LOCAL CONVERGENCE THEORY
OF INEXACT NEWTON METHODS
BASED ON STRUCTURED LEAST CHANGE UPDATES

JOSÉ MARIO MARTÍNEZ

ABSTRACT. In this paper we introduce a local convergence theory for Least Change Secant Update methods. This theory includes most known methods of this class, as well as some new interesting quasi-Newton methods. Further, we prove that this class of LCSU updates may be used to generate iterative linear methods to solve the Newton linear equation in the Inexact-Newton context. Convergence at a q-superlinear rate (or at an "ideal" linear rate, in the sense of Dennis-Walker) of the Inexact Newton methods generated in this way is proved, independently of the number of iterations used in the linear iterative subalgorithm. We apply the new theory to some particular methods.

1. INTRODUCTION

The problem we study is the following: Given $F: \Omega \subset \mathbf{R}^n \to \mathbf{R}^n$, $F = (f_1, \ldots, f_n)^T \in C^1(\Omega)$, and $J(x) = F'(x)$, find $x \in \Omega$ satisfying

$$(1.1) \qquad F(x) = 0.$$

The most popular method for solving this type of problem is Newton's method [13, 35, 40]. This is an iterative method which, starting from an arbitrary $x^0 \in \Omega$, computes the successive approximations x^k of the solution of (1.1) according to

$$(1.2) \qquad x^{k+1} = x^k - J(x^k)^{-1} F(x^k).$$

Hence, during each iteration of Newton's method, (a) the Jacobian matrix $J(x^k)$ must be calculated, and (b) the linear $n \times n$ system

$$(1.3) \qquad J(x^k)z = -F(x^k)$$

must be solved, in order to obtain x^{k+1}.

Newton's method has excellent local convergence properties (see [13, 35, 40]). This fact motivated many people in the last 25 years to introduce methods with similar local convergence properties, but which are less expensive, in the sense that either the task (a), the task (b), or both, are avoided.

There are situations where the Jacobian matrix $J(x^k)$ is easily available, but a direct method for obtaining z in (1.3) is very expensive in terms of time or storage. In these cases, Inexact Newton methods are recommended. At each iteration of an Inexact Newton method, equation (1.3) is solved only approximately, using some iterative algorithm for linear systems.

Ortega and Rheinboldt [35] analyze the case where a fixed number of iterations of the linear iterative algorithm is used at each main iteration. They obtain linear convergence results under a suitable assumption on the spectral radius of the iteration matrix at the solution (see [35, Theorem 10.3.10]), and they make special reference to the use of the SOR method as inner iteration.

Later, Dembo, Eisenstat, and Steihaug [7] studied the situation where an approximate solution of (1.3) is considered without specifying the inner algorithm used to obtain it. They prove that if

$$(1.4) \quad \|J(x^k)z + F(x^k)\| \le \theta_k \|F(x^k)\|, \qquad 0 \le \theta_k \le \theta < 1, \ k = 0, 1, 2, \ldots,$$

then local linear convergence is achieved, and convergence is superlinear if $\lim \theta_k = 0$.

On the other hand, quasi-Newton methods [11, 13] were essentially devised to handle the situation where the derivatives of F are not available, or are difficult to calculate. They are based on the iteration formula

$$(1.5) \qquad\qquad x^{k+1} = x^k - B_k^{-1} F(x^k),$$

where B_{k+1} is obtained from B_k using nonexpensive procedures which, in principle, do not involve derivatives. Alternatively, B_{k+1}^{-1} may be obtained directly from B_k^{-1}, or a suitable factorization of B_{k+1} may be obtained from the factorization of B_k (see [13, 16, 34]). Among many successful algorithms which were devised using the quasi-Newton idea, the best known are the first method of Broyden [2] (for small, unstructured problems), the BFGS method (see [13]) (for unconstrained minimization problems), the method of Dennis-Gay-Welsch [8, 13] (for nonlinear least squares), Schubert's method [3, 33, 39] (for large sparse nonlinear systems), and the Marwil-Toint sparse symmetric update [32, 33, 41, 42]. All of them belong to a more restricted class of quasi-Newton methods, the class of Least Change Secant Update methods (LCSU), analyzed in [12, 13]. In general, these methods have local superlinear convergence properties, although some authors introduced methods that do not have such properties, and yet still seem to have potential usefulness [9, 29, 30]. In fact, Dennis and Marwil [9] were the first to introduce a method where the (sparse) LU factorization of B_{k+1} is obtained directly from the LU factorization of B_k. Their method is not locally convergent unless implemented with a convenient restart procedure. However, using their idea, Johnson and Austria [24], and Chadee [5] introduced superlinearly convergent direct secant update methods, and Martínez [31] introduced a family of quasi-Newton methods with direct secant updates of matrix factorizations with superlinear convergence properties.

Our strong belief that in many large problems the evaluation of analytic derivatives is easy but that secant methods may still be useful to save on linear algebra is also supported by some recent work on the application of secant methods to infinite-dimensional problems (see, for example, [18, 25, 38]).

The motivation for this paper is the following: Consider the very frequent situation where the Jacobian matrix $J(x)$ is not difficult to evaluate but the direct solution of (1.3) is prohibitively time- or memory-consuming. Surely, an Inexact Newton method should be a good choice for solving (1.1). Now, iterative linear methods for solving (1.3) may also be prohibitively slow unless implemented using a suitable preconditioning scheme. Our idea is to use LCSU matrices, and their generalizations, as preconditioners for an application of an Inexact Newton method. Of course, this idea would be of no advantage if we tried to use a "Schubert-type" formula, where the whole matrix B_k is updated, and then factored in order to obtain the quasi-Newton iteration. But it may be advantageous in large-scale problems if we use direct secant updates of matrix factorizations, or even if we use classical low-rank updates (see [13]) storing a limited number of "past updating vectors".

In §2 of this paper we introduce the main model algorithm, and we discuss some examples. In §3 we prove the main convergence results. In §4 we prove convergence for the particular examples discussed in §2, using the theory of §3. Some conclusions are drawn in §5.

2. GENERAL DESCRIPTION OF THE METHODS AND EXAMPLES

2.1. The main model algorithm.
Let $F: \Omega \subset \mathbf{R}^n \to \mathbf{R}^n$, $F \in C^1(\Omega)$, Ω an open set.

Let X be a finite-dimensional linear space. For each $x, z \in \Omega$, let $\langle \, , \, \rangle_{x,z}$ be a scalar product on X and $\| \cdot \|_{x,z}$ its associated norm.

$| \cdot |$ will denote an arbitrary norm on $\mathbf{R}^n$ and its associated matrix norm throughout the paper.

For all $x, z \in \Omega$ let $V = V(x, z)$ be a linear manifold contained in X. For $x, z \in \Omega$, $E \in X$, let us call $P_{xz}(E)$ the orthogonal projection of E on $V(x, z)$, related to the norm $\| \cdot \|_{x,z}$.

Let $\varphi: \Omega \times X \to \mathbf{R}^{n \times n}$. The main model algorithm considered in this paper is described below.

Algorithm 2.1. Let x^0 be an arbitrary initial point, $E_0 \in X$. Given x^k and E_k, we compute x_Q^k, x^{k+1}, and E_{k+1} performing the following steps:

Step 1. Compute the quasi-Newton point x_Q^k:

$$(2.1) \qquad z_0 = -\varphi(x^k, E_k)^{-1} F(x^k),$$

$$(2.2) \qquad x_Q^k = x^k + z_0.$$

Step 2. Compute the new parameter approximation:

$$(2.3) \qquad E_{k+1} = P_{x^k x_Q^k}(E_k).$$

Step 3. Obtain x^{k+1} such that

$$(2.4) \qquad\qquad |x^{k+1} - x_N^k| \le |x_Q^k - x_N^k|,$$

where x_N^k is the Newton point, defined by

$$(2.5) \qquad\qquad x_N^k = x^k - J(x^k)^{-1}F(x^k).$$

A point x^{k+1} which satisfies (2.4) may be obtained applying some linear iterative algorithm to equation (1.3). Every convergent linear iterative algorithm starting from $x_Q^k - x^k$ produces, after a finite number of iterations, a point which satisfies (2.4), since $x_N^k - x^k$ is the exact solution of (1.3). However, this is not a practical observation, since we can compute only the residual of (1.3), and not the error, at each inner iteration. Fortunately, some linear iteration procedures have the property of decreasing the norm of the error monotonically, and, so, (2.4) is guaranteed using an arbitrary number of inner iterations. We mention below two of these procedures.

Conjugate gradient inner iteration. Choose $q_k \ge 0$. For $l = 0, 1, \ldots, q_k - 1$ compute z_{l+1}, the $(l+1)$st iteration of the Conjugate Gradient Method (see [1, 17, 23]) applied to the equation

$$(2.6) \qquad\qquad \varphi(x^k, E_{k+1})^{-1}[J(x^k)z + F(x^k)] = 0.$$

If $|\cdot|$ is the 2-norm in $\mathbf{R}^n$, $\{z_l\}$ satisfies the monotonically decreasing error property

$$(2.7) \qquad\qquad |z_{l+1} - J(x^k)^{-1}F(x^k)| \le |z_l - J(x^k)^{-1}F(x^k)|$$

(see [1]). Therefore, $x^{k+1} = x^k + z_{q_k}$ satisfies (2.4).

Iterative refinement inner iteration. Choose $q_k \ge 0$. For $l = 0, 1, \ldots, q_k - 1$, compute

$$(2.8) \qquad\qquad z_{l+1} = z_l - \varphi(x^k, E_{k+1})^{-1}[J(x^k)z_l + F(x^k)],$$

and set $x^{k+1} = x^k + z_{q_k}$.

We will prove in §3 that under some starting conditions this procedure computes a point which satisfies (2.4), independently of the norm used.

Remark. Observe that a "pure" quasi-Newton algorithm may be defined as a particular case of the inner iteration procedures by taking $q_k \equiv 0$, $k = 0, 1, 2, \ldots$.

The usefulness of considering $q_k > 0$ seems to be limited to situations where the computation of the linearization $J(x^k)z + F(x^k)$ is cheaper than the computation of the residual $F(x^k + z)$. In fact, if this is not the case, an algorithm based on

$$z_{l+1} = z_l - \varphi(x^k, E_{k+1})^{-1}F(x^k + z_l)$$

instead of (2.8) is probably more efficient. Observe that an analogous comment applies to most Inexact Newton procedures, since the iterative methods used

to solve the Newtonian linear system usually compute the linearized residuals. Nevertheless, there are many problems for which the evaluation of F is much more expensive than the computation of $J(x)z + F(x)$. Indeed, the cost of evaluating J, once F has been computed, is often negligible.

The conditions under which Algorithm 2.1 is well defined and locally convergent are given in §3. In this section we will exhibit several examples of methods which have the general form of the algorithms given above.

2.2. Using an updated factorization of a simplified Jacobian. Many practical problems of type (1.1) have the following characteristics:

(i) J is a large sparse matrix, but its LU (QR) factorization is not very sparse.

(ii) Neglecting some coefficients of J, we obtain a simplified matrix N whose LU (QR) factorization is pleasantly sparse.

(iii) $N(x)$ "dominates" $J(x)$ in a sense which will be made precise later (see (4.1)).

(iv) $J(x)$ is not difficult to compute.

Østerby and Zlatev [36] implicitly analyzed this type of problem in the linear case. In fact, they propose to solve linear systems with the characteristics (i)–(iii), dropping the small elements of the LU factorization of the coefficient matrix, and applying an iterative refinement procedure afterwards. The algorithm which we are going to introduce in this section may be viewed as a secant adaptation of their procedure for the nonlinear case.

Let us now define precisely the type of problems we want to consider. We assume that:

(a) For all $x \in \Omega$, $J(x) = N(x) + C(x)$.

(b) For all $x \in \Omega$, $N(x) = \mathscr{A}(x)^{-1}\mathscr{R}(x)$, where $(\mathscr{A}(x), \mathscr{R}(x)) \in S$, a linear manifold contained in $\mathbf{R}^{n \times n} \times \mathbf{R}^{n \times n}$.

Algorithm 2.2. Let $x^0 \in \Omega$ be an arbitrary initial point, and let $(A_0, R_0) \in S$. For $k = 0, 1, 2, \dots$, compute x^{k+1} and (A_{k+1}, R_{k+1}) as follows:

Step 1. $x_Q^k = x^k - (A_k^{-1}R_k)^{-1}F(x^k)$.

Step 2. Compute (A_{k+1}, R_{k+1}) as the solution of the problem:

$$
\begin{aligned}
\text{minimize} \quad & \|A - A_k\|_F^2 + \|R - R_k\|_F^2 \\
\text{s.t.} \quad & (A, R) \in S, \\
& Rs - Ay = 0 \\
& s = x_Q^k - x^k, \; y = N(x^k)s.
\end{aligned}
$$

Step 3. Choose $q_k \geq 0$. Compute x^{k+1} satisfying (2.4), using some linear iterative procedure, and
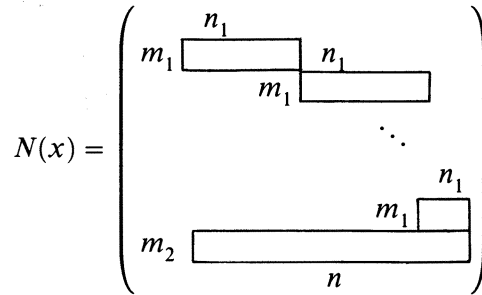
$$
\varphi(x, (A, R)) = A^{-1}R.
$$

For this algorithm, we define:

$$X = \mathbf{R}^{n \times n} \times \mathbf{R}^{n \times n},$$

$$\|(A, R)\|_{x,z}^2 = \|(A, R)\|^2 = \|A\|_F^2 + \|R\|_F^2 \quad \text{for all } x, z \in \Omega,$$

$$V(x, z) = S \cap \{(A, R) \in X | R(z - x) - AN(x)(z - x) = 0\}.$$

Therefore, Algorithm 2.2 is a particular case of Algorithm 2.1.

Example. Suppose that the structure of $N(x)$ is block-angular (see [15]) consisting of (say) k_1 rectangular $m_1 \times n_1$ "small" blocks ($m_1 \leq n_1$) and a "master" $m_2 \times n$ block, so that $k_1 m_1 + m_2 = n$. The configuration is depicted below:

Let $\mathscr{A}(x)$ be a lower triangular $n \times n$ matrix, and $\mathscr{R}(x)$ an orthogonal $n \times n$ matrix such that

$$\mathscr{A}(x)N(x) = \mathscr{R}(x).$$

Both $\mathscr{A}(x)$ and $\mathscr{R}(x)$ have a block-angular structure and may be calculated by standard procedures using $O(k_1 m_1^2 n_1 + m_2 m_1 n + m_2^2 n)$ flops (see [17, §6.2]). Now, the calculation of $N(x)s$ involves $k_1 m_1 n_1 + m_2 n$ flops, and the solution of the optimization problem at Step 2 uses $O(k_1 m_1 n_1 + m_2 n)$ flops (see [31]). Therefore, the computation of A_{k+1}, R_{k+1} using Algorithm 2.2 is, roughly, m_1 times less expensive than the factorization of $N(x)$. Similar considerations would apply if we were to use the LU factorization of $N(x)^T$.

Remark. The definition of y at Step 2 of Algorithm 2.2 may be replaced by some finite difference scheme (see [13, Chapter 11]), but this replacement is not advantageous if $N(x)$ is only mildly nonlinear. If $C(x) = 0$ for all $x \in \Omega$, Algorithm 2.2 represents the family of quasi-Newton methods with direct secant updates of matrix factorizations, introduced in [31]. In this case, we may define $y = F(x_Q^k) - F(x^k)$. Some of the best-known least change secant methods for solving systems of nonlinear equations, such as the first and second methods of Broyden, Schubert's method, Powell's symmetric method, the sparse PSB method of Marwil-Toint, the Johnson-Austria method and Chadee's method, belong to this family.

2.3. Partitioned quasi-Newton methods.
Griewank and Toint [19–22] introduced the family of Partitioned Quasi-Newton Methods for solving some classes of large-scale optimization problems or nonlinear systems of equations. We

will analyze here the application to nonlinear systems of equations described by Toint [43].

Let us assume that

$$F(x) = F_1(x) + \cdots + F_m(x),$$

where, for each $i = 1, \ldots, m$, there exist matrices $U_i \in \mathbf{R}^{n \times m_i}$ and $W_i \in \mathbf{R}^{n_i \times n}$ such that $F_i(x) = U_i G_i(W_i x)$ for a certain C^1-function $G_i : \mathbf{R}^{n_i} \to \mathbf{R}^{m_i}$. Therefore,

$$F_i'(x) = U_i G_i'(W_i x) W_i.$$

Problems of this type arise in the application of the finite element method to boundary value problems. In these cases, $F(x)$ is decomposed into a sum of functions related to each element of the discretization (see [43]), and the range and domain of an element function F_i have low dimension. The columns of U_i span the range of each F_i' and the rows of W_i span its domain, for all values of the variables.

With these hypotheses, Toint's partitioned quasi-Newton method may be described as follows:

Algorithm 2.3. Let $x^0 \in \Omega$ and $B_i^0 \in \mathbf{R}^{m_i \times n_i}$, $i = 1, \ldots, m$, be arbitrary initializations. For $k = 0, 1, 2, \ldots$, compute x^{k+1} and B_i^{k+1} by performing the following steps:

Step 1 (New point):

$$x^{k+1} = x^k - \left(\sum_{i=1}^{m} U_i B_i^k W_i \right)^{-1} F(x^k).$$

Step 2 (Updating): For $i = 1, \ldots, m$ define

$$y_i = F_i(x^{k+1}) - F_i(x^k), \qquad s = x^{k+1} - x^k,$$

and let B_i^{k+1} be the solution of

$$
\begin{aligned}
\text{minimize} \quad & \|B - B_i^k\|_F \\
\text{s.t.} \quad & U_i B W_i s = y_i.
\end{aligned}
$$

For this algorithm, we define

$$X = \{(B_1, \ldots, B_m) \mid B_i \in \mathbf{R}^{m_i \times n_i}\},$$

$$\|\cdot\|_{x,z} = \|\cdot\| \quad \text{for all } x, z,$$

$$\|(B_1, \ldots, B_m)\|^2 = \|B_1\|_F^2 + \cdots + \|B_m\|_F^2,$$

$$\varphi(x, (B_1, \ldots, B_m)) = \sum_{i=1}^{m} U_i B_i W_i,$$

and

$$V(x, z) = \{(B_1, \ldots, B_m) \in X \mid U_i B_i W_i(z - x) = F_i(z) - F_i(x), \; i = 1, \ldots, m\}.$$

Remark. The computational effort involved in Step 1 of Algorithm 2.3 is the same as that associated with the solution of a linear system which must be carried out at each iteration of Newton's method. Therefore, if the derivatives of F are easily available, Newton's method should be more efficient than Algorithm 2.3, because of its quadratic order of convergence. Hence, Algorithm 2.3 is interesting only if $F'(x)$ is not available, or very difficult to calculate. For this reason, the inner iteration step is not used in this algorithm, and we defined it setting $q_k \equiv 0$. The same observation applies to Schubert's method, as was mentioned in the Introduction, as well as to Algorithm 2.4 defined below.

2.4. Secant augmentations of the Gauss-Newton method. Let us consider the Nonlinear Least Squares problem (see [13]):

$$\text{minimize}_{x \in \mathbf{R}^n} \quad \frac{1}{2} \sum_{i=1}^{m} r_i(x)^2 .$$

This problem may be thought of as a system of nonlinear equations, writing

$$(2.9) \qquad R(x) = (r_1(x), \ldots, r_m(x))^T , \qquad F(x) = R'(x)^T R(x) .$$

Hence

$$(2.10) \qquad J(x) = R'(x)^T R'(x) + \sum_{i=1}^{m} r_i(x) \nabla^2 r_i(x) .$$

Unlike the first term, the second term in (2.10) is difficult to compute, so it is sometimes neglected, as in the classical Gauss-Newton Method, or replaced by a diagonal matrix, as in the Levenberg-Marquardt method [13, 27, 28]. Dennis, Gay, and Welsch [8] proposed to replace it by a weighted secant approximation, introducing a DFP type update. Later, Dennis and Walker [14] proposed a BFGS type update formula for the difference between $J(x)^{-1}$ and $[R'(x)^T R(x)]^{-1}$. Both methods are particular cases of the Model Algorithm 2.1. Let us describe the last one.

Algorithm 2.4. Let $x^0 \in \Omega$ and E_0 be an arbitrary symmetric initial matrix. For $k = 0, 1, 2, \ldots$, compute x^{k+1}, E_{k+1} by performing the following steps:
  Step 1 (New point):

$$x^{k+1} = x^k - \{[R'(x^k)^T R'(x^k)]^{-1} + E_k\} R'(x^k)^T R(x^k) .$$

Step 2 (Updating): Let G be a symmetric positive definite matrix such that

$$(2.11) \qquad Gs = y , \qquad s = x^{k+1} - x^k , \quad y = F(x^{k+1}) - F(x^k) .$$

Set $G = LL^T$, the Cholesky factorization of G. We define E_{k+1} to be the solution of

$$(2.12) \qquad \begin{array}{ll} \text{minimize} & \|L^T(E - E_k)L\|_F \\ \text{s.t.} & E \text{ symmetric}, \\ & Ey = s - [R'(x^{k+1})^T R'(x^{k+1})]^{-1} y . \end{array}$$

Let us develop a closed formula for the solution of (2.12). Writing $H = L^T E L$ and $H_k = L^T E_k L$, we may express (2.12) as:

$$\text{minimize} \quad \|H - H_k\|_F$$
(2.13)
$$\text{s.t.} \quad H \text{ symmetric}$$
$$HL^{-1}y = L^T s - L^T[R'(x^{k+1})^T R'(x^{k+1})]^{-1}y.$$

Using the symmetric PSB formula [13, pp. 195–198; 37], we see that the solution of (2.13) is

$$H_{k+1} = H_k + \frac{(\tilde{s} - H_k\tilde{y})\tilde{y}^T + \tilde{y}(\tilde{s} - H_k\tilde{y})^T}{\tilde{y}^T\tilde{y}} - \frac{(\tilde{s} - H_k\tilde{y})^T\tilde{y}\tilde{y}\tilde{y}^T}{(\tilde{y}^T\tilde{y})^2}$$

with $\tilde{y} = L^{-1}y$, $\tilde{s} = L^T s^{\#}$, $s^{\#} = s - [R'(x^{k+1})^T R'(x^{k+1})]^{-1}y$.
Thus,

$$L^T E_{k+1} L = L^T E_k L + \frac{L^T(s^{\#} - E_k y)y^T L^{-T} + L^{-1}y(s^{\#} - E_k y)^T L}{s^T y}$$
$$- \frac{(s^{\#} - E_k y)^T y L^{-1}yy^T L^{-T}}{(s^T y)^2}.$$

But, by (2.11), we have that $L^{-T}L^{-1}y = s$. Therefore,

$$E_{k+1} = E_k + \frac{(s^{\#} - E_k y)s^T + s(s^{\#} - E_k y)^T}{s^T y}$$
(2.14)
$$- \frac{(s^{\#} - E_k y)^T yss^T}{(s^T y)^2}.$$

So, E_{k+1} does not depend on the particular matrix chosen in (2.11). That is, we may use formula (2.14) as the definition of E_{k+1}, instead of (2.12).

For this algorithm, we define $X = \mathbf{R}^{n \times n}$. For each $x, z \in \Omega$, we define $\|E\|_{x,z} = \|L^T(x, z)EL(x, z)\|_F$, where $L(x, z)L(x, z)^T$ is the Cholesky factorization of

$$\tilde{G} = \int_0^1 J(x + t(z - x)) \, dt.$$

It is easy to verify that $\|\cdot\|_{x,z}$ is associated with a scalar product in X.
Finally,

$$\varphi(x, E) = \{[R'(x)^T R'(x)]^{-1} + E\}^{-1},$$

and

$$V(x, z) = S \cap \{E \in X | E[F(z) - F(x)] = z - x - [R'(z)^T R'(z)]^{-1}[F(z) - F(x)]\},$$

where S is the subspace of symmetric matrices of $\mathbf{R}^{n \times n}$.

As in Algorithm 2.3, the linear algebra calculations involved in the computation of an iteration of Algorithm 2.4 are the same as those involved in a Newton iteration. Therefore, if second derivatives of R are available, Newton's method

should be more efficient. For this reason, we also defined this algorithm without the inner iteration step.

Using similar arguments, we may verify that the BFGS and the DFP algorithms for unconstrained minimization (see [12, 13]) are also particular cases of Algorithm 2.1.

3. Main convergence results

In this section we prove a local convergence theorem for Algorithm 2.1. We use some basic assumptions for achieving the main results. The first assumption concerns the function F.

Assumption 1. Let Ω be an open, convex, and bounded set. Let $x^* \in \Omega$ be such that $F(x^*) = 0$ and $J(x^*)$ is nonsingular. We will assume that there exist $p, M > 0$ such that, for all $x \in \Omega$,

$$(3.1) \qquad |J(x) - J(x^*)| \leq M|x - x^*|^p.$$

Hence, by [4, Lemma 3.1], we have that, for all $x, z \in \Omega$

$$(3.2) \qquad |F(z) - F(x) - J(x^*)(z - x)| \leq M|z - x|\sigma(x, z)^p,$$

where

$$\sigma(x, z) = \max\{|x - x^*|, |z - x^*|\}.$$

The second assumption concerns the function φ.

Assumption 2. There exists $E^* \in X$ such that φ is continuous in a neighborhood of (x^*, E^*), $\varphi(x^*, E^*)$ is nonsingular, and

$$(3.3) \qquad |I - \varphi(x^*, E^*)^{-1} J(x^*)| \leq r^* < 1.$$

We will assume, without loss of generality, that $\varphi(x, E)$ is nonsingular for all (x, E) on its domain.

From now on, $\|\cdot\|$ will denote a norm on X associated with the scalar product $\langle \, , \, \rangle$.

Theorem 3.1. Let F satisfy Assumption 1, let φ satisfy Assumption 2, and let $r \in (r^*, 1)$. Then there exist bounded neighborhoods Ω_1 and $\mathcal{N}$ of x^* and E^*, respectively, such that, for all $x \in \Omega_1$, $E \in \mathcal{N}$, $|\varphi(x, E)^{-1}|$ and $\|E\|$ are uniformly bounded, and

$$(3.4) \qquad |x - \varphi(x, E)^{-1} F(x) - x^*| \leq r|x - x^*|.$$

Proof. The uniform boundedness of $\|E\|$, $|\varphi(x, E)|$, and $|\varphi(x, E)^{-1}|$ follows from the continuity of φ on suitable compact neighborhoods of x^* and E^*.

Now,

$$(3.5) \qquad \begin{aligned} |x - \varphi(x, E)^{-1} F(x) - x^*| &\leq |x - \varphi(x^*, E^*)^{-1} F(x) - x^*| \\ &\quad + |\varphi(x, E)^{-1} - \varphi(x^*, E^*)^{-1}||F(x)|. \end{aligned}$$

But, by (3.2) and (3.3),

$$|x - \varphi(x^*, E^*)^{-1} F(x) - x^*|$$

$$\leq |x - \varphi(x^*, E^*)^{-1} J(x^*)(x - x^*) - x^*|$$

(3.6)
$$+ |\varphi(x^*, E^*)^{-1}| \|[F(x) - J(x^*)(x - x^*)]\|$$

$$\leq \|[I - \varphi(x^*, E^*)^{-1} J(x^*)](x - x^*)| + |\varphi(x^*, E^*)^{-1}| M |x - x^*|^{p+1}$$

$$\leq [r^* + |\varphi(x^*, E^*)^{-1}| M |x - x^*|^p] |x - x^*|.$$

Moreover, by (3.2),

(3.7)
$$|F(x)| \leq |J(x^*)(x - x^*)| + M |x - x^*|^{p+1}$$

$$\leq (|J(x^*)| + M |x - x^*|^p) |x - x^*|.$$

Hence, the desired result follows in a straightforward way from (3.5), (3.6), and (3.7). □

From now on, we will denote Ω_1 by Ω in order to simplify our notation.

The third crucial assumption states that the manifolds $V(x, z)$ are close enough to E^*.

Assumption 3. Assume that, for all $x, z \in \Omega$, there exists $E \in V(x, z)$ such that

(3.8)
$$\|E - E^*\| \leq c_2 \sigma(x, z)^p,$$

where p is defined by (3.1) and $c_2 > 0$.

The fourth assumption concerns the relation between different norms in X.

Assumption 4. Assume that there exist $q, c_1 > 0$ such that, for all $x, z \in \Omega$ and $E \in X$,

(3.9)
$$\|E\|_{x, z} \leq [1 + c_1 \sigma(x, z)^q] \|E\|,$$

(3.10)
$$\|E\| \leq [1 + c_1 \sigma(x, z)^q] \|E\|_{x, z}.$$

The following lemma, a Bounded Deterioration Principle (see [4, 10, 11, 13, 14]), states that the distance between $P_{xz}(E)$ and E^* cannot be much larger than the distance between E and E^*.

Lemma 3.1. *Let* $F, \varphi, V,$ *and* E^* *satisfy Assumptions 1 to 4. Then there exist positive constants* c_3, c_4 *such that for all* $x, z \in \Omega$ *and* $E \in X$,

(3.11)
$$\|P_{xz}(E) - E^*\| \leq [1 + c_4 \sigma(x, z)^q] \|E - E^*\| + c_3 \sigma(x, z)^p.$$

Proof. By (3.10), we have

(3.12)
$$\|P_{xz}(E) - E^*\| \leq [1 + c_1 \sigma(x, z)^q] \|P_{xz}(E) - E^*\|_{x, z}.$$

Let $\hat{E}$ be the orthogonal projection of E^* on $V(x, z)$, related to the norm $\| \cdot \|$. Then, by (3.12),

(3.13)
$$\|P_{xz}(E) - E^*\| \leq [1 + c_1 \sigma(x, z)^q][\|P_{xz}(E) - \hat{E}\|_{x, z} + \|\hat{E} - E^*\|_{x, z}].$$

But P_{xz} is a projection on V, and $\hat{E} \in V$. So,

$$(3.14) \qquad \|P_{xz}(E) - \hat{E}\|_{x,z} \leq \|E - \hat{E}\|_{x,z} \leq \|E - E^*\|_{x,z} + \|\hat{E} - E^*\|_{x,z}.$$

Hence, by (3.9), (3.13), and (3.14),

$$\|P_{xz}(E) - E^*\| \leq [1 + c_1 \sigma(x,z)^q][\|E - E^*\|_{x,z} + 2\|\hat{E} - E^*\|_{x,z}]$$

$$\leq [1 + c_1 \sigma(x,z)^q]^2[\|E - E^*\| + 2\|\hat{E} - E^*\|].$$

Now, by Assumption 3, $\|\hat{E} - E^*\| \leq c_2 \sigma(x,z)^p$. Therefore,

$$\|P_{xz}(E) - E^*\| \leq [1 + c_1 \sigma(x,z)^q]^2[\|E - E^*\| + 2c_2 \sigma(x,z)^p]$$

$$= [1 + 2c_1 \sigma(x,z)^q + c_1^2 \sigma(x,z)^{2q}]\|E - E^*\|$$

$$+ 2[1 + c_1 \sigma(x,z)^q]^2 c_2 \sigma(x,z)^p.$$

Thus, setting

$$(3.15) \qquad d_1 = \sup\{|x - x^*| \,|\, x \in \Omega\},$$

we have

$$\|P_{xz}(E) - E^*\| \leq [1 + (2c_1 + c_1^2 d_1^q)\sigma(x,z)^q]\|E - E^*\|$$

$$+ 2[1 + c_1 d_1^q]^2 c_2 \sigma(x,z)^p.$$

So, (3.11) follows with $c_3 = 2[1 + c_1 d_1^q]^2 c_2$, $c_4 = 2c_1 + c_1^2 d_1^q$. □

Corollary 3.1. *Let* $s = \min\{p, q\}$. *There exists* $c_5 > 0$ *such that*

$$(3.16) \qquad \|P_{xz}(E) - E^*\| \leq \|E - E^*\| + c_5|x - x^*|^s$$

whenever $x, z \in \Omega$, $E \in \mathcal{N}$, *and* $|z - x^*| \leq |x - x^*|$.

Proof. Let us define d_1 as in (3.15) and $d_2 = \sup\{\|E - E^*\|, E \in \mathcal{N}\}$. Then, by (3.11),

$$(3.17) \qquad \|P_{xz}(E) - E^*\| \leq [1 + c_4|x - x^*|^q]\|E - E^*\| + c_3|x - x^*|^p$$

$$\leq \|E - E^*\| + c_4 d_2|x - x^*|^q + c_3|x - x^*|^p.$$

Thus, the desired result follows in a straightforward way from (3.17). □

Now, we are able to prove a local convergence theorem for Algorithm 2.1. Recall the definition of the "Newton point"

$$x_N^k = x^k - J(x^k)^{-1} F(x^k).$$

Theorem 3.2. *Let* F, φ, V, *and* E^* *satisfy Assumptions 1 to 4, assume that* $\{x^k\}$ *is defined by Algorithm 2.1, and let* $r_1 \in (r^*, 1)$. *Then there exist* $\varepsilon = \varepsilon(r_1)$, $\delta = \delta(r_1)$, *and* $c_6 > 0$ *such that, if* $|x^0 - x^*| \leq \varepsilon$ *and* $\|E_0 - E^*\| \leq \delta$, *the sequence generated by Algorithm 2.1 is well defined, and for all* $k = 0, 1, 2, \ldots$

$$|x^{k+1} - x^*| \leq r_1|x^k - x^*|.$$

Moreover, for all $k, j = 0, 1, 2, \ldots$

$$\|E_{k+j} - E^*\| \leq \|E_k - E^*\| + c_6|x^k - x^*|^s,$$

where $s = \min\{p, q\}$.

Proof. Set $r \in (r^*, r_1)$, and suppose that Ω and $\mathscr{N}$ are as in Theorem 3.1. Let $\delta_1, \delta, \varepsilon > 0$ be such that

(3.18)
$$\{E \in X \mid \|E - E^*\| \leq \delta_1\} \subset \mathscr{N},$$

(3.19)
$$\delta + c_5 \varepsilon^s / (1 - r_1^s) < \delta_1,$$

and

(3.20)
$$B_\varepsilon = \{x \in \mathbf{R}^n \mid |x - x^*| \leq \varepsilon\} \subset \Omega.$$

Moreover, assume that ε is small enough so that $x - J(x)^{-1} F(x)$ is well defined for all $x \in B_\varepsilon$, and

(3.21)
$$|x - J(x)^{-1} F(x) - x^*| < \frac{r_1 - r}{2} |x - x^*|.$$

Of course, the existence of an ε which satisfies (3.21) is guaranteed by the local convergence results of Newton's method [13, 35, 40], but it also follows easily as a corollary of Theorem 3.1.

Let us prove that for all $k = 0, 1, 2, \ldots$

(3.22)
   (i) x_Q^k, x_N^k, x^{k+1} are well defined,

   (ii) $|x_Q^k - x^*| \leq r |x^k - x^*|,$

   (iii) $|x^{k+1} - x^*| \leq r_1 |x^k - x^*|,$

   (iv) $|x^{k+1} - x^*| \leq r_1^{k+1} \varepsilon,$

   (v) $\|E_{k+1} - E^*\| \leq \delta + c_5 \varepsilon^s \sum_{j=0}^{k} r_1^{sj},$

   (vi) $\|E_{k+j} - E^*\| \leq \|E_k - E^*\|$
$$+ c_5 |x^k - x^*|^s \sum_{l=0}^{j-1} r_1^{sl} \quad \text{for all } j = 0, 1, 2, \ldots.$$

We prove (i)–(v) by induction on k.

By (3.18) and (3.19), $\|E_0 - E^*\| \leq \delta$ implies that $E_0 \in \mathscr{N}$, and, by (3.20), $x^0 \in \Omega$. Therefore, by Theorem 3.1, x_Q^0 is well defined and satisfies

(3.23)
$$|x_Q^0 - x^*| \leq r |x^0 - x^*|.$$

On the other hand, by (3.21), x_N^0 is well defined, and

$$|x_N^0 - x^*| \leq \frac{r_1 - r}{2} |x^0 - x^*|.$$

Thus, by (2.4) and (3.23),

$$|x^1 - x^*| \leq |x^1 - x_N^0| + |x_N^0 - x^*|$$
$$\leq |x_Q^0 - x_N^0| + |x_N^0 - x^*| \leq |x_Q^0 - x^*| + 2|x_N^0 - x^*|$$
$$\leq r |x^0 - x^*| + (r_1 - r)|x^0 - x^*| = r_1 |x^0 - x^*| \leq r_1 \varepsilon.$$

Therefore, by Corollary 3.1,

$$\|E_1 - E^*\| \leq \|E_0 - E^*\| + c_5 |x^0 - x^*|^s \leq \delta + c_5 \varepsilon^s.$$

Hence, the thesis is true for $k = 0$.

Assume now the inductive hypothesis for $k - 1$. Thus,

$$\|E_k - E^*\| \leq \delta + c_5 \varepsilon^s \sum_{j=0}^{k-1} r_1^{sj}$$

$$\leq \delta + c_5 \varepsilon^s \sum_{j=0}^{\infty} r_1^{sj} \leq \delta + \frac{c_5 \varepsilon^s}{1 - r_1^s} < \delta_1.$$

So, by (3.18), $E_k \in \mathcal{N}$.

But, by the inductive hypotheses, $|x^k - x^*| \leq r_1^k \varepsilon \leq \varepsilon$, so $x^k \in \Omega$.

Hence, by Theorem 3.1, x_Q^k is well defined, and

$$|x_Q^k - x^*| \leq r|x^k - x^*|.$$

Finally, we deduce that

$$|x^{k+1} - x^*| \leq r_1 |x^k - x^*| \leq r_1^{k+1} \varepsilon$$

and

$$\|E_{k+1} - E^*\| \leq \delta + c_5 \varepsilon^s \sum_{j=0}^{k} r_1^{sj},$$

using the same arguments as in the case $k = 0$. Therefore, (i)–(v) are proved for all $k = 0, 1, 2, \ldots$.

Let us prove (vi) by induction on j. For $j = 0$, the result is trivial, using the convention $\sum_{l=0}^{-1} = 0$.

Suppose now the assertion is true for $j - 1$.

By (3.16), we have

$$\|E_{k+j} - E^*\| \leq \|E_{k+j-1} - E^*\| + c_5 |x^{k+j-1} - x^*|^s.$$

Hence, using the inductive hypotheses,

$$(3.24) \quad \|E_{k+j} - E^*\| \leq \|E_k - E^*\| + c_5 |x^k - x^*|^s \sum_{l=0}^{j-2} r_1^{sl} + c_5 |x^{k+j-1} - x^*|^s.$$

Now, by (iii),

$$|x^{k+j-1} - x^*|^s \leq r_1^{s(j-1)} |x^k - x^*|^s.$$

Thus, (vi) follows from (3.24).

Therefore, by (vi),

$$\|E_{k+j} - E^*\| \leq \|E_k - E^*\| + \frac{c_5 |x^k - x^*|^s}{1 - r_1^s},$$

and the last inequality in the assertion follows with $c_6 = c_5/(1 - r_1^s)$. $\square$

Corollary 3.2. *There exists $c_7 > 0$ such that, for all $k, j = 0, 1, 2, \ldots$*

$$(3.25) \qquad \|E_{k+j} - E^*\|^2 \leq \|E_k - E^*\|^2 + c_7 |x^k - x^*|^s.$$

Proof. Trivial, using Theorem 3.2 and the boundedness of $\|E_k - E^*\|$ and $|x^k - x^*|$. □

Theorem 3.3. *Under the assumptions of Theorem 3.2,*

$$\lim_{k \to \infty} \|E_{k+1} - E_k\| = 0.$$

Proof. Suppose the assertion is not true. Then there exists an infinite set K_1 of indices such that

$$\|E_{k+1} - E_k\| \geq \gamma > 0$$

for all $k \in K_1$. Hence, by (3.10),

$$[1 + c_1 \sigma(x^k, x_Q^k)^q]\|E_{k+1} - E_k\|_k \geq \gamma$$

for $k \in K_1$. Thus, for k large enough and $k \in K_1$,

$$\|E_{k+1} - E_k\|_k \geq \gamma/2.$$

So,

$$(3.26) \qquad \|E_{k+1} - E_k\|_k^2 \geq \gamma^2/4$$

for k belonging to an infinite set K_2 of indices.

Let $\hat{E}$ and $\hat{E}_k$ be the projections of E^* on $V(x^k, x_Q^k)$ related to the norms $\|\cdot\|$ and $\|\cdot\|_k$, respectively.

By (3.8), and Theorem 3.2, we have

$$\|\hat{E} - E^*\| \leq c_2|x^k - x^*|^p.$$

Thus, by (3.9),

$$\|\hat{E} - E^*\|_k \leq (1 + c_1|x^k - x^*|^q)c_2|x^k - x^*|^p$$
$$\leq (1 + c_1\varepsilon^q)c_2|x^k - x^*|^p = c_8|x^k - x^*|^p$$

with $c_8 = (1 + c_1\varepsilon^q)c_2$.

Therefore, by the definition of $\hat{E}_k$,

$$(3.27) \qquad \|\hat{E}_k - E^*\|_k \leq c_8|x^k - x^*|^p,$$

and, consequently,

$$(3.28) \qquad \|\hat{E}_k - E^*\|_k^2 \leq c_8^2|x^k - x^*|^{2p}.$$

Let $k \in K_2$. By (3.28) and the Pythagorean Theorem we have

$$\|E_{k+1} - E^*\|_k^2 = \|E_{k+1} - \hat{E}_k\|_k^2 + \|\hat{E}_k - E^*\|_k^2$$
$$\leq \|E_{k+1} - \hat{E}_k\|_k^2 + c_8^2|x^k - x^*|^{2p}$$
$$= \|E_k - \hat{E}_k\|_k^2 - \|E_{k+1} - E_k\|_k^2 + c_8^2|x^k - x^*|^{2p}.$$

So, by (3.26), (3.27), and (3.9),

$$\|E_{k+1} - E^*\|_k^2 \leq \|E_k - \hat{E}_k\|_k^2 - \frac{\gamma^2}{4} + c_8^2 |x^k - x^*|^{2p}$$

$$\leq (\|E_k - E^*\|_k + \|E^* - \hat{E}_k\|_k)^2 - \frac{\gamma^2}{4} + c_8^2 |x^k - x^*|^{2p}$$

$$\leq (\|E_k - E^*\|_k + c_8 |x^k - x^*|^p)^2 - \frac{\gamma^2}{4} + c_8^2 |x^k - x^*|^{2p}$$

$$= \|E_k - E^*\|_k^2 + 2c_8 \|E_k - E^*\|_k |x^k - x^*|^p + 2c_8^2 |x^k - x^*|^{2p} - \frac{\gamma^2}{4}$$

$$\leq \|E_k - E^*\|_k^2 + 2c_8 (1 + c_1 |x^k - x^*|^q) \|E_k - E^*\| |x^k - x^*|^p$$

$$+ 2c_8^2 |x^k - x^*|^{2p} - \frac{\gamma^2}{4}$$

$$\leq \|E_k - E^*\|_k^2 + c_9 |x^k - x^*|^p - \frac{\gamma^2}{4}$$

with $c_9 = 2c_8 (1 + c_1 \varepsilon^q) \delta_1 + 2c_8^2 \varepsilon^p$.

Hence, there exists $\overline{k}$ such that, for $k \in K_2$ and $k \geq \overline{k}$,

$$\|E_{k+1} - E^*\|_k^2 \leq \|E_k - E^*\|_k^2 - \frac{\gamma^2}{8}.$$

So, by (3.9), (3.10), and Theorem 3.2, we have, for large enough $\overline{\overline{k}}$,

$$\|E_{k+1} - E^*\|^2 \leq (1 + c_1 |x^k - x^*|^q)^2 \|E_{k+1} - E^*\|_k^2$$

$$\leq (1 + c_1 |x^k - x^*|^q)^2 \left(\|E_{k+1} - E^*\|_k^2 - \frac{\gamma^2}{8} \right)$$

(3.29)

$$\leq (1 + c_1 |x^k - x^*|^q)^2 \left[(1 + c_1 |x^k - x^*|^q)^2 \|E_k - E^*\|^2 - \frac{\gamma^2}{8} \right]$$

$$\leq \|E_k - E^*\|^2 - \frac{\gamma^2}{16}$$

for $k \in K_2$, $k \geq \overline{\overline{k}}$.

Let now $k_0 \geq \overline{\overline{k}}$ be such that, for all $k \geq k_0$,

(3.30) $$c_7 |x^k - x^*|^s \leq \frac{\gamma^2}{32}.$$

Define

$$K_3 = \{k \in K_2 | k \geq k_0\} = \{k_1, k_2, k_3, \dots\}, \qquad k_i < k_{i+1}, \quad i = 1, 2, 3, \dots.$$

Then, for all $j = 1, 2, 3, \dots$, we have, by (3.29),

(3.31) $$\|E_{k_j+1} - E^*\|^2 \leq \|E_{k_j} - E^*\|^2 - \frac{\gamma^2}{16}.$$

Now, by (3.25), (3.30), and (3.31),

$$(3.32) \quad \|E_{k_{j+1}} - E^*\|^2 \leq \|E_{k_j+1} - E^*\|^2 + c_7|x^k - x^*|^s$$

$$\leq \|E_{k_j} - E^*\|^2 - \frac{\gamma^2}{16} + \frac{\gamma^2}{32} = \|E_{k_j} - E^*\|^2 - \frac{\gamma^2}{32}.$$

But (3.32) holds for all $j = 1, 2, 3, \ldots$. Hence,

$$(3.33) \quad \|E_{k_j} - E^*\|^2 \leq \|E_{k_1} - E^*\|^2 - (j-1)\frac{\gamma^2}{32}.$$

But (3.33) implies that $\|E_{k_j} - E^*\|^2 < 0$ for large enough j, which is a contradiction. $\square$

Corollary 3.3. *There holds*

$$(3.34) \quad \lim_{k \to \infty} |\varphi(x^{k+1}, E_{k+1}) - \varphi(x^k, E_k)| = 0.$$

Proof. (3.34) follows from Theorem 3.3 and the uniform continuity of φ on any compact neighborhood of (x^*, E^*). $\square$

The following theorem states a Dennis-Moré-type condition (see [10]) for the ideal convergence of a linearly convergent sequence generated by Algorithm 2.1.

Theorem 3.4. *Let F satisfy Assumption 1 and let φ satisfy Assumption 2. Assume that $\{x^k\}$, generated by Algorithm 2.1, is well defined, and that, for some $r \in (r^*, 1)$, we have*

$$|x_Q^k - x^*| \leq r|x^k - x^*|$$

for $k = 0, 1, 2, \ldots$. Assume that

$$(3.35) \quad \lim_{k \to \infty} \frac{|[\varphi(x^k, E_k) - \varphi(x^*, E^*)](x_Q^k - x^k)|}{|x_Q^k - x^k|} = 0.$$

Then,

$$(3.36) \quad \limsup \frac{|x_Q^k - x^*|}{|x^k - x^*|} \leq r^*$$

and

$$(3.37) \quad \limsup \frac{|x^{k+1} - x^*|}{|x^k - x^*|} \leq r^*.$$

Proof. We write $B_k = \varphi(x^k, E_k)$, $B^* = \varphi(x^*, E^*)$. By (3.35), we have

$$(3.38) \quad \lim_{k \to \infty} \frac{|(I - (B^*)^{-1}B_k)(x_Q^k - x^k)|}{|x_Q^k - x^k|} = 0.$$

And, by (3.2),

$$(3.39) \quad |F(x^k) - J(x^*)(x^k - x^*)| \leq M|x^k - x^*|^{p+1}.$$

Hence, by (3.3), (3.38), and (3.39), we have

$$
\begin{aligned}
\frac{|x_Q^k - x^*|}{|x^k - x^*|} &= \frac{|x^k - B_k^{-1}F(x^k) - x^*|}{|x^k - x^*|} \\
&\le \frac{|x^k - x^* - (B^*)^{-1}F(x^k)|}{|x^k - x^*|} + \frac{|[(B^*)^{-1} - B_k^{-1}]F(x^k)|}{|x^k - x^*|}
\end{aligned}
$$

(3.40)
$$
\begin{aligned}
&\le \frac{|(I - (B^*)^{-1}J(x^*))(x^k - x^*)|}{|x^k - x^*|} + M|(B^*)^{-1}||x^k - x^*|^p \\
&\quad + \frac{|[(B^*)^{-1} - B_k^{-1}]B_k(x_Q^k - x^k)|}{|x^k - x^*|} \\
&\le r^* + M|(B^*)^{-1}||x^k - x^*|^p + \frac{|(I - (B^*)^{-1}B_k)(x_Q^k - x^k)|}{|x^k - x^*|}.
\end{aligned}
$$

Now, by (3.22),

(3.41)
$$
\begin{aligned}
\frac{|(I - (B^*)^{-1}B_k)(x_Q^k - x^k)|}{|x^k - x^*|} &= \frac{|(I - (B^*)^{-1}B_k)(x_Q^k - x^k)|}{|x_Q^k - x^k|} \frac{|x_Q^k - x^k|}{|x^k - x^*|} \\
&\le (1 + r)\frac{|(I - (B^*)^{-1}B_k)(x_Q^k - x^k)|}{|x_Q^k - x^k|}.
\end{aligned}
$$

Let $r_2 \in (r^*, r_1)$ and $r_3 \in (r^*, r_2)$. By (3.38), (3.40), and (3.41), we find that there exists $\bar{k}_0 \in \mathbf{N}$ such that, for all $k \ge \bar{k}_0$,

(3.42)
$$
|x_Q^k - x^*| \le r_3|x^k - x^*|.
$$

Let $k_0 \ge \bar{k}_0$ be such that, for $k \ge k_0$, by the local convergence results of Newton's method,

(3.43)
$$
|x_N^k - x^*| \le \frac{r_2 - r_3}{2}|x^k - x^*|.
$$

Then, by (3.42) and (3.43), we have, for $k \ge k_0$,

$$
\begin{aligned}
|x^{k+1} - x^*| &\le |x^{k+1} - x_N^k| + |x_N^k - x^*| \le |x_Q^k - x_N^k| + |x_N^k - x^*| \\
&\le |x_Q^k - x^*| + 2|x_N^k - x^*| \le r_2|x^k - x^*|.
\end{aligned}
$$

So, the desired result is proved. $\square$

Corollary 3.4. *Under the hypotheses of Theorem 3.4, if $r^* = 0$, we have*

$$
\lim \frac{|x_Q^k - x^*|}{|x^k - x^*|} = \lim \frac{|x^{k+1} - x^*|}{|x^k - x^*|} = 0.
$$

That is, the sequence converges at a q-superlinear rate.

Let us now introduce a final assumption in our framework.

Assumption 5 represents the fundamental property which defines a secant method. It states that the algorithm is making an effort to approximate an "ideal" iteration, and not merely trying to avoid excessive deterioration. This condition is usually achieved through the classical secant equation (see [13, Chapter 8]) if a Lipschitz condition is assumed on J.

Assumption 5. The sequence generated by Algorithm 2.1 satisfies

$$(3.44) \qquad \lim_{k \to \infty} \frac{|[\varphi(x^{k+1}, E_{k+1}) - \varphi(x^*, E^*)](x_Q^k - x^k)|}{|x_Q^k - x^k|} = 0.$$

Observe that, if the secant equation

$$\varphi(x^{k+1}, E_{k+1})(x_Q^k - x^k) = F(x_Q^k) - F(x^k)$$

holds, the statement (3.44) follows clearly from (3.2) and from the local linear convergence theorem. Therefore, in spite of its asymptotic form, (3.44) is easily verified for many practical algorithms.

Theorem 3.5. *Let $\{x^k\}$ be generated by Algorithm 2.1 and let F, φ, V, E^*, and $\{x^k\}$ satisfy Assumptions 1 to 5. Then there exist $\varepsilon, \delta > 0$ such that, if $|x^0 - x^*| \leq \varepsilon$ and $\|E_0 - E^*\| \leq \delta$, the sequence $\{x^k\}$ converges to x^*,*

$$\limsup \frac{|x_Q^k - x^*|}{|x^k - x^*|} \leq r^*,$$

and

$$\limsup \frac{|x^{k+1} - x^*|}{|x^k - x^*|} \leq r^*.$$

Proof. It is easy to prove that (3.34) and (3.44) imply (3.35). □

As a final result of this section, let us prove that the iterative refinement procedure may be used as an inner iteration for Algorithm 2.1.

Proposition 3.1. *If ε, δ are small enough, the iterative refinement procedure (2.8) produces points which satisfy (2.4).*

Proof. Set $z^* = -J(x^k)^{-1}F(x^k)$. The sequence $\{z_l\}$ in (2.8) satisfies

$$|z_{l+1} - z^*| \leq |I - \varphi(x^k, E_{k+1})^{-1}J(x^k)| |z_l - z^*|.$$

So, the desired result follows, using Assumption 2 and the continuity of φ. □

4. LOCAL CONVERGENCE OF SOME PARTICULAR ALGORITHMS

The theory presented in §3 is of wide applicability. Most quasi-Newton algorithms for nonlinear systems of equations, unconstrained minimization, and nonlinear least squares problems may be analyzed within this framework. Algorithms 2.2, 2.3, and 2.4 were selected as typical examples to illustrate the fact

that the abstractions used to define the model Algorithm 2.1 have a practical motivation. For example, X is not $\mathbf{R}^{n \times n}$ in Algorithms 2.2 and 2.3, φ depends explicitly on x in Algorithm 2.4, r^* may be different from 0 in Algorithm 2.2, etc.

We will now apply the results of §3 to the Algorithms 2.2, 2.3, and 2.4. We show that Algorithms 2.3 and 2.4 are locally and superlinearly convergent and that Algorithm 2.2 converges at the ideal rate.

Theorem 4.1. *In addition to hypotheses* (a)–(b) *of* §2.2, *assume that F satisfies Assumption* 1 *and*

$$(4.1) \qquad |I - N(x^*)^{-1} J(x^*)| \leq r^* < 1.$$

(*Observe that, according to* [35, Theorem 2.2.8] *assumption* (4.1) *holds for some suitable norm* $|\cdot|$ *if and only if* $I - N(x^*)^{-1} J(x^*)$ *has a spectral radius less than* 1.)

Assume, further, that, for all $x \in \Omega$,

$$(4.2) \qquad |\mathscr{A}(x) - \mathscr{A}(x^*)| \leq c_2'|x - x^*|^p,$$

$$(4.3) \qquad |\mathscr{R}(x) - \mathscr{R}(x^*)| \leq c_2'|x - x^*|^p.$$

Then there exist $\varepsilon, \delta > 0$ *such that, if* $|x^0 - x^*| \leq \varepsilon$, $|A_0 - \mathscr{A}(x^*)| \leq \delta$, *and* $|R_0 - \mathscr{R}(x^*)| \leq \delta$, *the sequence defined by Algorithm* 2.2 *is well defined, converges to* x^*, *and satisfies*

$$\limsup \frac{|x^{k+1} - x^*|}{|x^k - x^*|} \leq r^*.$$

Proof. Define

$$E^* = (A_*, R_*) = (\mathscr{A}(x^*), \mathscr{R}(x^*)).$$

Then, Assumption 2 is a direct consequence of (4.1).

Since $(\mathscr{A}(x), \mathscr{R}(x)) \in V(x, z)$ for all $x, z \in \Omega$, Assumption 3 follows from (4.2) and (4.3). Assumption 4 is clearly satisfied.

Now, $\varphi(x^{k+1}, E_{k+1})(x_Q^k - x^k) = N(x^k)(x_Q^k - x^k)$ by the definition of $V(x, z)$. So, Assumption 5 follows from the continuity of N. Therefore, the proof follows by applying Theorem 3.5. □

Theorem 4.2. *Let F satisfy Assumption* 1. *Define $G_1, \ldots, G_m$ as in the beginning of* §2.3. *Assume that G_i satisfies a Lipschitz condition*

$$(4.4) \qquad |G_i(W_i x) - G_i(W_i x^*)| \leq L_i|x - x^*|$$

for all $x \in \Omega$, $i = 1, \ldots, m$.

Then, there exist $\varepsilon, \delta > 0$ *such that, if* $|x^0 - x^*| \leq \varepsilon$ *and* $|B_i^0 - G_i'(W_i x^*)| \leq \delta$, $i = 1, \ldots, m$, *the sequence defined by Algorithm* 2.3 *converges to* x^* *at a q-superlinear rate.*

Proof. Define

$$E^* = (B_1^*, \ldots, B_m^*) = (G_1'(W_1(x^*)), \ldots, G_m'(W_m(x^*))).$$

Then, Assumption 2 holds with $r^* = 0$. Clearly, Assumption 4 is trivially satisfied.

Now,

$$F_i(z) - F_i(x) = \left[\int_0^1 F_i'(x + t(z - x)) \, dt \right] (z - x)$$

$$= U_i \left\{ \int_0^1 G_i'[W_i(x + t(z - x))] \, dt \right\} W_i(z - x).$$

Hence,

$$\tilde{E} = \left(\int_0^1 G_1'[W_1(x + t(z - x))] \, dt, \ldots, \int_0^1 G_m'[W_m(x + t(z - x))] \, dt \right)$$

$$\in V(x, z).$$

Thus, Assumption 3 follows easily using (4.4).

Finally, the definition of B_i^{k+1} guarantees that $\varphi(x^{k+1}, E_{k+1})(x^{k+1} - x^k) = F(x^{k+1}) - F(x^k)$, so that Assumption 5 is also satisfied and thus, the algorithm is superlinearly convergent. $\square$

Theorem 4.3. *Assume that F is given by (2.9)–(2.10) and satisfies Assumption 1. Moreover, assume that $J(x^*)$ and $R'(x^*)^T R'(x^*)$ are positive definite and*

$$(4.5) \qquad |R'(x)^T R'(x) - R'(x^*)^T R'(x^*)| \le M'|x - x^*|$$

for all $x \in \Omega$.

Define

$$E^* = J(x^*)^{-1} - [R'(x^*)^T R'(x^*)]^{-1}.$$

Then, there exist $\varepsilon, \delta > 0$ such that, if $|x^0 - x^*| \le \varepsilon$ and $|E_0 - E^*| \le \delta$, the sequence defined by Algorithm 2.4 converges to x^* at a q-superlinear rate.

Proof. By the definition of E^* it is clear that Assumption 2 is satisfied with $r^* = 0$.

We define $\|E\| = \|L_*^T E L_*\|_F$, where $L_* L_*^T$ is the Cholesky factorization of $J(x^*)$.

Now, examining the steps of the classical algorithm for the Cholesky factorization, we verify that each coefficient of L is a C^1-function of the coefficients of the factored matrix. This observation, together with (3.1), leads to

$$|L(x, z) - L^*| \le K\sigma(x, z)^p,$$

and, thus, to Assumption 4 with $q = p$.

Define

$$\tilde{E} = \int_0^1 [J(x + t(z - x)) \, dt]^{-1} - [R'(z)^T R'(z)]^{-1}.$$

Obviously, $\tilde{E} \in V(x, z)$. Now, using (3.1), (4.5), and Banach's Lemma [17, p. 28], we verify that

$$(4.6) \qquad |\tilde{E} - E^*| \le c_2' \sigma(x, z)^p.$$

Hence, Assumption 3 follows from (4.6) using the equivalence of norms in finite-dimensional linear spaces.

According to the definition of φ and $V(x, z)$, we have

$$\varphi(x^{k+1}, E_{k+1})(x^{k+1} - x^k) = \{[R'(x^{k+1})^T R'(x^{k+1})]^{-1} + E_{k+1}\}^{-1}(x^{k+1} - x^k).$$

But

$$[(R'(x^{k+1})^T R'(x^{k+1}))^{-1} + E_{k+1}][F(x^{k+1}) - F(x^k)] = x^{k+1} - x^k,$$

therefore

$$\varphi(x^{k+1}, E_{k+1})(x^{k+1} - x^k) = F(x^{k+1}) - F(x^k),$$

and so Assumption 5 follows using (3.2). □

5. FINAL REMARKS

In this paper we presented a large family of quasi-Newton methods for solving systems of algebraic nonlinear equations, as well as an inexact-Newton extension of this family. The convergence proofs in §3 apply to most known quasi-Newton methods in the literature. It is not difficult to recognize these methods as members of the family, using, in some cases, their identification with least change secant update methods.

We make one assumption on the function and three assumptions on the algorithms to guarantee local linear convergence to an isolated solution. According to our approach, projections take place in a "parameter space" X, which may be different from $\mathbf{R}^{n \times n}$. This allows us to cover a broader class of methods than the theory of Dennis-Walker [14] does. Clearly, our Assumption 2 corresponds to Assumption 3.5 of [14], and our Assumption 3 plays the role of Assumption 3.6 of the Dennis-Walker paper. However, our Assumption 3 is not associated with "choice rules" or secant type equations. This fact has a theoretical and pedagogical significance, since it shows that linear convergence may not be related to the secant approach in potentially useful methods. Our Assumption 4 states the relation between the norm used for variable projections of the iteration parameters E_k at each stage of the calculation. No assumption with this generality is considered in the Dennis-Walker theory. In fact, variable norms in [14] are always weighted Frobenius norms defined by a matrix which satisfies a secant equation. We do not know if there exist practical methods where the norms $\| \cdot \|_k$ are defined in a different way, but we feel that stressing the independence of the essential properties of $\| \cdot \|_k$ from the secant approach has some theoretical interest.

In our theory, the secant equation appears only in Assumption 5. Assumptions 1 to 4 are enough to prove not only the linear convergence result, but also the key Theorem 3.3. Using this theorem, Assumption 5 and linear convergence, we prove convergence at the ideal rate r^* (superlinear if $r^* = 0$).

We think that the main message of the new theory is also implicit in the Dennis-Walker theory and in other works on quasi-Newton methods. This is:

given a particular class of problems, put all the desirable characteristics of the ideal parameter E^* in the manifolds $V(x, z)$. The resulting method is probably locally linearly convergent and, if some secant type equation is also satisfied, ideal convergence may be obtained.

Finally, as in the Dennis-Walker theory, our convergence results are strictly local, not only with respect to x but also with respect to E. The possibility of relaxing the condition on E_0 and x^0 through line searches or trust regions (see [13, pp. 139–143]) deserves further research.

BIBLIOGRAPHY

1. Å. Björck, *Least squares methods*, Handbook of Numerical Analysis, vol 1 (P. G. Ciarlet and J. L. Lions, eds.), Elsevier, North-Holland, 1987.

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

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

4. C. G. Broyden, J. E. Dennis, and J. J. Moré, *On the local and superlinear convergence of quasi-Newton methods*, J. Inst. Math. Appl. **12** (1973), 223–245.

5. F. F. Chadee, *Sparse quasi-Newton methods and the continuation problem*, T. R. SOL 85-8, Dept. of Operations Research, Stanford University, Stanford, CA, 1985.

6. W. C. Davidon, *Variable metric method for minimization*, Report ANL-5990 Rev (1959), Argonne National Laboratory, Argonne, Ill., 1959.

7. R. S. Dembo, S. C. Eisenstat and T. Steihaug, *Inexact Newton methods*, SIAM J. Numer. Anal. **19** (1982), 400–408.

8. J. E. Dennis, Jr., D. M. Gay, and R. E. Welsch, *An adaptive nonlinear least-squares algorithm*, ACM Trans. Math. Software **7** (1981), 348–368.

9. J. E. Dennis, Jr. and E. S. Marwil, *Direct secant updates of matrix factorizations*, Math. Comp. **38** (1982), 459–476.

10. J. E. Dennis, Jr. and J. J. Moré, *A characterization of superlinear convergence and its application to quasi-Newton methods*, Math. Comp. **28** (1974), 549–560.

11. _____, *Quasi-Newton methods, motivation and theory*, SIAM Rev. **19** (1977), 46–89.

12. J. E. Dennis, Jr. and R. B. Schnabel, *Least change secant updates for quasi-Newton methods*, SIAM Rev. **21** (1979), 443–459.

13. _____, *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice-Hall Series in Comput. Math., Prentice-Hall, N.J., 1983.

14. J. E. Dennis, Jr. and H. F. Walker, *Convergence theorems for least-change secant update methods*, SIAM J. Numer. Anal. **18** (1981), 949–987.

15. I. S. Duff, A. M. Erisman, and J. K. Reid, *Direct methods for sparse matrices*, Clarendon Press, Oxford, 1986.

16. P. E. Gill, G. H. Golub, W. Murray, and M. A. Saunders, *Methods for modifying matrix factorizations*, Math. Comp. **28** (1974), 505–535.

17. G. H. Golub and Ch. F. Van Loan, *Matrix computations*, The Johns Hopkins University Press, Baltimore, 1983.

18. A. Griewank, *The local convergence of Broyden-like methods on Lipschitzian problems in Hilbert spaces*, SIAM J. Numer. Anal. **24** (1987), 684–705.

19. A. Griewank and Ph. L. Toint, *On the unconstrained optimization of partially separable functions*, Nonlinear Optimization, 1981 (M. J. D. Powell, ed.), Academic Press, New York, 1982.

20. _____, *Partitioned variable metric for large structured optimization problems*, Numer. Math. **39** (1982), 119–137.

21. _____, *Local convergence analysis for partitioned quasi-Newton updates*, Numer. Math. **39** (1982), 429–448.

22. _____, *Numerical experiments with partially separable optimization problems*, Numerical Analysis (Proceedings Dundee 1983) (D. F. Griffiths, ed.), Lecture Notes in Math., vol. 1066, Springer-Verlag, Berlin, 1984, pp. 203–220.

23. M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Nat. Bur. Standards B **49** (1952), 409–436.

24. G. W. Johnson and N. H. Austria, *A quasi-Newton method employing direct secant updates of matrix factorizations*, SIAM J. Numer. Anal. **20** (1983), 315–325.

25. C. T. Kelley and E. W. Sachs, *Broyden's method for approximate solution of nonlinear integral equations*, J. Integral Equations **9** (1985), 25–43.

26. K. Levenberg, *A method for the solution of certain nonlinear problems in least squares*, Quart. J. Appl. Math. **2** (1944), 164–168.

27. D. G. Luenberger, *Linear and nonlinear programming*, 2nd ed., Addison-Wesley, Reading, Mass., 1984.

28. D. W. Marquardt, *An algorithm for least-squares estimation of nonlinear parameters*, SIAM J. Appl. Math. **11** (1963), 431–441.

29. J. M. Martínez, *A Quasi-Newton method with a new updating for the LDU factorization of the approximate Jacobian*, Mat. Apl. Comput. **2** (1983), 131–142.

30. _____, *Quasi-Newton methods with factorization scaling for solving sparse nonlinear systems of equations*, Computing **38** (1987), 131–141.

31. _____, *A family of Quasi-Newton methods with direct secant updates of matrix factorizations*, SIAM J. Numer. Anal. (to appear).

32. E. S. Marwil, *Exploiting sparsity in Newton-type methods*, Cornell Applied Mathematics Ph.D. thesis, 1978.

33. _____, *Convergence results for Schubert's method for solving sparse nonlinear equations*, SIAM J. Numer. Anal. **16** (1979), 588–604.

34. J. J. Moré and J. A. Trangenstein, *On the global convergence of Broyden's method*, Math. Comp. **30** (1976), 523–540.

35. J. M. Ortega and W. C. Rheinboldt, *Iterative solution of nonlinear equations in several variables*, Academic Press, New York, 1970.

36. O. Østerby and Z. Zlatev, *Direct methods for sparse matrices*, Lecture Notes in Comput. Sci., vol. 157, Springer-Verlag, Berlin, Heidelberg, New York, and Tokyo, 1983.

37. M. J. D. Powell, *A new algorithm for unconstrained optimization*, Nonlinear Programming (J. B. Rosen, O. L. Mangasarian, and K. Ritter, eds.), Academic Press, New York, 1970, pp. 31–65.

38. E. Sachs, *Broyden's method in Hilbert space*, Math. Programming **35** (1986), 71–82.

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

40. H. Schwetlick, *Numerische Lösung nichtlinearer Gleichungen*, Deutscher Verlag der Wissenschaften, Berlin, 1978.

41. Ph. L. Toint, *On sparse and symmetric matrix updating subject to a linear equation*, Math. Comp. **13** (1977), 954–961.

42. Ph. L. Toint, *Some numerical results using a sparse matrix updating formula in unconstrained optimization*, Math. Comp. **32** (1978), 839–851.

43. _____, *Numerical solution of large sets of algebraic nonlinear equations*, Math. Comp. **46** (1986), 175–189.

APPLIED MATHEMATICS LABORATORY, IMECC–UNICAMP, CAIXA POSTAL 6065, 13.081–CAMPINAS–SP, BRASIL