To show that $\mathbf{B}_{\cap}(v, S) \subseteq \mathbf{B}(v, \hat{y})$, let $w \in \mathbf{B}_{\cap}(v, S)$. It must be shown that for $i = 1, \cdots, n$, one has

$$\varepsilon_i^T w = t_i \varepsilon_i^T v + (1 - t_i)\varepsilon_i^T \hat{y}, \tag{2.13}$$

where $-1 \leqq t_i \leqq 1$. Let $y_1, \cdots, y_n, y^1, \cdots, y^n$ be such that $l_i = \varepsilon_i^T y_i$ and $u_i = \varepsilon_i^T y^i$ for $i = 1, \cdots, n$. For some i, suppose that $\varepsilon_i^T \hat{y} = l_i$. Since $w \in \mathbf{B}(v, y_i)$, one can write $\varepsilon_i^T w = t_i \varepsilon_i^T v + (1 - t_i)\varepsilon_i^T y_i$, where $-1 \leqq t_i \leqq 1$. since $\varepsilon_i^T \hat{y} = l_i = \varepsilon_i^T y_i$, (2.13) holds with this t_i. Since w also belongs to $\mathbf{B}(v, y^i)$, (2.13) holds similarly for an appropriate t_i if $\varepsilon_i^T \hat{y} = u_i$. The remaining case is that in which $l_i \leqq \varepsilon_i^T v = \varepsilon_i^T \hat{y} \leqq u_i$. In this case, one has for some $t_i, t^i \in [-1, 1]$ that both $\varepsilon_i^T w = t_i \varepsilon_i^T v + (1 - t_i)\varepsilon_i^T y_i = t_i \varepsilon_i^T v + (1 - t_i) l_i \leqq \varepsilon_i^T v$ and, similarly, $\varepsilon_i^T w = t^i \varepsilon_i^T v + (1 - t^i) u_i \geqq \varepsilon_i^T v$. consequently, $\varepsilon_i^T w = \varepsilon_i^T v$ and (2.13) holds with $t_i = 1$. This completes the proof.

With this lemma, we can easily obtain from Theorem 2.2 a local convergence result suitable for application in any circumstances in which one can determine for each k some set Y_k containing an accurate secant condition.

COROLLARY 2.6. *Let the hypotheses of Theorem* 2.2 *hold. If* $r \in (r_*, 1)$, *then there are positive constants* ε_r, δ_r *such that for* $x_0 \in \Omega$ *and* $A_0 \in \mathbf{Z}$ *satisfying* $|x_0 - x_*| < \varepsilon_r$ *and* $|A_0 - A_*| < \delta_r$, *the sequence* $\{x_k\}$ *is well defined by* $B_0 = A_0 + C(x_0)$ *and*

$$\begin{aligned} &x_{k+1} = x_k - B_k^{-1} F(x_k), \\ &\textit{a choice of } Y_k \subseteq \mathbf{R}^n \textit{ which satisfies } Y_k \cap \chi(x_k, x_{k+1}) \neq \varnothing, \\ &s_k = x_{k+1} - x_k,\ y_k \in \mathbf{B}(A_k s_k, \hat{y}_k), \\ &B_{k+1} \in \{(A_k)_+ + C(x_{k+1}), (A_k)_+ + C(x_k)\}, \end{aligned} \tag{2.14}$$

and satisfies $|x_{k+1} - x_*| \leqq r|x_k - x_*|$ *for* $k = 0, 1, 2, \cdots$, *where* $\hat{y}_k = \hat{y}(A_k s_k, Y_k)$ *is determined as in* (2.10) *with* $v = A_k s_k$ *and* $S = Y_k$ *and* $(A_k)_+$ *is the sparse secant update of* A_k *with respect to* s_k *and* y_k *given by* (2.1). *Furthermore,* $\{\|B_k\|\}$ *and* $\{\|B_k^{-1}\|\}$ *are uniformly bounded.*

Proof. Since we assume that $\chi(x_k, x_{k+1}) \cap Y_k \neq \varnothing$ for each k, one sees that $\mathbf{B}_{\cap}(A_k s_k, Y_k) \subseteq \mathbf{B}_{\cup}(A_k s_k, \chi(x_k, x_{k+1}))$ provided $x_k, x_{k+1} \in \Omega$. Since $\mathbf{B}(A_k s_k, \hat{y}_k) = \mathbf{B}_{\cap}(A_k s_k, Y_k)$ by Lemma 2.5, the corollary now follows immediately from Theorem 2.2.

This local convergence result merits some discussion. We begin by considering the properties of the vectors $\hat{y}_k = \hat{y}(A_k s_k, Y_k)$ as secant conditions. We are not recommending the use of these vectors in (2.14); indeed, we will argue at the end of § 3 that the vectors $\bar{y}_k \in \mathbf{B}(A_k s_k, \hat{y}_k)$ described there and pictured below in Fig. 2.2 should lead to faster convergence. Nevertheless, the $\hat{y}_k$ have certain very appealing properties that explain the usefulness of the sets $\mathbf{B}(A_k s_k, \hat{y}_k)$ in choosing secant conditions.

Let us first note that if $A \in \mathbf{Z}$ and if $s, y \in \mathbf{R}^n$ with $s \neq 0$, then A_+ given by (2.1) satisfies

$$\|A_+ - A\|^2 = \sum_{i=1}^{n} (|P_i s|^+ |\varepsilon_i^T (y - As)|)^2. \tag{2.15}$$

Suppose that y is allowed to range over a subset $S \subseteq \mathbf{R}^n$. One sees that $\|A_+ - A\|$ is minimized over S by any $y \in S$ for which the weighted $\mathbf{l}_2$-norm on the right-hand side of (2.15) is minimal. In the case of the full Broyden update in which $\mathbf{Z} = \mathbf{R}^{n \times n}$ and

$P_i = I$ for $i = 1, \cdots, n$, (2.15) becomes

$$(2.16) \qquad \|A_+ - A\|^2 = |y - As|^2/|s|^2.$$

Thus in this case, $\|A_+ - A\|$ is minimized by any $y \in S$ for which $|y - As|$ is minimal. For general $\mathbf{Z}$, it may not be so clear which vectors in S minimize $\|A_+ - A\|$. However, this norm will certainly be minimized if a $y \in S$ can be found for which each $|\varepsilon_i^T(y - As)|$ is minimal for $i = 1, \cdots, n$. Such a y may not exist for some sets S, but suppose that S is a box with sides parallel to the coordinate axes. In this case, there is a unique $y \in S$ for which each $|\varepsilon_i^T(y - As)|$ is minimal for $i = 1, \cdots, n$, namely $\hat{y} = \hat{y}(As, S)$ determined as in (2.10); and this $\hat{y}$ is also the unique minimizer of $|y - As|$ over S.

Returning to the context of iteration (2.14), we denote by S_k the smallest box containing Y_k and having sides parallel to the coordinate axes. One sees from the above discussion that $\hat{y}_k$ is the unique vector which minimizes $\|(A_k)_+ - A_k\|$ over S_k. In this sense, it represents the most conservative secant condition among all those determined by vectors in a set, namely S_k, which is known to contain a vector associated with an accurate secant condition. We find it attractive that $\hat{y}_k$ and therefore any $y_k \in \mathbf{B}(A_k s_k, \hat{y}_k)$ conservatively impose on their $(A_k)_+$ the presumably useful but less than totally trustworthy secant information in vectors in S_k. In one form or another, the principle of making conservative use of current secant information has consistently led to successful updates in the past. Our procedure for specifying secant conditions in (2.14) can be viewed as a filter on the secant information that produces a "box" of secant conditions. All of them seem to lead to the same radius of local convergence, although not the same speed of convergence, as we will see in the next section.

There is another way in which one can regard any $y_k \in \mathbf{B}(A_k s_k, \hat{y}_k)$ as making conservative yet effective use of current secant information. Specifically, consider an "accurate" $y_k^\chi \in \chi(x_k, x_{k+1}) \cap Y_k$ and observe that $|\varepsilon_i^T(y_k - y_k^\chi)| \leqq |\varepsilon_i^T(A_k s_k - y_k^\chi)|$ and hence,

$$(2.17) \qquad |\varepsilon_i^T[(A_k)_+ s_k - y_k^\chi]| \leqq |\varepsilon_i^T(A_k s_k - y_k^\chi)| \quad \text{for } i = 1, \cdots, n,$$

where $(A_k)_+$ is the sparse secant update of A_k in $\mathbf{Z}$ determined by y_k. In other words, if $(A_k)_+$ is determined by y_k, then each component of $(A_k)_+ s_k$ is at least as close as its counterpart in $A_k s_k$ to the corresponding component of an "accurate" y_k^χ.

Since Y_k is the subset of S_k in which vectors specifying accurate secant conditions are known to lie, one might wonder if a better condition might be obtained by minimizing $\|(A_k)_+ - A_k\|$ over Y_k. We think not. To support our position, we note that if $(A_k)_+$ is determined by a vector in Y_k such that $\|(A_k)_+ - A_k\|$ is minimal, then (2.17) may not hold; whether or not (2.17) does hold in this case depends on the geometry of Y_k. To illustrate this and to lend some perspective to this discussion, let us recall that Barrera and Dennis (1979) consider full Broyden updating with $C(x) = 0$ when the accurate secant condition determined by $y_k^\chi = F(x_{k+1}) - F(x_k)$ is not observable because of inaccuracy in computed values of F. As we indicated previously, they assume that the inaccuracy in computed F-values is bounded so that y_k^χ is contained in Y_k given by (2.9). In Barrera and Dennis (1979), the secant condition studied is that determined by the vector in Y_k which is closest to $A_k s_k$ in the $\mathbf{l}_2$-norm, and which we denote here by y_k^{BD}. One sees from (2.16) that y_k^{BD} is the unique minimizer for full Broyden updating of $\|(A_k)_+ - A_k\|$in Y_k. In Fig. 2.2 below, we have shown a circumstance in which (2.17) does not hold for $(A_k)_+$ determined by y_k^{BD}, while, of course, it does hold for $(A_k)_+$ determined by $y_k \in \mathbf{B}(A_k s_k, \hat{y}_k)$. In Fig. 2.2, $\tilde{y}_k$ denotes the observed (inaccurate) value of y_k^χ as before.

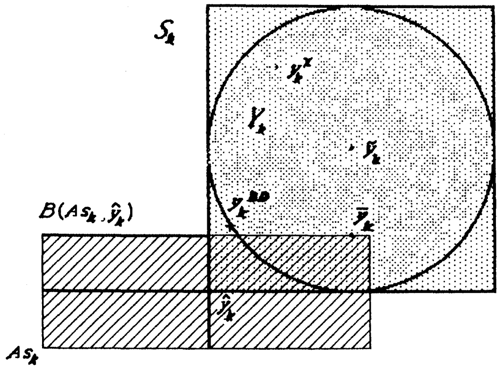

FIG. 2.2

Barrera and Dennis (1979) report on numerical experiments which suggest that the vectors y_k^{BD} are useful in practice, although they give no supporting convergence analysis. We feel that the analysis given above reinforces their encouraging computational results, even though their methods do not actually fall within its scope. Indeed, the circumstance illustrated in Fig. 2.2 not withstanding, it seems very unlikely that a vector y_k^{BD} will fail to lie within the box $\mathbf{B}(A_k s_k, \hat{y}_k)$ unless one is very near the solution. Thus, updates determined by these vectors seem very likely to exhibit bounded deterioration in practice, at least until some stopping criterion is met, and thus to lead to methods with satisfactory local q-linear convergence properties. Similarly, the analysis here strongly suggests that in most traditional applications, the usual secant conditions determined through the use of inaccurate computed function values should be completely adequate.

3. Speed of convergence. In the preceding section, it was shown that under the hypotheses of Theorem 2.2 sequences of iterates produced by iteration (2.6) exhibit local q-linear convergence. In this section, we analyze the behavior of these iteration sequences in greater depth. Specifically, we consider q-linearly convergent sequences produced by (2.6) with the objective of estimating their asymptotic speeds of convergence.

When accurate secant conditions are used in all iterations of (2.6), the resulting asymptotic speed of convergence has been determined by Dennis and Walker (1981). To recall those results, suppose that the hypotheses of Theorem 2.2 hold and that $\{x_k\}$ is an iteration sequence produced by (2.6) which converges q-linearly to x_*. One can easily modify Theorem 3.3 of Dennis and Walker (1981) to include the case $B_{k+1} = (A_k)_+ + C(x_k)$. From this modification, one sees that if the secant condition is determined by $y_k \in \chi(x_k, x_{k+1})$ at each iteration of (2.6), then

$$\overline{\lim_{k\to\infty}} \frac{|x_{k+1}-x_*|}{|x_k-x_*|} = \overline{\lim_{k\to\infty}} \left| [I - B_*^{-1}F'(x_*)] \frac{x_k - x_*}{|x_k - x_*|} \right|. \tag{3.1}$$

In particular,

$$\overline{\lim_{k\to\infty}} \frac{|x_{k+1}-x_*|}{|x_k-x_*|} \leqq |I - B_*^{-1}F'(x_*)| \leqq r_*. \tag{3.2}$$

This is to say that if an accurate secant condition is used at each iteration, then the asymptotic q-linear convergence rate constant is the same as that of the ideal stationary iteration of the form (1.2) which takes each B_k as close as possible to $F'(x_*)$.

We offer two results below concerning the extent to which inaccuracy in secant conditions affects asymptotic speeds of convergence. These results are given as Corollaries 3.2 and 3.3 of Theorem 3.1, which can be regarded as the central technical result of this section. One can certainly obtain other asymptotic convergence results as corollaries of Theorem 3.1, but we feel that the two given here are particularly appealing and useful. Corollary 3.2 can be interpreted as saying that if at least some minimal positive proportion of accurate secant information is imposed on the update at each iteration, then the optimal asymptotic speed of convergence given by (3.1) and (3.2) will result. Corollary 3.3 shows precisely how much given amounts of inaccuracy in secant conditions can cause the asymptotic speed of convergence to deteriorate from this optimal asymptotic speed. It is in the tradition of previous results in Dennis and Moré (1974) and Dennis and Walker (1981) in that it relates the asymptotic speed of convergence to the extent to which the successive B_k's approximate the action of B_* in the directions of the s_k's. From it, one sees that the asymptotic speed of convergence can still be regarded as optimal, given the inaccuracy in the secant conditions used to update the B_k's.

Throughout this section, we assume that $\chi: \mathbf{R}^n \times \mathbf{R}^n \rightarrow 2^{\mathbf{R}^n}$ is a choice rule which is accurate in the sense of Definition 2.1. The statements and proofs of our results rely heavily on the characteristic property of each $y_k \in \mathbf{B}_\cup(A_k s_k, \chi(x_k, x_{k+1}))$ in (2.6) that it can be written as

$$y_k = T_k A_k s_k + (I - T_k) y_k^\chi, \quad \text{for some } y_k^\chi \in \chi(x_k, x_{k+1}) \tag{3.3}$$

$$\text{and} \quad T_k = \operatorname{diag}(t_{1k}, \cdots, t_{nk}). \quad t_{ik} \in [-1, 1] \quad \text{for } i = 1, \cdots, n. \tag{3.4}$$

We use (3.3) and (3.4) freely below with minimal explanation.

THEOREM 3.1. *Let the hypotheses of Theorem 2.1 hold and assume that for some $x_0 \in \mathbf{R}^n$ and $A_0 \in \mathbf{Z}$, $\{x_k\}$ is a sequence defined by (2.6) which converges q-linearly to x_* with $x_k \neq x_*$ for all k. Assume further that $\{\|B_k\|\}$ is uniformly bounded. For each T_k in (3.3) and (3.4), set $\bar{t}_{ik} = t_{ik}|P_i s_k||P_i s_k|^+$ for $i = 1, \cdots, n$. Then*

$$\lim_{k\rightarrow\infty} \sqrt{1 - \bar{t}_{ik}^2}\,|P_i s_k|^+ \varepsilon_i^T (B_k - B_*) P_i s_k = 0, \tag{3.5}$$

$$\lim_{k\rightarrow\infty} (1 - |\bar{t}_{ik}|)|P_i s_k|^+ \varepsilon_i^T (B_k - B_*) P_i s_k = 0, \tag{3.6}$$

for $i = 1, \cdots, n$. It follows that

$$\lim_{k\rightarrow\infty} \frac{(I - \bar{T}_k^2)^{1/2} (B_k - B_*) s_k}{|s_k|} = 0, \tag{3.7}$$

$$\lim_{k\rightarrow\infty} \frac{(I - \bar{\bar{T}}_k)(B_k - B_*) s_k}{|s_k|} = 0, \tag{3.8}$$

where $\bar{\mathrm{T}}_k \equiv \operatorname{diag}(\bar{t}_{1k}, \cdots, \bar{t}_{nk})$ and $\bar{\bar{T}}_k$ is the matrix of absolute values of entries of $\bar{T}_k$.

Proof. For convenience, set $C(x_k) = C_k$, $C(x_*) = C_*$, $\mathbf{Q}(0, s_k) = \mathbf{N}_k$, $\sigma_k = \max\{|x_k - x_*|, |x_{k+1} - x_*|\}$, and $E_k = B_k - B_*$. Note that

$$E_k = A_k + C_k + A_* - C_* = P_{\mathbf{Z}} E_k + P_{\mathbf{Z}}^\perp (C_k - C_*) \tag{3.9}$$

and, as in the proof of Lemma 2.4,

$$\begin{aligned} E_{k+1} &= P_{\mathbf{Z}\cap\mathbf{N}_k} E_k + T_k P_{\mathbf{Z}\cap\mathbf{N}_k}^\perp E_k + O(\sigma_k^p) = P_{\mathbf{Z}\cap\mathbf{N}_k} E_k + T_k (I - P_{\mathbf{Z}\cap\mathbf{N}_k}) E_k + O(\sigma_k^p) \\ &= T_k E_k + (I - T_k) P_{\mathbf{Z}\cap\mathbf{N}_k} E_k + O(\sigma_k^p) = T_k P_{\mathbf{Z}} E_k + (I - T_k) P_{\mathbf{Z}\cap\mathbf{N}_k} E_k + O(\sigma_k^p). \end{aligned} \tag{3.10}$$

The ith row of this equation is

$$\begin{aligned}\varepsilon_i^T E_{k+1} &= t_{ik} P_i \varepsilon_i^T E_k + (1-t_{ik})\{P_i \varepsilon_i^T E_k - (P_i \varepsilon_i^T E_k) P_i s_k [|P_i s_k|^2]^+ (P_i s_k)^T\} + O(\sigma_k^p)\\ &= P_i \varepsilon_i^T E_k - (1-\bar t_{ik})\{(P_i \varepsilon_i^T E_k) P_i s_k [|P_i s_k|^2]^+ (P_i s_k)^T\} + O(\sigma_k^p) \qquad (3.11)\\ &\equiv v_{ik} + O(\sigma_k^p).\end{aligned}$$

We want to bound $|v_{ik}|$, and direct computation gives

$$|v_{ik}|^2 = |P_i \varepsilon_i^T E_k|^2 - (1-\bar t_{ik}^2)[(P_i \varepsilon_i^T E_k) P_i s_k |P_i s_k|^+]^2.$$

Since $(P_i \varepsilon_i^T E_k) P_i s_k = (\varepsilon_i^T E_k) P_i s_k$ and $|P_i| \leqq 1$,

$$|v_{ik}|^2 \leqq |\varepsilon_i^T E_k|^2 - (1-\bar t_{ik}^2)[(\varepsilon_i^T E_k) P_i s_k |P_i s_k|^+]^2 \equiv \eta_{ik}^2 - \psi_{ik}^2.$$

Now we use the general inequality

$$(a^2-b^2)^{1/2} \leqq a - \frac{b^2}{2a}$$

to obtain

$$|v_{ik}| \leqq \eta_{ik} - \frac{\psi_{ik}^2}{2\eta_{ik}}.$$

It follows from (3.11) that

$$\eta_{i,k+1} \leqq \eta_{ik} - \frac{\psi_{ik}^2}{2\eta_{ik}} + O(\sigma_k^p).$$

By a standard argument (see, e.g., Dennis and Walker (1981, p. 966)), this implies that $\sum_k \psi_{ik}^2 < \infty$ and hence, $0 = \lim_{k\to\infty} \psi_{ik}$, which is (3.5). To get (3.6), we note that

$$\psi_{ik} \geqq [(1-|\bar t_{ik}|)/(1+|\bar t_{ik}|)]^{1/2} \psi_{ik} \geqq 0.$$

Finally, we prove (3.8) from (3.6), since (3.7) follows in just the same way from (3.5). For any $1 \leqq i \leqq n$, (3.9) implies

$$\begin{aligned}\varepsilon_i^T \frac{(I-\bar{\bar{T}}_k)E_k s_k}{|s_k|} &= \frac{(1-|\bar t_{ik}|)\varepsilon_i^T[P_Z E_k s_k + P_Z^\perp (C_k - C_*) s_k]}{|s_k|} = \frac{(1-|\bar t_{ik}|)\varepsilon_i^T P_Z E_k s_k}{|s_k|} + O(\sigma_k^p)\\ &= \frac{(1-|\bar t_{ik}|)\varepsilon_i^T E_k P_i s_k}{|s_k|} + O(\sigma_k^p).\end{aligned}$$

Since $|s_k| \geqq |P_i s_k|$, (3.8) follows from (3.6).

COROLLARY 3.2. *Let $\{x_k\}$ be an iteration sequence generated under the hypotheses of Theorem* 3.1. *If for some τ, $0 \leqq \tau < 1$, T_k in* (3.3) *and* (3.4) *satisfies $0 \leqq |\bar t_{ik}| \leqq \tau$ for $1 \leqq i \leqq n$ and $k = 0, 1, \cdots$, then*

$$\overline{\lim_{k\to\infty}} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} = |I - B_*^{-1} F'(x_*)| \leqq r_*. \qquad (3.12)$$

Proof. It is immediate from (3.8) and the existence of the bound $\tau < 1$ that

$$\lim_{k\to\infty} \frac{E_k s_k}{|s_k|} = 0.$$

Since $\{x_k\}$ converges q-linearly to x_*,

$$\lim_{k\to\infty} \frac{|s_k|}{|x_k - x_*|} < \infty,$$

and so the result follows from Theorem A.1 of the appendix.

COROLLARY 3.3. *Let $\{x_k\}$ be an iteration sequence generated under the hypotheses of Theorem* 3.1. *Then for y_k as in* (3.3),

$$\lim_{k\to\infty} \left[\frac{(B_k - B_*)s_k}{|s_k|} - \bar{D}_k \left(\frac{y_k - y_k^{\chi}}{|s_k|} \right) \right] = 0 \tag{3.13}$$

and

$$\lim_{k\to\infty} \left[\frac{(B_k - B_*)s_k}{|x_k - x_*|} - \bar{D}_k \left(\frac{y_k - y_k^{\chi}}{|x_k - x_*|} \right) \right] = 0, \tag{3.14}$$

where $\bar{D}_k = \operatorname{diag}(d_{1k}, \cdots, d_{nk})$ with $d_{ik} = \bar{t}_{ik} |\bar{t}_{ik}|^+$ for $i = 1, \cdots, n$. If

$$\overline{\lim_{k\to\infty}} \frac{|B_*^{-1} \bar{D}_k (y_k - y_k^{\chi})|}{|x_k - x_*|} \leqq \lambda, \tag{3.15}$$

then

$$\overline{\lim_{k\to\infty}} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} \leqq r_* + \lambda. \tag{3.16}$$

If

$$\overline{\lim_{k\to\infty}} \frac{|B_*^{-1} \bar{D}_k (y_k - y_k^{\chi})|}{|s_k|} \leqq \mu < 1, \tag{3.17}$$

then

$$\overline{\lim_{k\to\infty}} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} \leqq \frac{r_* + \mu}{1 - \mu}. \tag{3.18}$$

Proof. We first establish (3.13) from (3.8). Since $\{x_k\}$ converges q-linearly to x_*, (3.14) then follows from (3.13). For each k, we define E_k, $\mathbf{N}_k$, and σ_k as in the proof of Theorem 3.1. From (3.8), we have that

$$\lim_{k\to\infty} \left[\frac{E_k s_k}{|s_k|} - \frac{\bar{\bar{T}}_k E_k s_k}{|s_k|} \right] = 0.$$

Since

$$E_{k+1} s_k = T_k P^{\perp}_{\mathbf{Z} \cap \mathbf{N}_k} E_k s_k + O(\sigma_k^p) s_k = T_k E_k s_k + O(\sigma_k^p) s_k$$

by (3.10), it follows that

$$\lim_{k\to\infty} \left[\frac{E_k s_k}{|s_k|} - \bar{D}_k \frac{E_{k+1} s_k}{|s_k|} \right] = 0.$$

But $E_{k+1} = (A_k)_+ - A_* + O(\sigma_k^p)$, so

$$\lim_{k\to\infty} \left[\frac{E_k s_k}{|s_k|} - \bar{D}_k \frac{[(A_k)_+ - A_*] s_k}{|s_k|} \right] = 0.$$

Let $G_k \in \mathbf{M}(\mathbf{Z}, \mathbf{Q}(y_k, s_k))$ and $G_k^\chi \in \mathbf{M}(\mathbf{Z}, \mathbf{Q}(y_k^\chi, s_k))$. Note that if $P_i s_k \neq 0$, then $\varepsilon_i^T y_k = \varepsilon_i^T G_k s_k$ and $\varepsilon_i^T y_k^\chi = \varepsilon_i^T G_k^\chi s_k$. From this and from Dennis and Walker (1981), p. 959, it follows that

$$\begin{aligned}\bar{D}_k[(A_k)_+ - A_*]s_k &= \bar{D}_k[P_{\mathbf{Z}\cap\mathbf{N}_k}^\perp(G_k - A_* \pm G_k^\chi)]s_k = \bar{D}_k[P_{\mathbf{Z}\cap\mathbf{N}_k}^\perp(G_k - G_k^\chi)]s_k + O(\sigma_k^p)s_k \\ &= \bar{D}_k(y_k - y_k^\chi) + O(\sigma_k^p)s_k;\end{aligned}$$

and (3.13) follows.

It also follows from continuity that

$$\text{(3.19)} \qquad \lim_{k\to\infty}\left[\frac{B_*^{-1}E_k s_k}{|s_k|} - \frac{B_*^{-1}\bar{D}_k(y_k - y_k^\chi)}{|s_k|}\right] = 0$$

and

$$\text{(3.20)} \qquad \lim_{k\to\infty}\left[\frac{B_*^{-1}E_k s_k}{|x_k - x_*|} - \frac{B_*^{-1}\bar{D}_k(y_k - y_k^\chi)}{|x_k - x_*|}\right] = 0.$$

If (3.15) holds, then (3.16) follows immediately from (3.20) and (A.2) of the appendix. If (3.17) holds, then (3.19) implies

$$\overline{\lim_{k\to\infty}} \frac{B_*^{-1}E_k s_k}{|s_k|} \leqq \mu.$$

We see from (A.2) of the appendix that for

$$r = \overline{\lim_{k\to\infty}} \frac{|x_{k+1} - x_*|}{|x_k - x_*|},$$

we have

$$r \leqq r_* + \mu \overline{\lim_{k\leftarrow\infty}} \frac{|s_k|}{|x_k - x_*|} \leqq r_* + \mu(1 + r);$$

and $r \leqq (r_* + \mu)/(1 - \mu)$ follows if $\mu < 1$.

Remark. It is clear from (3.15)-(3.18) that we would like $y_k \in \mathbf{B}(A_k s_k, \hat{y}_k)$ to be as near as possible to y_k^χ in the $\mathbf{l}_2$ norm scaled by $B_*^{-1}\bar{D}_k$. We are unlikely to be able to make practical use of this information because of the scaled norm and because y_k^χ may be anywhere in Y_k. It does suggest that an interesting secant condition is $\bar{y}_k$, the b that solves

$$\text{(3.21)} \qquad \min_{b\in\mathbf{B}(A_k s_k, \hat{y}_k)} \max_{y\in Y_k} |b - y|.$$

In many cases, e.g. (2.9), Y_k will be derived from an observation $\tilde{y}_k$ of y_k^χ in such a way that $\bar{y}_k$ is simply the b that solves

$$\text{(3.22)} \qquad \min_{b\in\mathbf{B}(A_k s_k, \hat{y}_k)} |b - \tilde{y}_k|.$$

In fact, if Y_k in (2.9) is obtained from $\tilde{y}_k$ by allowing only for small rounding errors, then Y_k is probably so small compared to $\mathbf{B}(A_k s_k, \hat{y}_k)$ that $\bar{y}_k = \tilde{y}_k$ will hold until the iteration is stopped. This is consistent with the good performance in practice of Broyden's method with the traditional secant condition.

4. An application.

We conclude with an illustration of how one might make use of the local convergence analysis given in this report. Assume the following:

(1) The updates are sparse Broyden updates with $C(x) = 0$ and $F'(x) \in \mathbf{Z}$ for each $x \in \Omega$, and so $y_k^\chi = F(x_{k+1}) - F(x_k)$ is an accurate secant conditions for each k.

(2) For each $x \in \Omega$, one computes an inaccurate value $\tilde{F}(x) = F(x) + N(x)$, where $|N(x)| \leqq \varepsilon_F |F(x)|$.

(3) For each k, the observed secant conditions is $\tilde{y}_k = y_k^\chi + N(x_{k+1}) - N(x_k)$, and so $y_k^\chi \in Y_k$, where Y_k is given by (2.9) with $\eta_k = \varepsilon_F(|F(x_{k+1})| + |F(x_k)|)$.

To emphasize the role of the inaccuracy in secant conditions most clearly, we maintain our policy of disregarding all other inaccuracy. In particular, it is assumed that the accurate value $F(x_k)$ is used explicitly in (1.2) for each k. Although this leaves us considering a somewhat artificial situation, to do otherwise would cloud the issues of interest. We leave to future work the analysis of the effects of all inaccuracy on the performance of methods emplying sparse Broyden updates.

According to Corollary 2.6, a method (1.2) employing sparse secant updates determined by the $\bar{y}_k$ given by (3.21)-(3.22) will enjoy local q-linear convergence. Let us show how one might apply the results of § 3 to determine asymptotic speeds of convergence which can be obtained by using the $\bar{y}_k$. Let $\{x_k\}$ be a q-linearly convergent iteration sequence. From our assumption, one has for each k,

$$\overline{\lim_{k\to\infty}} \frac{|\bar{y}_k - y_k^\chi|}{|x_k - x_*|} \leqq (1+\sqrt{n})\varepsilon_F |F'(x_*)| \left(1 + \overline{\lim_{k\to\infty}} \frac{|x_{k+1} - x_*|}{|x_k - x_*|}\right).$$

Now, $B_* = F'(x_*)$ and $r_* = 0$; and so it follows from (3.15) and (3.16) of Corollary 3.3 that

$$\overline{\lim_{k\to\infty}} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} \leqq \overline{\lim_{k\to\infty}} \frac{|F'(x_*)^{-1}\bar{D}_k(\bar{y}_k - y_k^\chi)|}{|x_k - x_*|} \leqq (1+\sqrt{n})\varepsilon_F K_* \left(1 + \overline{\lim_{k\to\infty}} \frac{|x_{k+1} - x_*|}{|x_k - x_*|}\right), \tag{4.1}$$

where $K_* = |F'(x_*)^{-1}| |F'(x_*)|$ is the condition number of $F'(x_*)$ in the norm $|\cdot|$. If $(1+\sqrt{n})\varepsilon_F K_* < 1$, then one obtains from this the estimate

$$\overline{\lim_{k\to\infty}} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} \leqq \frac{(1+\sqrt{n})\varepsilon_F K_*}{1 - (1+\sqrt{n})\varepsilon_F K_*} \tag{4.2}$$

of the asymptotic speed of convergence.

The estimate (4.2) and the underlying analysis suggest that if one has some control over the amount of inaccuracy in computed values of F used in determining secant conditions, then the results of § 3 might be used to provide guidelines for exercising this control in order to achieve a reasonable balance of convergence speed and computational expense. One sees from (4.2) that a q-linearly convergent iteration sequence $\{x_k\}$ will exhibit asymptotic q-linear convergence which is as fast as desired, provided the relative inaccuracy in computed functions values used to determine secant conditions is kept sufficiently small. In some applications, it might too costly to maintain a sufficiently high level of accuracy in computed function values at all iterations. A natural strategy in such applications is to use relatively inaccurate computed function values in the early iterations and then to increase function evaluation accuracy in the later stages as the solution is neared. The analysis leading to (4.2) bears out the validity of this strategy in the case under consideration. To see that this is so, suppose that for each k, the relative inaccuracy in the computed value of $F(x_k)$ is bounded by $(\varepsilon_F)_k$. Then by a derivation similar to that of (4.2), one can show that a q-linearly convergent iteration sequence $\{x_k\}$ satisfies

$$\overline{\lim_{k\to\infty}} \frac{|x_{k+1} - x_*|}{|x_k - x_*|} \leqq \frac{(1+\sqrt{n})\, \overline{\lim}_{k\to\infty} (\varepsilon_F)_k K_*}{1 - (1+\sqrt{n})\, \overline{\lim}_{k\to\infty} (\varepsilon_F)_k K_*}. \tag{4.3}$$

Note that (4.3) implies that $\{x_k\}$ converges q-superlinearly if $\lim_{k\to\infty} (\varepsilon_F)_k = 0$.

Although (4.3) provides a satisfactory guarantee that adequately fast asymptotic convergence will result whenever $\overline{\lim}_{k\to\infty}(\varepsilon_F)_k$ is sufficiently small, there remains the question of how one should choose $(\varepsilon_F)_k$ which is suggested by Corollary 3.2. Suppose that an initial $(\varepsilon_F)_0>0$ and suitable values $\tau\in(0,1)$ and $(\varepsilon_F)_{\min}>0$ are given at the outset. If one has arrived at the kth iteration for some $k>0$, then take $(\varepsilon_F)_k=(\varepsilon_F)_{k-1}$ as a trial value and set $\eta_k=(\varepsilon_F)_k\{|F(x_{k+1})|+|F(x_k)|\}$. Even though the matrix $T_k=\operatorname{diag}(t_{1k},\cdots,t_{nk})$ of (3.3) and (3.4) appropriate for $y_k=\bar{y}_k$ cannot be observed, if

$$(4.4)\qquad |\varepsilon_i^T B_k s_k-\varepsilon_i^T\tilde{y}_k|\geqq\left(1+\frac{1}{\tau}\right)\eta_k,\qquad 1=1,\cdots,n,$$

then one necessarily has $0\leqq|\bar{t}_{ik}|\leqq\tau$ for $i=1,\cdots,n$, where $\bar{t}_{ik}$ is as in Theorem 3.1 and Corollary 3.2; furthermore, $\bar{y}_k=\tilde{y}_k$ in this case. If (4.4) holds, then accept $(\varepsilon_F)_k$ and $\bar{y}_k=\tilde{y}_k$ and proceed. Otherwise, reduce $(\varepsilon_F)_k$ and redefine η_k until either (4.4) holds or $(\varepsilon_F)_k=(\varepsilon_F)_{\min}$. Note that without the restriction $(\varepsilon_F)_k\geqq(\varepsilon_F)_{\min}$, (4.4) will holds for sufficiently small $(\varepsilon_F)_k>0$ except in the unlikely event that $B_k s_k=y_k^\chi$. The restriction $(\varepsilon_F)_k\geqq(\varepsilon_F)_{\min}$ guards against this event and also prevents requiring excessive accuracy in function evaluations.

This manner of choosing each $(\varepsilon_F)_k$ has the appeal that if $\{x_k\}$ is assumed to converge q-linearly to x_*, then unless or until $(\varepsilon_F)_k=(\varepsilon_F)_{\min}$ for some k, one can use the traditional secant conditions and presumably be on the way to enjoying the q-superlinear convergence guaranteed by Corollary 3.2. Actually, it always happens that $(\varepsilon_F)_k=(\varepsilon_F)_{\min}$ eventually, as we show in Proposition 3.4 below. However, this should be all right, because $(\varepsilon_F)_{\min}$ can be chosen so small in most applications that the asymptotic speed of convergence given by (4.3) with $\overline{\lim}_{k\to\infty}(\varepsilon_F)_k=(\varepsilon_F)_{\min}$ will be adequately fast.

PROPOSITION 4.1. *If* $\{x_k\}$ *converges q-linearly to* x_*, *then* $(\varepsilon_F)_k=(\varepsilon_F)_{\min}$ *for sufficiently large k.*

Proof. Suppose the contrary, i.e., that $(\varepsilon_F)_k>(\varepsilon_F)_{\min}$ for all k. Then (4.4) holds for all k, and it follows that

$$(4.5)\qquad \left(1+\frac{1}{\tau}\right)\overline{\lim_{k\to\infty}}\frac{\eta_k}{|s_k|}\leqq\overline{\lim_{k\to\infty}}\frac{|\varepsilon_i^T B_k s_k-\varepsilon_i^T\tilde{y}_k|}{|s_k|}$$

for $i=1,\cdots,n$. It also follows from Corollary 3.2 that $\{x_k\}$ converges q-superlinearly to x_*, and so

$$0=\lim_{k\to\infty}\frac{[B_k-F'(x_*)]s_k}{|s_k|}=\lim_{k\to\infty}\frac{(B_k s_k-y_k^\chi)}{|s_k|}.$$

See Dennis and Moré (1974) or Dennis and Walker (1981). Consequently, for $i=1,\cdots,n$,

$$(4.6)\qquad \overline{\lim_{k\to\infty}}\frac{|\varepsilon_i^T B_k s_k-\varepsilon_i^T\tilde{y}_k|}{|s_k|}\leqq\overline{\lim_{k\to\infty}}\frac{|B_k s_k-y_k^\chi|}{|s_k|}+\overline{\lim_{k\to\infty}}\frac{|\varepsilon_i^T y_k^\chi-\varepsilon_i^T\tilde{y}_k|}{|s_k|}\leqq\overline{\lim_{k\to\infty}}\frac{\eta_k}{|s_k|}.$$

Since $\{x_k\}$ converges q-linearly to x_* and the $(\varepsilon_F)_k$'s are bounded away from 0, one has that

$$0<\overline{\lim_{k\to\infty}}\frac{\eta_k}{|s_k|}<\infty.$$

It follows that (4.5) and (4.6) are contradictory, and the proposition is proved.

Appendix. In this appendix, we establish the result referenced in the proofs of Corollaries 3.2 and 3.3. This result is an extension of Dennis and Walker (1981, Thm. A3.1), and we suspect that it might have uses beyond those in this paper. Our interest here is in a general quasi-Newton iteration (1.2) for solving the problem (1.1) with no presumption about the form of each B_k or the manner in which it is generated. We suppose that F satisfies the standard hypothesis given in the introduction, although we have no particular interest in a computed part C of F'. We use $|\cdot|$ to denote both a vector norm on $\mathbf{R}^n$ and its subordinate operator norm on $\mathbf{R}^{n\times n}$. For an iteration sequence $\{x_k\}$, we denote $s_k = x_{k+1} - x_k$ as before. It is also convenient to set $e_k = x_k - x_*$ for each k.

THEOREM A.1. *Suppose that F satisfies the standard hypothesis and that $\{x_k\}$ is a sequence generated by* (1.2) *which converges to x_* with $x_k \neq x_*$ for all k. If $B_* \in \mathbf{R}^{n\times n}$ is any invertible matrix then for any norm $|\cdot|$, one has*

$$\lim_{k\to\infty} \left| \frac{e_{k+1}}{|e_k|} - [I - B_*^{-1}F'(x_*)]\frac{e_k}{|e_k|} + \frac{B_*^{-1}(B_k - B_*)s_k}{|e_k|} \right| = 0. \tag{A.1}$$

Setting $r_* = |I - B_*^{-1}F'(x_*)|$, one has in particular that

$$\begin{aligned} \overline{\lim_{k\to\infty}} \frac{|e_{k+1}|}{|e_k|} &\leqq \overline{\lim_{k\to\infty}} \left| [I - B_*^{-1}F'(x_*)]\frac{e_k}{|e_k|} + \frac{B_*^{-1}(B_k - B_*)s_k}{|e_k|} \right| \\ &\leqq r_* + \overline{\lim_{k\to\infty}} \frac{|B_*^{-1}(B_k - B_*)s_k|}{|e_k|} \\ &\leqq r_* + \overline{\lim_{k\to\infty}} \frac{|B_*^{-1}(B_k - B_*)s_k|}{|s_k|} \overline{\lim_{k\to\infty}} \frac{|s_k|}{|e_k|}. \end{aligned} \tag{A.2}$$

Proof. From the standard hypothesis on F and (1.2), one has that

$$B_k s_k = -F(x_k) = -F'(x_*)e_k + O(e_k^{p+1}).$$

It follows from multiplication by B_*^{-1} that

$$s_k + B_*^{-1}F'(x_*)e_k + B_*^{-1}(B_k - B_*)s_k = O(e_k^{p+1}).$$

With $e_k \neq 0$, one verifies from this equation that

$$\frac{e_{k+1}}{|e_k|} - [I - B_*^{-1}F'(x_*)]\frac{e_k}{|e_k|} + \frac{B_*^{-1}(B_k - B_*)s_k}{|e_k|} = O(e_k^p),$$

and (A.1) and (A.2) follow immediately.

5. Acknowledgment. An improvement in the presentation of these results was facilitated by two conscientious referee reports.

REFERENCES

P. BARRERA AND J. E. DENNIS JR. (1979), *When to stop making quasi-Newton updates*, presented at the Tenth International Symposium on Mathematical Programming, Montreal.

C. G. BROYDEN (1965), *A class of methods for solving nonlinear simultaneous equations*, Math. Comp., 19, pp. 577–593.

———, (1971), *The convergence of an algorithm for solving sparse nonlinear systems*, Math. Comp., 25, pp. 285–294.

J. E. DENNIS JR. AND J. J. MORÉ (1974), *A characterization of superlinear convergence and its application to quasi-Newton methods*, Math. Comp., 28, pp. 549–560.

J. E. DENNIS JR. AND R. B. SCHNABEL (1979), *Least change secant updates for quasi-Newton methods*, SIAM Rev., 21, pp. 443-459.

———, (1983), *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs, NJ.

J. E. DENNIS JR. AND H. F. WALKER (1981), *Convergence theorems for least-change secant update methods*, this Journal, 18, pp. 949-987.

———, (1984), *Inaccuracy in Quasi-Newton methods: local improvement theorems*, Rice MASC TR 83-11, Math. Prog. Stud. 22, 1984, pp. 70-85.

E. S. MARWIL (1979), *Convergence results for Schubert's method for solving sparse nonlinear equations*, this Journal, 16, pp. 588-604.

J. J. MORÉ, B. S. GARBOW AND K. E. HILLSTROM (1980), *User guide for* MINPACK-1, Argonne National Labs Report ANL-80-74.

J. M. ORTEGA AND W. C. RHEINBOLDT (1970), *Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, New York.

J. K. REID (1973), *Least squares solution of sparse systems of nonlinear equations by a modified Marquardt algorithm*, In Proc. NATO Conf. at Cambridge, July 1972, North-Holland, Amsterdam, pp. 437-445.

L. K. SCHUBERT (1970), *Modification of a quasi-Newton method for nonlinear equations with a sparse Jacobian*, Math. Comp., 24, pp. 27-30.

T. J. YPMA (1983), *The effect of rounding errors on Newton-like methods*, IMA J. Numer. Anal., 3, pp. 109-118.