Algorithms for Solving Nonlinear Systems of Equations[(*)]

José Mario Martínez

Department of Applied Mathematics
IMECC-UNICAMP, CP 6065,
University of Campinas
13081 Campinas SP
Brazil

ABSTRACT
In this paper we survey numerical methods for solving nonlinear systems of equations $F(x) = 0$, where $F : I\!R^n \to I\!R^n$. We are especially interested in large problems. We describe modern implementations of the main local algorithms, as well as their globally convergent counterparts.

1. INTRODUCTION

Nonlinear systems of equations appear in many real - life problems. Moré [1989] has reported a collection of practical examples which include: Aircraft Stability problems, Inverse Elastic Rod problems, Equations of Radiative Transfer, Elliptic Boundary Value problems, etc.. We have also worked with Power Flow problems, Distribution of Water on a Pipeline, Discretization of Evolution problems using Implicit Schemes, Chemical Plant Equilibrium problems, and others. The scope of applications becomes even greater if we include the family of Nonlinear Programming problems, since the first-order optimality conditions of these problems are nonlinear systems.

Given $F : I\!R^n \to I\!R^n, F = (f_1, \ldots, f_n)^T$, our aim is to find solutions of

[0(*)] Work supported by FAPESP (PT N$^{\underline{o}}$ 90-3724-6) FAEP-UNICAMP, FINEP and CNPq. This paper appeared in *Continuous Optimization: the state or art*, edited by E. Spedicato, Kluwer, 81-108, 1994.

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

We assume that F is well defined and has continuous partial derivatives on an open set of $I\!R^n$. We denote $J(x)$ the matrix of partial derivatives of F (Jacobian matrix). So,

$$J(x) \equiv F'(x) \equiv \begin{bmatrix} f_1'(x) \\ \vdots \\ f_n'(x) \end{bmatrix} \equiv \begin{bmatrix} \nabla f_1(x)^T \\ \vdots \\ \nabla f_n(x)^T \end{bmatrix} \equiv \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(x) & \dots & \frac{\partial f_1}{\partial x_n}(x) \\ \vdots & & \vdots \\ \frac{\partial f_n}{\partial x_1}(x) & \dots & \frac{\partial f_n}{\partial x_n}(x) \end{bmatrix}.$$

We are mostly interested in problems where n is large and $J(x)$ is *structurally sparse*. This means that most entries of $J(x)$ are zero *for all x in the domain of F*. Sparsity is a particular case of the more general notion of *structure*. Jacobian matrices can be symmetric, antisymmetric, positive definite, combination of other matrices with some particular structure, etc.. Many times we can take advantage of particular structures of $J(x)$ in order to obtain efficient algorithms for solving (1.1).

Most popular methods for solving nonlinear systems are *local*. A local method is an iterative scheme that converges if the initial approximation is close enough to a particular solution. Frequently, we are also able to prove *rate of convergence* results for these methods, which tell something about the asymptotic velocity of convergence of the process. Fortunately, in many practical cases the domain of convergence of local methods is large, so that these methods are useful. However, when the initial estimate of the solution is very poor, local methods must be modified in order to improve their *global convergence properties*.

We say that a method for solving (1.1) is *globally convergent* if at least one limit point of the sequence generated by the method is a solution or, at least, a *stationary point*, where $\nabla \, ||F(x)||^2 = 0$. Many times *all* the limit points are solutions or stationary points and, frequently, the whole sequence converges to a solution. In general, global methods are modifications of local methods that preserve the local convergence properties of the original algorithm.

This paper is organized as follows. In Section 2, 3, 4 and 5 we survey local methods, namely: Newton's method, Quasi-Newton methods, Inexact-Newton methods and Decomposition methods. In Sections 6 and 7 we survey the main procedures currently used for globalization: Optimization and Homotopies. Conclusions are given in Section 8.

2. NEWTON'S METHOD

Newton's method is the most widely used algorithm for solving nonlinear systems of equations. Given an initial estimation x^0 of the solution of (1.1), this method considers, at each iteration, the approximation

$$F(x) \approx L_k(x) \equiv F(x^k) + J(x^k)(x - x^k) \tag{2.1}$$

and computes x^{k+1} as a solution of the linear system $L_k(x) = 0$. This solution exists and is unique if $J(x^k)$ is nonsingular. Therefore, an iteration of Newton's method is described by

$$J(x^k)s^k = -F(x^k), \tag{2.2}$$

$$x^{k+1} = x^k + s^k. \tag{2.3}$$

At each iteration of Newton's method, we must compute the Jacobian $J(x^k)$ and solve the linear system (2.2). Using modern techniques of automatic differentiation (see Rall [1984, 1987], Griewank [1992], and references therein) we can compute $F(x)$ and $J(x)$ in a reliable and economical way. If, instead of the true Jacobian in (2.2), we use an approximation by differences of $J(x^k)$, which is generally expensive, we obtain the *Finite-Difference Newton's Method*, whose convergence properties are very similar to those of Newton's method.

Now, (2.2) is a linear system of equations. If n is small, this system can be solved using the LU factorization with partial pivoting or the QR factorization. See Golub and Van Loan [1989]. Using these linear solvers, the cost of solving (2.2) is $O(n^3)$ floating point operations. If n is large this cost becomes prohibitive. However, in many situations, where the matrix $J(x^k)$ is sparse, we can solve (2.2) using LU factorizations. In fact, many times the structure of the matrix is such that the factors L and U of its factorization are also sparse, and can be computed using a moderate amount of operations. Computer algorithms for sparse LU factorizations are surveyed in Duff, Erisman and Reid [1989]. In Gomes-Ruggiero, Martínez and Moretti [1992] we describe the first version of the NIGHTINGALE package for solving sparse nonlinear systems. In NIGHTINGALE, we use the sparse linear solver of George and Ng [1987]. The George-Ng method performs the LU factorization with partial pivoting of a sparse matrix A using a *static data structure* defined before beginning numerical computations. In Newton's method we solve a sequence of linear systems with the same structure, so, the *symbolic phase* that defines the data structure is executed only once.

The system (2.2) has a unique solution if and only if $J(x^k)$ is nonsingular. If the Jacobian is singular, the iteration must be modified. Moreover, if $J(x^k)$ is nearly singular, it is also convenient to modify the iteration in order to prevent numerical instability. Many modifications are possible to keep this phenomenon controlled. In the NIGHTINGALE package, when a very small pivot, relative to the size of the matrix, occurs, it is replaced by a nonzero scalar whose modulus is sufficiently large. Moreover, nearly singular or ill-conditioned matrices usually cause very large increments s^k. So, $||s^k||$ must also be controlled. Computer algorithms usually normalize the stepsize by

$$s^k \leftarrow \min\{1, \frac{\Delta}{||s^k||}\}s^k.$$

where Δ is a parameter given by the user.
The main convergence result relative to Newton's method is given by the following theorem.

Theorem 2.1. Let us assume that $F : \Omega \subset I\!R^n \to I\!R^n$, Ω an open and convex set, $F \in C^1(\Omega)$, $F(x^*) = 0$, $J(x^*)$ nonsingular, and that there exist $L, p > 0$ such that for all $x \in \Omega$

$$||J(x) - J(x^*)|| \leq L||x - x^*||^p. \tag{2.4}$$

Then there exists $\varepsilon > 0$ such that if $||x^0 - x^*|| \leq \varepsilon$, the sequence $\{x^k\}$ generated by (2.2) – (2.3) is well defined, converges to x^*, and satisfies

$$||x^{k+1} - x^*|| \leq c||x^k - x^*||^{p+1}. \tag{2.5}$$

Proof. See Ortega and Rheinboldt [1970], Dennis and Schnabel [1983], etc. □

Property (2.5) (quadratic convergence if $p = 1$) depends on the Hölder condition (2.4). If (2.4) is not assumed, only superlinear convergence of $\{x^k\}$ can be proved $\left(\lim_{k\to\infty} \frac{||x^{k+1} - x^*||}{||x^k - x^*||} = 0\right)$. Clearly, (2.5) implies superlinear convergence but the converse is not true.

3. QUASI-NEWTON METHODS

In this survey, we call Quasi-Newton methods those methods for solving (1.1) whose general form is

$$x^{k+1} = x^k - B_k^{-1}F(x^k). \tag{3.1}$$

Newton's method, studied in Section 2, belongs to this family. Most Quasi-Newton methods use less expensive iterations than Newton, but their convergence properties are not very different. In general, Quasi-Newton methods avoid either the necessity of computing derivatives, or the necessity of solving a full linear system per iteration or both tasks.

The most simple Quasi-Newton method is the *Stationary Newton Method*, where $B_k = J(x^0)$ for all $k \in I\!N$. In this method, derivatives are computed at the initial point and we only need the LU factorization of $J(x^0)$. A variation of this method is the *Stationary Newton Method with restarts,* where $B_k = J(x^k)$ if k is a multiple of a fixed integer m and $B_k = B_{k-1}$ otherwise. The number of iterations used by this method tends to increase with m, but the average computer time per iteration decreases. In some situations we can determine an optimal choice for m (Shamanskii [1967]).

An obvious drawback of the stationary Newton methods is that, except when $k \equiv 0 \pmod m$, B_k does not incorporate information about x^k and $F(x^k)$. Therefore, the adequacy of the model $L_k(x) \equiv F(x^k) + B_k(x - x^k)$ to the real function $F(x)$ can decrease rapidly as k grows. Observe that, due to (3.1), in Quasi-Newton methods x^{k+1} is defined as the solution of $L_k(x) = 0$, which exists and is unique if B_k is nonsingular. One way to incorporate new information about F on the linear model is to impose the interpolatory conditions

$$L_{k+1}(x^k) = F(x^k), \tag{3.2}$$
$$L_{k+1}(x^{k+1}) = F(x^{k+1}). \tag{3.3}$$

Defining

$$y^k = F(x^{k+1}) - F(x^k) \tag{3.4}$$

and substracting (3.2) from (3.3) we obtain the *Secant Equation*

$$B_{k+1} s^k = y^k. \tag{3.5}$$

Reciprocally, if B_{k+1} satisfies (3.5), L_{k+1} interpolates F at x^k and x^{k+1}. We give the name *Secant Methods* to the family of Quasi-Newton methods based on (3.1) and (3.5).

If $n \geq 2$, there exist infinite many possible choices of B_{k+1} satisfying (3.5). If, in addition to (3.5), we impose that

$$B_{k+1} s^{k-j} = y^{k-j}, j = 1, \ldots, n-1 \tag{3.6}$$

we obtain the Sequential Secant Method (Wolfe [1959], Barnes [1965], Gragg and Stewart [1976], Martínez [1979a], etc.). If the set of increments $\{s^k, s^{k-1}, \ldots, s^{k-n+1}\}$ is linearly independent there exists only one matrix B_{k+1} that satisfies (3.6). In this case,

$$B_{k+1} = (y^k, y^{k-1}, \ldots, y^{k-n+1})(s^k, s^{k-1}, \ldots, s^{k-n+1})^{-1} \tag{3.7}$$

and

$$B_{k+1}^{-1} = (s^k, s^{k-1}, \ldots, s^{k-n+1})(y^k, y^{k-1}, \ldots, y^{k-n+1})^{-1}. \tag{3.8}$$

B_{k+1}^{-1} can be obtained from B_k^{-1} using $O(n^2)$ floating point operations. However, in order to ensure numerical stability, the definition of the increments s^j that appear in (3.7) and (3.8) must sometimes be modified. When these modifications are not necessary, the Sequential Secant Method has the following interpolatory property:

$$L_{k+1}(x^{k+j}) = F(x^{k+j}),\ j = 0, -1, \ldots, -n. \tag{3.9}$$

The Sequential Secant Method is useful in many situations where n is small. When n is not very small, it is not worthwhile to waste time trying to preserve the interpolatory condition (3.9) for $j \approx -n$. It is more profitable to maintain the Secant Equation (3.5), using the degrees of freedom inherent to this equation to guarantee numerical stability. Broyden's "good" method (Broyden [1965]) and the Column Updating Method (COLUM) (Martínez [1984]) are two examples of this idea. In both methods

$$B_{k+1} = B_k + \frac{(y^k - B_k s^k)(z^k)^T}{(z^k)^T s^k} \tag{3.10}$$

where

$$z^k = s^k \tag{3.11}$$

for Broyden's method and

$$\begin{aligned} z^k &= e^{j_k}, &(3.12)\\ |(e^{j_k})^T s^k| &= ||s^k||_\infty &(3.13)\end{aligned}$$

for COLUM. $\{e^1, \ldots, e^n\}$ is the canonical basis of $I\!R^n$.

Applying the Sherman-Morrison formula to (3.10) (Golub and Van Loan [1989, p. 51]) we obtain

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s^k - B_k^{-1} y^k)(z^k)^T}{(z^k)^T B_k^{-1} y^k} B_k^{-1}. \tag{3.14}$$

Formula (3.14) shows that B_{k+1}^{-1} can be obtained from B_k^{-1} using $O(n^2)$ floating point operations in the dense case. Moreover,

$$B_{k+1}^{-1} = (I + u^k (z^k)^T) B_k^{-1}, \tag{3.15}$$

where $u^k = (s^k - B_k^{-1} y^k)/(z^k)^T B_k^{-1} y^k$, so

$$B_k^{-1} = (I + u^{k-1}(z^{k-1})^T) \ldots (I + u^0 (z^0)^T) B_0^{-1}, \tag{3.16}$$

for $k = 1, 2, 3 \ldots$

Formula (3.16) is used when n is large. In this case, the vectors $u^0, z^0, \ldots, u^{k-1}, z^{k-1}$ are stored and the product $B_k^{-1} F(x^k)$ is computed using (3.16). In this way, the computer time of iteration k is $O(kn)$ plus the computer time of computing $B_0^{-1} F(x^k)$. If k is large the process must be periodically restarted taking $B_k \approx J(x^k)$.

An intermediate method between Broyden's method and the Sequential Secant Method is *Broyden's Method with Projected Updates*, which was introduced by Gay and Schnabel [1978]. See also Martínez [1979b] and Lopes and Martínez [1980]. This method is probably more efficient than Broyden for small problems, but we have not heard about large-scale implementations.

Broyden's method is a particular case of the family of Least Change Secant Update (LCSU) methods (Dennis and Schnabel [1979, 1983], Dennis and Walker [1981], Martínez [1990b, 1992a]), which include many algorithms that are useful for particular structures (Hart-Soul algorithms for boundary value problems (Hart and Soul [1973], Kelley and Sachs [1987]), Partitioned Quasi-Newton methods for separable problems (Griewank and Toint [1982a, 1982b, 1982c, 1984], Toint [1986]), Methods of Direct Updating of Factorizations (Dennis and Marwil [1982], Johnson and Austria [1983], Chadee [1985], Martínez [1990a]), BFGS and DFP algorithms for unconstrained minimization (see Dennis and Schnabel [1983]), etc.

Let us survey the main convergence results related to Quasi-Newton algorithms. We are going to assume, as in Theorem 2.1, that $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$, Ω open and convex, $F \in C^1(\Omega)$, $F(x^*) = 0$, $J(x^*)$ nonsingular and that the Hölder condition (2.4) is satisfied. The first result is the "Theorem of two neighborhoods".

Theorem 3.1. Given $r \in (0,1)$, there exists $\varepsilon, \delta > 0$ such that if $||x^0 - x^*|| \le \varepsilon$ and $||B_k - J(x^*)|| \le \delta$ for all $k \in \mathbb{N}$ then the sequence $\{x^k\}$ generated by (3.1) is well defined, converges to x^*, and satisfies

$$||x^{k+1} - x^*|| \le r||x^k - x^*|| \tag{3.17}$$

for all $k \in \mathbb{N}$.

Proof. See, for instance, Dennis and Walker [1981]. □.

Using Theorem 3.1 we can prove that the Stationary Newton Method and its variations with restarts have local convergence at a linear rate. The main tool for proving superlinear convergence of Quasi-Newton methods, is the following theorem, due to Dennis and Moré.

Theorem 3.2. Assume that the sequence $\{x^k\}$ generated by (3.1) is well defined and convergent to x^*. Then, the two following properties are equivalent.

$$\text{(a)} \qquad \lim_{k\to\infty} \frac{||[B_k - J(x^*)](x^{k+1} - x^k)||}{||x^{k+1} - x^k||} = 0. \tag{3.18}$$

$$\text{(b)} \qquad \lim_{k\to\infty} \frac{||x^{k+1} - x^*||}{||x^k - x^*||} = 0. \tag{3.19}$$

Proof. See Dennis and Moré [1974]. □

(3.18) is called the Dennis-Moré condition. Using (3.18), we can prove that the Stationary Newton Method with periodic restarts (for which $\lim_{k\to\infty} B_k = J(x^*)$) is superlinearly convergent. The Dennis-Moré condition says that the effect of $B_k - J(x^*)$ on the normalized increment tends to be null when $k \to \infty$. This condition is weaker than saying that B_k tends to $J(x^*)$. The Dennis-Moré condition is closely related to the Secant Equation. In fact, from (2.4) we can deduce that, for all $x \in \Omega$,

$$||F(z) - F(x) - J(x^*)(z - x)|| \leq L||z - x|| \max\{||x - x^*||^p, ||z - x^*||^p\} \tag{3.20}$$

(Broyden, Dennis and Moré [1973]). So, writing $x = x^k,\ z = x^{k+1}$, and assuming that the sequence converges,

$$\lim_{k\to\infty} \frac{||y^k - J(x^*)s^k||}{||s^k||} = 0. \tag{3.21}$$

Therefore, if the Secant Equation (3.5) holds,

$$\lim_{k\to\infty} \frac{||[B_{k+1} - J(x^*)]s^k||}{||s^k||} = 0. \tag{3.22}$$

This allows us to prove the following result.

Lemma 3.3. If the sequence generated by a Secant Method converges to x^* and, in addition,

$$\lim_{k\to\infty} ||B_{k+1} - B_k|| = 0. \tag{3.23}$$

then the convergence is superlinear.

Proof. Use (3.18) and (3.22). □

Theorem 3.1 does not guarantee local convergence of all Secant Methods. In fact, the hypothesis of this theorem is that *all the* B_k*'s* belong to a neighborhood of $J(x^*)$ of radius δ. Observe that, even if the first B_0 belongs to this neighborhood it could be possible that $||B_k - J(x^*)|| \gg ||B_0 - J(x^*)||$, destroying convergence. Fortunately, for LCSU methods (including Broyden) we are able to prove that exists $\delta' > 0$ such that $||B_k - J(x^*)|| \leq \delta$ for all $k \in I\!N$, if $||B_0 - J(x^*)|| \leq \delta'$. This is a *Bounded Deterioration Property.* Moreover, the Successive Projection scheme that characterizes LCSU methods guarantees also (3.23). Summing up, the following result holds for Broyden, and other LCSU methods.

Theorem 3.4. There exists $\varepsilon, \delta > 0$ such that, if $||x^0 - x^*|| \leq \varepsilon$ and $||B_0 - J(x^*)|| \leq \delta$, the sequence generated by Broyden's method is well defined, converges to x^* and satisfies

(3.19).

Proof. See Broyden, Dennis and Moré [1973]. For an extension to "all" LCSU methods see Martínez [1990b, 1992a]. □

For Broyden's method, we also have the following result, which states that the convergence is 2n-quadratic.

Theorem 3.5. Under the hypotheses of Theorem 3.4, if $p \geq 1$, there exists $c > 0$ such that the sequence generated by Broyden's method satisfies

$$||x^{k+2n} - x^*|| \leq c||x^k - x^*||^2$$

for all $k \in I\!N$.
Proof. See Gay [1979]. □

Local superlinear convergence results also hold for suitable implementations of the Sequential Secant Method and its variations (Gay-Schnabel and Lopes-Martínez). Under slightly stronger assumptions, we can prove stronger convergence results for these methods. For the Sequential Secant Method and the Lopes-Martínez Method the R-order of convergence (Ortega and Rheinboldt [1970]) is the positive root of $t^{n+1} - t^n - 1 = 0$. The Gay-Schnabel method has R-order equal to the positive root of $t^{2n} - t^{2n-1} - 1 = 0$ (Martínez [1979b]). Both numbers tend to 1 when $n \to \infty$.

After nearly 10 years of practical experience, we verified that COLUM has practically the same behavior as Broyden, in terms of reliability and number of iterations. Many times, COLUM uses less computer time than Broyden because the iteration is slightly less expensive. However, COLUM is not an LCSU method, and local and superlinear convergence cannot be proved using the classical Bounded Deterioration techniques. The following convergence results have been proved for COLUM.

Theorem 3.6. Assume that the sequence $\{x^k\}$ is generated by COLUM, except that when $k \equiv 0 \pmod{m}$, $B_k = J(x^k)$. Then, there exists $\varepsilon > 0$ such that, if $||x^0 - x^*|| \leq \varepsilon$, the sequence converges superlinearly to x^*.

Proof. See Martínez [1984]. For a similar result, concerning an Inverse Column Updating Method, see Martínez and Zambaldi [1992]. □

Theorem 3.7. Assume that $n = 2$. Let $r \in (0, 1)$. Then, there exists $\varepsilon, \delta > 0$ such that, if $||x^0 - x^*|| \leq \varepsilon$ and $||B_0 - J(x^*)|| \leq \delta$, the sequence $\{x^k\}$ generated by COLUM is well defined, converges to x^*, and satisfies (3.17).

Proof. See Martínez [1992c]. □

Theorem 3.8. Assume that the sequence $\{x^k\}$ generated by COLUM is well defined, converges to x^* and satisfies (3.17). Then

$$\lim_{k\to\infty} \frac{||x^{k+2n} - x^*||}{||x^k - x^*||} = 0 \tag{3.24}$$

and

$$\lim_{k\to\infty} ||x^k - x^*||^{1/k} = 0. \tag{3.25}$$

Proof. See Martínez [1992c]. Property (3.25) is called *R-superlinear convergence.* □

A large gap between practice and theory remains to be filled in relation to COLUM. Some of the relevant questions are:

Does COLUM with restarts (as in the hypothesis of Theorem 3.6), satisfy

$$||x^{k+m} - x^*|| \leq c||x^k - x^*||^{p+q} \tag{3.26}$$

for some $c > 0$, q strictly greater than m? The motivation of this question is that the Stationary Newton method with restarts every m iterations satisfies (3.26) with $q = m$.

Does Theorem 3.7 hold for $n > 2$?

Is COLUM superlinearly convergent in the sense of (3.19)?

4. INEXACT-NEWTON METHODS

Many times, large nonlinear systems can be solved using Newton's method, employing a sparse LU factorization for solving (2.2). Most frequently, even if Newton's method is applicable, more efficient algorithms are obtained using COLUM or Broyden, with Newton restarts. In the NIGHTINGALE package, an automatic restart procedure has been incorporated, by means of which a Newton iteration is performed only when it is expected that its efficiency should be greater than the efficiency of previous Quasi-Newton iterations.

Sometimes, the structure of the Jacobian matrix is unsuitable for LU factorizations. That is, a lot of fill-in appears due to that structure and so the iteration becomes very expensive. In many of these cases a strategy of "false Jacobians" works well. By this we mean that we use a Quasi-Newton iteration with restarts, where, at the restarted iterations, B_k is not $J(x^k)$ but a "simplified Jacobian" $\widetilde{J}(x^k)$ such that its LU factorization can be performed without problems.

Unhappily, in many cases, $||\widetilde{J}(x^k) - J(x^k)||$ is excessively large, and the Quasi-Newton method looses its local convergence properties. In these cases, it is strongly recommendable to use Inexact-Newton methods.

The idea is the following. Since we cannot solve (2.2) using a direct (LU) method, we use an Iterative Linear Method for solving (2.2). Usually, iterative linear methods based on Krylov subspaces are preferred (Golub and Van Loan [1989], Hestenes and Stiefel [1952], Saad and Schulz [1986], etc). Iterative Linear Methods are interesting for solving large-scale systems of equations because of their low memory requirements. When we solve (2.2) using an iterative linear method, we need a stopping criterion for deciding when to finish the calculation. A very reasonable stopping criterion is

$$||J(x^k)s^k + F(x^k)|| \leq \theta_k ||F(x^k)||, \tag{4.1}$$

where $\theta_k \in (0,1)$. The condition $\theta_k < 1$ is necessary because, otherwise, the null increment $s^k \equiv 0$ could be accepted as an approximate solution of (2.2). On the other hand, if $\theta_k \approx 0$, the number of iterations needed by the Iterative Linear Method to obtain (4.1) could be excessively large. Therefore, in practice, an intermediate value $\theta_k \approx 0.1$ is recommended.

Dembo, Eisenstat and Steihaug introduced the criterion (4.1) and proved the main local convergence properties of the algorithms based on this criterion.

Theorem 4.1 Assume that $F(x^*) = 0, J(x^*)$ is nonsingular and continuous at x^*, and $\theta_k \leq \theta_{\max} < \theta < 1$. Then there exists $\varepsilon > 0$ such that, if $||x^0 - x^*|| \leq \varepsilon$, the sequence $\{x^k\}$ obtained using (4.1) and (2.3) converges to x^* and satisfies

$$||x^{k+1} - x^k||_* \leq \theta ||x^k - x^*||_* \tag{4.2}$$

for all $k \geq 0$, where $||y||_* = ||J(x^*)y||$. If $\lim_{k\to\infty} \theta_k = 0$ the convergence is superlinear.

Proof. See Dembo, Eisenstat and Steihaug [1982]. □

Krylov subspace methods for solving systems like (2.2) are usually implemented using some preconditioning scheme. See Axelsson [1985]. By this we mean that the original system is replaced by an equivalent one, which is easier to solve by the Iterative Linear Solver. In the case of (2.2), we wish to replace the linear system by

$$B_k^{-1} J(x^k) s^k = -B_k^{-1} F(x^k) \tag{4.3}$$

where B_k^{-1} (or, at least, the product $B_k^{-1}z$) must be easy to compute and $B_k \approx J(x^k)$. For general linear systems, many useful preconditioners B_k have been introduced. Most of them are based on Incomplete LU Factorizations, or on Stationary Linear iterations. A very cheap and popular procedure is to use the diagonal of the original matrix as preconditioner. Many other preconditioners for specific problems can be found in the papers published in Spedicato [1991]. A common feature to different preconditioning schemes applied to a linear system $Az = b$ is that the first iteration of the preconditioned Iterative Linear Solver is $z^1 = \lambda B^{-1} b$, where B is the preconditioner. So, in the case of the system (2.2), the first increment tried should be of the form $-\lambda B_k^{-1} F(x^k)$. This increment will be accepted if it satisfies (4.1). However, (2.2) is not an isolated linear system of equations. In fact, probably $J(x^k) \approx J(x^{k-1})$ specially when $k \to \infty$. Therefore, we are motivated to use information about $B_k, F(x^k), F(x^{k+1}), x^{k+1}, x^k$ when we choose the preconditioner B_{k+1}. This idea leads to impose a Secant Condition to the preconditioner. So, we would like to introduce an algorithm based on (4.1), where the sequence of preconditioners B_k are chosen in order to satisfy

$$B_{k+1} s^k = y^k \tag{4.4}$$

for all $k \in I\!N$.

We saw in Section 2 that there exist infinite many possible choices of B_{k+1} satisfying (4.4). Nazareth and Nocedal [1978] and Nash [1985] suggested to use the classical BFGS formula in order to precondition (2.2), when we deal with minimization problems. Our preference is to define

$$B_{k+1} = C_{k+1} + D_{k+1} \tag{4.5}$$

where C_{k+1} is a classical preconditioner and D_k is chosen to satisfy (4.4).

The main appeal of Secant Preconditioners is that it has been shown (Martínez [1992b]) that using them it is possible to obtain stronger convergence results than the one mentioned in Theorem 4.1. In fact, the main drawback of this result is the necessity of $\theta_k \to 0$ for obtaining superlinear convergence. The following Preconditioned Inexact Newton Method was introduced by Martínez [1992b] with the aim of obtaining superlinear convergence without imposing a precision tending to infinity in the iterative resolution of (2.2).

Algorithm 4.2. Let $\theta_k \in (0, \theta)$ for all $k \in I\!N, \theta \in (0, 1)$ and $\lim_{k \to \infty} \theta_k = 0$. Assume that $x^0 \in I\!R^n$ is an initial approximation to the solution of (1.1) and $B_0 \in I\!R^{n \times n}$ is an initial nonsingular preconditioner. Given $x^k \in I\!R^n$ and B_k nonsingular, the steps for obtaining x^{k+1}, B_{k+1} are the following.

Step 1. Compute

$$s_Q^k = -B_k^{-1} F(x^k). \tag{4.6}$$

Step 2. If

$$||J(x^k) s_Q^k + F(x^k)|| \leq \theta ||F(x^k)|| \tag{4.7}$$

define

$$s^k = s_Q^k. \tag{4.8}$$

Else, find an increment s^k such that (4.1) holds, using some iterative method.

Step 3. Define $x^{k+1} = x^k + s^k$. $\odot$

The following theorem states the main convergence result relative to Algorithm 4.2.

Theorem 4.3. Assume that $F : \Omega \subset \mathbb{R}^n \rightarrow \mathbb{R}^n$, Ω an open and convex set, $F \in C^1(\Omega)$, $J(x^*)$ nonsingular, $F(x^*) = 0$ and (2.4) holds for some $L \geq 0, p \geq 1$. Suppose that $||B_k||$ and $||B_k^{-1}||$ are bounded and that the Dennis-Moré condition (3.18) is satisfied. Then, there exists $\varepsilon > 0$ such that, if $||x^0 - x^*|| \leq \varepsilon$, the sequence $\{x^k\}$ generated by Algorithm 4.2 converges superlinearly to x^*. Moreover, there exists $k_0 \in \mathbb{N}$ such that $s^k = s_Q^k$ for all $k \geq k_0$.

Proof. See Martínez [1992b]. $\square$

Theorem 4.3 states that, if we use preconditioners satisfying the Dennis-Moré condition, superlinear convergence is obtained without $\lim_{k\to\infty} \theta_k = 0$. In fact, the first iteration s_Q^k of the preconditioned iterative linear method will satisfy (4.6), and so it will be accepted as the new increment s^k, preserving superlinearity. Least-Change Secant-Update formulae can be used for obtaining preconditioners satisfying the hypotheses of Theorem 4.3.

Recently, Abaffy [1992] considered the possibility of using, for solving (2.2), iterative linear algorithms that work componentwise, that is, without evaluating the whole residual at each step. For this type of algorithms, he introduced a new stopping criterion, different from (4.1), which also ensures superlinear convergence.

5. DECOMPOSITION METHODS

The methods studied in the previous sections evaluate *all* the components of the function F at the same points. This is not always the best possible strategy. In many practical problems, given a guess of the solution, the evaluation of a few components of F is enough to suggest a new useful estimate. Methods that evaluate different components at different points are called *Decomposition Methods.*

The (Block) SOR-Newton method proceeds as follows. Assume that the components of F are divided into blocks $F_1, \ldots, F_m$ and that we decompose $x = \begin{pmatrix} x_1 \\ \vdots \\ x_m \end{pmatrix}$, where

$x_i \in I\!R^{n_i},\ F_i(x) \in I\!R^{n_i},\ i = 1, \ldots, m$. We denote $\dfrac{\partial F_i}{\partial x_j}$ the matrix of partial derivatives of F_i with respect to the variables x_j. The SOR-Newton method is defined by

$$x_i^{k+1} = x_i^k - w \left(\frac{\partial F_i}{\partial x_i}\right)^{-1} (x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i^k, \ldots, x_m^k) F_i(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i^k, \ldots, x_m^k) \quad (5.1)$$

for $i = 1, \ldots, m$, where $w \in (0, 2)$ is a *relaxation parameter*. A generalization of (5.1) is

$$x_i^{k+1} = x_i^k - w B_{i,k}^{-1} F_i(x_1^{k+1}, \ldots, x_{i-1}^{k+1}, x_i^k, \ldots, x_m^k). \quad (5.2)$$

Methods of type (5.2) are called SOR-Quasi-Newton methods (Martínez [1992d, 1992e]). If $B_{i,k+1}$ satisfies

$$B_{i,k+1}(x_i^{k+1} - x_i^k) = F_i(x_1^{k+1}, \ldots, x_i^{k+1}, \ldots, x_m^k) - F_i(x_1^{k+1}, \ldots, x_i^k, \ldots, x_m^k) \quad (5.3)$$

we say that (5.2) defines a SOR-Secant method (Martínez [1992e]). The local convergence analysis of (5.1), (5.2) and (5.3) has been made in Martínez [1992e]. Essentially, the condition for the local linear convergence of (5.1) - (5.3) is the same convergence condition of the linear SOR method for a linear system with matrix $J(x^*)$. If n_i is small for all $i = 1, \ldots, n$, SOR methods have low storage requirements and so, they are useful for large-scale problems.

The *asynchronous* generalization of (5.2) can be very useful in parallel computer environments. Assume that we work with m parallel processors, where processor i is specialized in the computation of new guesses of the component x_i. That is, processor i only computes $x_i - w B_i^{-1} F_i(x)$ for given B_i and x. Different processors use different times for their respective computations and, as soon as a processor produces a result, this result is communicated to the other processors. In this way, all the processors are working all the time and full advantage is taken from parallelism. See Bertsekas and Tsitsiklis [1989]. Let us now formalize these ideas. The index $k \in I\!N$ represents the times at which at least one component of x is changed. We define $T_i \subset I\!N$ by

$$T_i = \text{Set of times at which } x_i \text{ is updated.}$$

Then, the Asynchronous SOR method is defined by

$$x_i^{k+1} = x_i^k - w B_{i,k}^{-1} F_i(x_1^{\nu(i,1,k)}, \ldots, x_i^k, \ldots, x_m^{\nu(i,m,k)}) \quad (5.4)$$

if $k + 1 \in T_i$, and

$$x_i^{k+1} = x_i^k \quad (5.5)$$

if $k+1 \notin T_i$, where $0 \le \nu(i,j,k) \le k$ for all $k \in I\!N, i, j = 1, \ldots, m$. (5.2) is a particular case of (5.4) - (5.5) defining $T_i = \{i + jm, j \in I\!N\}, \nu(i,j,k) = k$ for all $i, j, \ldots, n, i \neq j, k \in I\!N$.

The Jacobi-Quasi-Newton method corresponds to $T_i = \{1, 2, 3, \ldots\}, \nu(i, j, k) = k$. Secant Asynchronous methods based on (5.4) can also be considered.

The SOR and Jacobi methods for linear systems are strongly related with the projection methods of Kaczmarz [1937] and Cimmino [1938] respectively. The advantage of Kaczmarz and Cimmino is that convergence is guaranteed for every linear system, while SOR and Jacobi require a condition on the spectral radius of the transformation. However, in general, when SOR and Jacobi converge, they are far more efficient than Kaczmarz and Cimmino. Nonlinear generalizations of Kaczmarz and Cimmino may be found in Tompkins [1955], McCormick [1977], Meyn [1983] , Martínez [1986a, 1986b, 1986c] and Diniz - Ehrhardt and Martínez [1992].

Different decomposition methods are motivated by direct methods for solving linear systems, for example, the family of Nonlinear ABS algorithms developed in Abaffy and Spedicato [1989], Abaffy, Broyden and Spedicato [1984], Abaffy, Galantai and Spedicato [1987], Spedicato, Chen and Deng [1992], etc.. These methods generalize classical algorithms due to Brown [1969], Brent [1973], Gay [1975] and Martínez [1979c, 1980]). The idea is the following. Divide the components of F into m groups $F_1, \ldots, F_m$. Assume that x^k has been computed. We generate $x^{k,1}, \ldots, x^{k,m}$ by:

$$x^{k,0} = x^k, \tag{5.6}$$

$$J_i(x^{k,i})(x^{k,i+1} - x^{k,i}) = -F_i(x^{k,i}), \tag{5.7}$$

$$J_j(x^{k,j})(x^{k,i+1} - x^{k,i}) = 0, j = 1, \ldots, i-1, \tag{5.8}$$

where $J_j(x) = F'_j(x), i = 0, 1, \ldots, m-1$.

$$x^{k+1} = x^{k,m}. \tag{5.9}$$

Clearly, the scheme (5.6) - (5.9) solves a linear system of equations in one cycle. However, ther are infinite many ways to choose the intermediate points $x^{k,1}, \ldots, x^{k,m-1}$. Different choices of these points originate different methods of the ABS class. The first motivation given in the papers of Brown and Brent for methods of type (5.6) - (5.9) was that, using suitable factorizations, the derivatives can be approximated by differences in a more economic way than in the Finite Difference Newton method. Methods of this class have, in general, the same local convergence properties of Newton, though the proofs are technically complicated.

Many other decomposition algorithms have been introduced with the aim of taking advantage of particular structures. For systems that are reducible to block lower triangular form see Eriksson [1976] and Dennis, Martínez and Zhang [1992]. For Block Tridiagonal Systems, see Hoyer, Schmidt and Shabani [1989]. Much theory about decomposition methods has been produced by the German school (Schmidt [1987], Burmeister and Schmidt [1988], Hoyer [1987], Hoyer and Schmidt [1984], Schmidt, Hoyer and Haufe [1985], etc.)

6. GLOBALIZATION BY OPTIMIZATION

In the previous sections we studied local methods, that is, algorithms that converge, usually with a high rate of convergence, if the initial point is close enough to the solution. Luckily, in many cases the domain of convergence of local algorithms is large enough to guarantee practical efficiency. However, locally convergent methods may not converge if the starting point is very poor, or if the system is highly nonlinear. By this reason, local methods are usually modified in order to improve their global convergence properties. The most usual way to do this is to transform (1.1) into an Optimization Problem, with the objective function $f(x) = \frac{1}{2}||F(x)||^2$. Then, (1.1) becomes the problem of finding a global minimizer of f.

However, the decision of merely using a method to minimize f, in order to solve (1.1), is not satisfactory. In fact, sometimes efficient local methods converge rapidly to a solution but the generated sequence $\{x^k\}$ does not exhibit monotonic behavior in $f(x^k)$. In these cases, the pure local method is much more efficient than the f-minimization method. Often, the minimization method converges to a local (nonglobal) minimizer of f, while the local method converges to a solution of (1.1). By these reasons, it is necessary to give a chance to the local method before calling the minimization algorithm. Different solutions have been proposed to this problem (Grippo, Lampariello e Lucidi [1986]). Here we describe a strategy that combines local algorithms and minimization methods introduced in the NIGHTINGALE package. We define "ordinary iterations" and "special iterations". By an ordinary iteration we understand an iteration produced by any of the methods described in sections 2 to 4 of this paper. Decomposition methods can also be considered, with some modifications. A special iteration is an iteration produced by a minimization algorithm applied to f. We define, for all $k \in I\!N$,

$$a^k = \text{ Argmin } \{f(x^0), \dots, f(x^k)\}. \tag{6.1}$$

Ordinary and special iterations are combined by the following strategy.

Algorithm 6.1. Initialize $k \leftarrow 0$, FLAG $\leftarrow 1$. Let $q \geq 0$ be an integer, $\gamma \in (0, 1)$.

Step 1. If FLAG $= 1$, obtain x^{k+1} using an ordinary iteration. Else, obtain x^{k+1} using a special iteration.

Step 2. If

$$f(a^{k+1}) \leq \gamma f(a^{k-q}) \tag{6.2}$$

set FLAG $\leftarrow 1, k \leftarrow k + 1$ and go to Step 1. Else, re-define $x^{k+1} \leftarrow a^{k+1}$. Set FLAG $\leftarrow -1, k \leftarrow k + 1$ and go to Step 1. $\odot$

If the test (6.2) is satisfied infinite many times, then there exists a subsequence of $\{x^k\}$ such that $\lim_{k\to\infty} ||F(x^k)|| = 0$. So, if the sequence is bounded, we will be able to find a solution of (1.1) up to any prescribed accuracy. Conversely, if (6.2) does not hold for all $k \geq k_0$, then all the iterations starting from the k_0-th will be special, and the convergence properties of the sequence will be those of the minimization algorithm.

In principle, we can use as minimization algorithm to define special iterations any method for minimization described in the literature (Dennis and Schnabel [1983], Fletcher [1987], etc.). For large-scale problems, the NIGHTINGALE package uses a strategy based on Trust Regions derived from the Inexact Newton approach (Friedlander, Gomes-Ruggiero, Martínez and Santos [1993]). This strategy is described by the following algorithm.

Algorithm 6.2. Assume that $\Delta_{\min} > 0, \alpha \in (0,1)$ are given independently of the iteration k. Define $\psi_k(x) = ||F(x^k) + J(x^k)(x - x^k)||^2$, $\Delta \geq \Delta_{\min}$.

Step 1. Compute an approximate minimizer $\overline{x}$ of $\psi_k(x)$ on the box $||x - x^k||_\infty \leq \Delta$ such that $\psi_k(x) \leq \psi_k(x_Q^k)$, x_Q^k is the projection of $x^k - 2J(x^k)^T F(x^k)/M_k$ on the box and $M_k \geq 2||J(x_k)||_1 ||J(x_k)||_\infty$.

Step 2. If

$$||F(\overline{x})||^2 \leq ||F(x^k)||^2 + \alpha(\psi_k - \psi_k(\overline{x})) \tag{6.3}$$

define $x^{k+1} = \overline{x}$. Else, choose $\Delta_{\text{new}} \in [0.1||\overline{x} - x^k||, 0.9\Delta]$, replace Δ by Δ_{new} and go to Step 1. $\odot$

Most of the work of Algorithm 6.2 is concentrated on the approximate solution of

$$\left.\begin{array}{rl} & \text{Minimize } \psi_k(x) \\ s.t. & ||x - x^k||_\infty \leq \Delta \end{array}\right\} \tag{6.4}$$

(6.4) is the problem of minimizing a convex quadratic with box constraints. For this problem, algorithms based on combinations of Krylov Subspace methods with Gradient Projection strategies are currently preferred. In NIGHTINGALE, the approximate solution of (6.4) is defined as a point that satisfies $\psi_k(x) \leq \psi_k(x_Q^k)$ and where, in addition, the norm of projected gradient of $\psi_k(x)$ is less than $0.1||J(x^k)^T F(x^k)||$. We also choose: $\Delta_{\min} = 0.001\times$ (typical $||x||$), initial choice of $\Delta \equiv \Delta_0 = ||x^0||, \Delta_{\text{new}} = 0.5||\overline{x} - x^k||$, further choice is $\Delta = 4 \times \Delta$.

The convergence properties of Algorithm 6.2 were given in Friedlander, Gomes - Ruggiero, Martínez and Santos [1993]. Every limit point x^* of a sequence $\{x^k\}$ generated by this algorithm satisfies $J(x^*)^T F(x^*) = 0$. Therefore, x^* is a solution of (1.1) if $J(x^*)$

is nonsingular. Unhappily, if $J(x^*)$ is singular, it is possible that $F(x^*) \neq 0$. This is the main weakness of algorithms based on optimization for the globalization of (1.1).

An advantage of special iterations based on box trust regions is that they can be easily adapted to situations where we have natural bounds for the solution of (1.1). Other recent methods based on the inexact Newton approach with global convergent properties were given by Deuflhard [1991] and Eisenstat and Walker [1993].

7. GLOBALIZATION BY HOMOTOPIES

In Section 6 we saw that local methods for solving $F(x) = 0$ can be "globalized" through their transformation into minimization problems. In this section we study another popular technique to solve (1.1) when the initial approximation is poor. This technique is based on homotopies. A homotopy associated to this problem is a function $H(x,t) : \mathbb{R}^n \times \mathbb{R} \to \mathbb{R}$ such that

$$\left.\begin{array}{rcr} H(x,1) & = & F(x) \\ H(x^0,0) & = & 0 \end{array}\right\} \tag{7.1}$$

If H satisfies (7.1) we can expect that $\Gamma \equiv \{(x,t) \in \mathbb{R}^n \times \mathbb{R} \mid H(x,t) = 0, 0 \le t \le 1\}$ will be an arc connecting the initial approximation x^0 with a solution x^*. Homotopic techniques consist in tracing Γ from $t = 0$ to $t = 1$ in a reliable and efficient way. The first to propose the homotopic idea were Lahaye [1935] and Davidenko [1955]. The homotopic principle is very popular. In fact, this is one of the techniques mentioned in the article *Numerical Analysis* of *Encyclopaedia Britannica*.

Sometimes, the homotopy has an interest by itself. In other cases we are interested only in the solution of $H(x,1) = 0$. In this section we deal with the latter situation. This has a practical consequence. If we are interested only in the solution of $H(x,1) = 0$, it makes sense to interrupt the tracing of Γ when t is close to 1, trying to apply a local method from the current point. In general, this will be more efficient than to insist in carefully tracing Γ.

Unhappily, Γ could not be an arc of finite length. It is possible that, if we trace Γ starting from $(x^0, 0)$, we never arrive to $(x^*, 1)$. However, some classical results of Differential Geometry help us to identify situations in which tracing Γ leads to the solution of (1.1). If $H'(x,t) \equiv (H'_x(x,t), H'_t(x,t))$ has full rank for all $(x,t) \in H^{-1}(\{0\})$, then $H^{-1}(\{0\})$ is a discrete union of curves homomorphic to $\mathbb{R}$ or to the one-dimensional sphere S^1. See Milnor [1969]. In this case, each component of $H^{-1}(\{0\})$ can be traced efficiently using numerical methods. The condition that $H'(x,t)$ has full rank results, in many practical cases, from the application of a theorem of Chow, Mallet-Paret and Yorke [1979]. See also Watson [1979]. However, this does not guarantee that the component of $H^{-1}(\{0\})$ that passes through $(x^0, 0)$ should arrive to $(x^*, 1)$.

Now, if the unique solution of $H(x, 0) = 0$ is x^0 and $H'_x(x^0, 0)$ is nonsingular, it can be proved that the component of $H^{-1}(\{0\})$ that passes through $(x^0, 0)$ is homomorphic to $\mathbb{R}$. In fact, even weaker hypotheses are sufficient (Ortega and Rheinboldt [1970, Ch. 6]). These hypotheses are easy to verify and guarantee that, if we trace Γ starting from $(x^0, 0)$, it is not possible to return to the x^0. However this does not guarantee yet that $(x^*, 1)$ will be reached.

To be sure that this homotopy is useful we must verify, in addition to the previous assumptions, that Γ is bounded. In this case, Γ will be necessarily a segment of arc joining $(x^0, 0)$ with $(x^*, 1)$. When we want to make sure that a given homopopy will solve a problem, the difficult part is precisely to prove boundedness of Γ.

Sometimes, "natural" homotopies are used. This mean, homotopies where the parameter t that is suggested by the problem under consideration. However, there exist useful "artificial" homotopies. Let us mention here two of them. The homotopy of Reduction of Residual is

$$H(x,t) = F(x) + (t-1)F(x^0).$$

The "regularizing homotopy", used in the well known package HOMPACK (Watson, Billups and Morgan [1987]) is

$$H(x,t) = tF(x) + (1-t)(x - x^0).$$

After the choice of H, we must use a numerical method to trace Γ. First, Γ must be parametrized. Frequently, we describe Γ as a function of the parameter t. However, if, for some t_0, we have that $H'_x(x,t_0)$ is singular, x cannot be expressed as a function of t in a neighborhood of t_0, and, instead of increasing t, we must decrease this parameter, in order to advance in Γ. By this reason, it is usual to trace Γ using s, the arclength, as parameter.

If this is the case, the procedure generally recommended to trace Γ is of Predictor-Corrector type. Given a set of points $(x(s_1), t(s_1)), \ldots, (x(s_m), t(s_m))$ computed consecutively on Γ, and an increment $\Delta > 0$, we calculate a polynomial interpolating those points and, using this polynomial, we compute a predictor point $(\overline{x}(s_m + \Delta), \overline{t}(s_m + \Delta))$. This new point does not necessarily belong to Γ. So, using this point as initial approximation, we compute a point on Γ using a local method for the nonlinear system $H(x,t) = 0$. If, in this system, we consider that t is also a variable, we have n equations with $n+1$ unknowns. Special local algorithms for underdetermined nonlinear systems were developed by Walker and Watson [1989], Martínez [1991], etc. This is the Corrector Phase. When we arrive to a point (x,t) with t close to 1, we try the direct application of a local method for $F(x) = 0$.

An interesting discussion on homotopy methods is given in a recent survey by Forster [1993].

8. CONCLUSIONS

Many practical applications give rise to large scale nonlinear systems of equations. From the local point of view, the most interesting techniques for this case are variations of the Inexact Newton method. To develop preconditioners for Krylov Subspace solvers, taking into account the structure of these methods and problems, is a challenging problem. Since solving a linear system is a particular case of minimizing a quadratic on a box, it turns out that to solve efficiently this last problem is crucial. Moreover the problem appears again in the development of methods for globalization using optimization, and in the Corrector Phase of Homotopy methods.

For very small problems Newton's method continues to be the best choice, and for medium to large problems a combination Newton-Quasi-Newton seems to be better. Many decomposition methods are interesting when they are induced by characteristics of the problem structure , or when decomposition is suggested by the computer architecture.

In this paper, we surveyed methods based on first-order approximations of the nonlinear system. Methods based on approximations of higher order have also been developed (Schnabel and Frank [1984]) and are useful in the presence of singularities of the Jacobians. Large - scale implementations of these methods seem to be hard.

References

Abaffy, J. [1992]: Superlinear convergence theorems for Newton-type methods for nonlinear systems of equations, *JOTA* , 73, pp. 269 - 277.

Abaffy, J.; Broyden, C.G.; Spedicato, E. [1984]: A class of direct methods for linear equations, *Numerische Mathematik* 45, pp. 361-376.

Abaffy, J.; Galantai, A. ; Spedicato, E. [1987]: The local convergence of ABS method for nonlinear algebraic system, *Numerische Mathematik* 51, pp. 429 - 439.

Abaffy, J.; Spedicato, E. [1989]: *ABS Projection Algorithms, Mathematical Techniques for Linear and Nonlinear Equations*, Ellis Horwood, Chichester.

Axelsson, O. [1985]: A survey of preconditioned iterative methods for linear systems of equations, *BIT* 25, pp. 166 - 187.

Barnes, J.G.P. [1965]: An algorithm for solving nonlinear equations based on the secant method, *Computer Journal* 8, pp. 66 - 72.

Bertsekas, D.P.; Tsitsiklis, C. [1989]: *Parallel and Distributed Computation. Numerical Methods*, The Johns Hopkins University Press, New York.

Brent, R.P. [1973]: Some efficient algorithms for solving systems of nonlinear equations, *SIAM Journal on Numerical Analysis* 10, pp. 327 - 344.

Brown, K.M. [1969]: A quadratically convergent Newton - like method based upon Gaussian elimination, *SIAM Journal on Numerical Analysis* 6, pp. 560 - 569.

Broyden, C.G. [1965]: A class of methods for solving nonlinear simultaneous equations, *Mathematics of Computation* 19, pp. 577-593.

Broyden, C.G.; Dennis Jr., J.E.; Moré, J.J. [1973]: On the local and superlinear convergence of quasi-Newton methods, *Journal of the Institute of Mathematics and its Applications* 12, pp. 223-245.

Burmeister, W.; Schmidt, J.W. [1978]: On the k-order of coupled sequences arising in single - step type methods, *Numerische Mathematik* 33, pp. 653 - 661.

Chadee, F.F. [1985]: Sparse quasi-Newton methods and the continuation problem, T.R. S.O.L. 85-8, Department of Operations Research, Stanford University.

Chow, S.N.; Mallet-Paret, J.; Yorke, J.A. [1978]: Finding zeros of maps: Homotopy methods that are constructive with probability one, Mathematics of Computation 32, pp. 887-899.

Cimmino, G. [1938]: Calcolo approsimato per la soluzione dei sistemi di equazioni lineari, *La Ricerca Scientifica Ser II, Ano IV 1* pp. 326-333.

Coleman,T. F.; Garbow, B. S.; Moré,J. J. [1984]: Software for estimating sparse Jacobian matrices, *ACM Trans. Math. Software* 11, pp. 363-378.

Coleman,T. F.; Moré,J. J. [1983]: Estimation of sparse Jacobian matrices and graph coloring problems, *SIAM Journal on Numerical Analysis* 20, pp. 187-209.

Curtis,A. R.; Powell, M. J. D.; Reid, J. K. [1974], On the estimation of sparse Jacobian matrices, *Journal of the Institute of Mathematics and its Applications* 13, pp. 117-120.

Davidenko, D.F. [1953]: On the approximate solution of nonlinear equations, *Ukrain. Mat. Z.* 5, pp. 196 -206.

Dembo, R.S.; Eisenstat, S.C.; Steihaug, T. [1982]: Inexact Newton methods, *SIAM Journal on Numerical Analysis* 19, pp. 400–408.

Dennis Jr., J.E.; Martínez, J.M.; Zhang, X. [1992]: Triangular Decomposition Methods for Reducible Nonlinear Systems of Equations , Technical Report, Department of Mathematical Sciences, Rice University.

Dennis Jr., J.E.; Marwil, E.S. [1982]: Direct secant updates of matrix factorizations, *Mathematics of Computation* 38, pp. 459–476.

Dennis Jr., J.E.; Moré, J.J. [1974]: A characterization of superlinear convergence and its application to quasi - Newton methods, *Mathematics of Computation* 28, pp. 549 -560.

Dennis Jr., J.E.; Moré, J.J. [1977]: Quasi-Newton methods, motivation and theory, *SIAM Review* 19, pp. 46-89.

Dennis Jr.,J.E.; Schnabel,R.B. [1979]: Least change secant updates for quasi-Newton methods, *SIAM Review* 21, pp. 443-459.

Dennis Jr.,J.E.; Schnabel,R.B. [1983]:*Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice-Hall, Englewood Cliffs.

Dennis Jr., J.E. ; Walker, H.F. [1981]: Convergence theorems for least-change secant update methods, *SIAM Journal on Numerical Analysis* 18, pp. 949-987.

Deuflhard, P. [1991]: Global inexact Newton methods for very large scale nonlinear problems *Impact of Computing in Science and Engineering* 3, pp. 366–393.

Diniz-Ehrhardt, M.A.; Martínez, J.M. [1992]: A parallel projection method for overdetermined nonlinear systems of equations, to appear in *Numerical Algorithms*.

Duff, I.S. [1977]: MA28 – a set of Fortran subroutrines for sparse unsymmetric linear equations. AERE R8730, HMSO, London.

Duff, I.S.; Erisman, A.M.; Reid, J.K. [1989]:*Direct Methods for Sparse Matrices*, Oxford Scientific Publications.

Eisenstat, S.C.; Walker, H.F. [1993]: Globally convergent inexact Newton methods, to appear in *SIAM Journal on Optimization*.

Eriksson, J. [1976]: A note on the decomposition of systems of sparse nonlinear equations, *BIT* 16, pp. 462 - 465.

Fletcher, R. [1987]: *Practical Methods of Optimization* (2nd edition), John Wiley and Sons, New York.

Forster, W. [1993]: Homotopy methods, to appear in *Handbook of Global Optimization*, Kluwer.

Friedlander, A.; Gomes-Ruggiero, M.A.; Martínez, J.M.; Santos, S.A. [1993]: A globally convergent method for solving nonlinear systems using a trust - region strategy, Technical Report, Department of Applied Mathematics, University of Campinas.

Gay, D.M. [1975]: Brown's method and some generalizations with applications to minimization problems, Ph D Thesis, Computer Science Department, Cornell University, Ithaca , New York.

Gay, D.M. [1979]: Some convergence properties of Broyden's method, *SIAM Journal on Numerical Analysis* 16, pp. 623 - 630.

Gay, D.M.; Schnabel, R.B. [1978]: Solving systems of nonlinear equations by Broyden's method with projected updates, in *Nonlinear Programming 3*, edited by O. Mangasarian, R. Meyer and S. Robinson, Academic Press, New York, pp. 245-281.

George, A.; Ng, E. [1987]: Symbolic factorization for sparse Gaussian elimination with partial pivoting, *SIAM Journal on Scientific and Statistical Computing* 8, pp. 877-898.

Golub, G.H.; Van Loan, Ch.F. [1989]: *Matrix Computations*, The Johns Hopkins University Press, Baltimore and London.

Gomes–Ruggiero, M.A.; Martínez, J.M. [1992]: The Column–Updating Method for solving nonlinear equations in Hilbert space, *Mathematical Modelling and Numerical Analysis* 26, pp 309-330.

Gomes–Ruggiero, M.A.; Martínez, J.M.; Moretti, A.C. [1992]: Comparing algorithms for solving sparse nonlinear systems of equations, *SIAM Journal on Scientific and Statistical Computing* 13, pp. 459 - 483.

Gragg, W.B.; Stewart, G.W. [1976]: A stable variant of the secant method for solving nonlinear equations, *SIAM Journal on Numerical Analysis* 13, pp. 127 - 140.

Griewank, A. [1986]: The solution of boundary value problems by Broyden based secant methods, *CTAC – 85. Proceedings of the Computational Techniques and Applications Conference*, J. Noye and R. May (eds.), North Holland, Amsterdam.

Griewank, A. [1992]: Achieving Logarithmic Growth of Temporal and Spacial Complexity in Reverse Automatic Differentiation, *Optimization Methods and Software* 1, pp. 35 - 54.

Griewank, A.; Toint, Ph.L. [1982a]: On the unconstrained optimization of partially separable functions, in *Nonlinear Optimization 1981*, edited by M.J.D. Powell, Academic Press, New York.

Griewank, A.; Toint, Ph.L. [1982b]: Partitioned variable metric for large structured optimization problems, *Numerische Mathematik* 39, pp. 119 - 137.

Griewank, A.; Toint, Ph.L. [1982c]: Local convergence analysis for partitioned quasi-Newton updates, *Numerische Mathematik* 39, pp. 429-448.

Griewank, A.; Toint, Ph.L. [1984]: Numerical experiments with partially separable optimization problems, in *Numerical Analysis Proceedings Dundee 1983*, edited by D.F. Griffiths, Lecture Notes in Mathematics vol. 1066, Springer - Verlag, Berlin, pp. 203-220.

Grippo, L.; Lampariello, F.; Lucidi, S. [1986]: A nonmonotone line search technique for Newton's method, *SIAM Journal on Numerical Analysis* 23, pp. 707 - 716.

Hart, W.E.; Soul, S.O.W. [1973]: Quasi-Newton methods for discretized nonlinear boundary value problems, *J. Inst. Math. Applics.* 11, pp. 351 - 359.

Hestenes, M.R.; Stiefel, E. [1952]: Methods of conjugate gradients for solving linear systems, *Journal of Research of the National Bureau of Standards* B49, pp. 409 - 436.

Hoyer, W. [1987]: Quadratically convergent decomposition algorithms for nonlinear systems with special structure, Inform. Tech. Univ. Dresden 07 , pp. 23-87.

Hoyer, W.; Schmidt, J.W. [1984]: Newton-type decomposition methods for equations arising in network analysis, *Z. Angew-Math. Mech.* 64, pp. 397 - 405.

Hoyer, W.; Schmidt, J.W.; Shabani, N. [1989]: Superlinearly convergent decompositon methods for block - tridiagonal nonlinear systems of equations, *Numerical Functional Analysis and Optimization* 10 (9 & 10), pp. 961 - 975.

Iri, M. [1984]: Simultaneous computations of functions, partial derivatives and estimates of rounding errors. Complexity and Practicality , *Japan Journal of Applied Mathematics* 1, pp. 223 - 252.

Johnson, G.W.; Austria, N.H. [1983]: A quasi-Newton method employing direct secant updates of matrix factorizations, *SIAM Journal on Numerical Analysis* 20, pp. 315-325.

Kaczmarz, S. [1937]: Angenaherte Auflösung von Systemen linearer Gleichungen, *Bull. Acad. Polon. Sci. Lett. A35*, pp. 355-357.

Kelley, C.T.; Sachs, E.W. [1987]: A quasi-Newton method for elliptic boundary value problems, *SIAM Journal on Numerical Analysis* 24, pp. 516 - 531.

Lahaye, E. [1934]: Une méthode de résolution d'une catégorie d'equations transcendantes, Comptes Rendus Acad. Sci. Paris 198, pp. 1840-1842.

Lopes, T.L.; Martínez, J.M. [1980]: Combination of the Sequential Secant Method and Broyden's method with projected updates, *Computing* 25, pp. 379-386.

Martínez, J.M. [1979a]: Three new algorithms based on the sequential secant method, *BIT* 19, pp. 236-243.

Martínez, J.M. [1979b]: On the order of convergence of Broyden - Gay - Schnabel's method, *Commentationes Mathematicae Universitatis Carolinae* 19, pp. 107-118.

Martínez, J.M. [1979c]: Generalization of the methods of Brent and Brown for solving nonlinear simultaneous equations, *SIAM Journal on Numerical Analysis* 16, pp. 434 - 448.

Martínez, J.M. [1980]: Solving nonlinear simultaneous equations with a generalization of Brent's method, *BIT* 20, pp. 501 - 510.

Martínez, J.M. [1983]: A quasi–Newton method with a new updating for the LDU factorization of the approximate Jacobian, *Matemática Aplicada e Computacional* 2, pp. 131–142.

Martínez, J.M. [1984]: A quasi–Newton method with modification of one column per iteration, *Computing* 33, pp. 353–362.

Martínez, J.M. [1986a]: The method of Successive Orthogonal Projections for solving nonlinear simultaneous equations, *Calcolo* 23, pp. 93 - 105.

Martínez, J.M. [1986b]: Solving systems of nonlinear simultaneous equations by means of an accelerated Successive Orthogonal Projections Method, *Computational and Applied Mathematics* 165, pp. 169 - 179.

Martínez, J.M. [1986c]: Solution of nonlinear systems of equations by an optimal projection method, *Computing* 37, pp. 59 - 70.

Martínez, J.M. [1987]: Quasi–Newton Methods with Factorization Scaling for Solving Sparse Nonlinear Systems of Equations, *Computing* 38, pp. 133–141.

Martínez, J.M. [1990a]: A family of quasi-Newton methods for nonlinear equations with direct secant updates of matrix factorizations, *SIAM Journal on Numerical Analysis* 27, pp. 1034-1049.

Martínez, J.M. [1990b]: Local convergence theory of inexact Newton methods based on structured least change updates, *Mathematics of Computation* 55, pp. 143-168.

Martínez, J.M. [1991]: Quasi-Newton Methods for Solving Underdetermined Nonlinear Simultaneous Equations, *Journal of Computational and Applied Mathematics* 34, pp. 171–190.

Martínez, J.M. [1992a]: On the relation between two local convergence theories of least change secant update methods, *Mathematics of Computation* 59, pp. 457–481.

Martínez, J.M. [1992b]: A Theory of Secant Preconditioners, to appear in *Mathematics of Computation*.

Martínez, J.M. [1992c]: On the Convergence of the Column-Updating Methods, Technical Report, Department of Applied Mathematics, University of Campinas.

Martínez, J.M. [1992d]: Fixed-Point Quasi-Newton Methods, *SIAM Journal on Numerical Analysis* 29, pp. 1413–1434.

Martínez, J.M. [1992e]: SOR - Secant Methods, to appear in *SIAM Journal on Numerical Analysis.*

Martínez, J.M.; Zambaldi, M.C. [1992]: An inverse Column-Updating Method for solving Large-Scale Nonlinear Systems of Equations, to appear in *Optimization Methods and Software.*

Matthies, H.; Strang, G. [1979]: The solution of nonlinear finite element equations, *International Journal of Numerical Methods in Engineering* 14, pp. 1613 - 1626.

Mc Cormick, S.F. [1977]: The method of Kaczmarz and row-orthogonalization for solving linear equations and least-squares problems in Hilbert space, *Indiana University Mathematical Journal* 26, pp. 1137 - 1150.

Meyn, K-H [1983]: Solution of underdetermined nonlinear equations by stationary iteration, *Numerische Mathematik* 42, pp. 161-172.

Milnor, J.W. [1969]:*Topology from the differential viewpoint*, The University Press of Virginia, Charlottesville, Virginia.

Moré, J.J. [1989]: A collection of nonlinear model problems, Preprint MCS - P60 - 0289, Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, Illinois.

Nazareth, L.; Nocedal, J. [1978]: A study of conjugate gradient methods, Report SOL 78-29, Department of Operations Research, Stanford University.

Nash, S.G. [1985]: Preconditioning of Truncated Newton methods, *SIAM Journal on Scientific and Statistical Computing* 6, pp. 599 -616.

Ortega, J.M.; Rheinboldt, W.G. [1970]:*Iterative Solution of Nonlinear Equations in Several Variables*, Academic Press, NY.

Ostrowski, A.M. [1973]: *Solution of Equations in Euclidean and Banach Spaces*, Academic Press, New York.

Rall, L.B. [1984]: Differentiation in PASCAL - SC: Type Gradient, *ACM Transactions on Mathematical Software* 10, pp. 161-184.

Rall, L.B. [1987]: Optimal Implementation of Differentiation Arithmetic, in *Computer Arithmetic, Scientific Computation and Programming Languages*, U. Külisch (ed.), Teubner, Stuttgart.

Rheinboldt, W.C. [1986]: *Numerical Analysis of Parametrized Nonlinear Equations*, J. Wiley, Interscience, New York.

Saad, Y.; Schultz, M.H. [1986]: GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM Journal on Numerical Analysis* 7, pp. 856-869.

Schmidt, J.W. [1987]: A class of superlinear decomposition methods in nonlinear equations, *Numerical Functional Analysis and Optimization* 9, pp. 629 - 645.

Schmidt J.W.; Hoyer, W.; Haufe, Ch. [1985]: Consistent approximations in Newton - type decomposition methods, *Numerische Mathematik* 47, pp. 413 - 425.

Schnabel, R.B.; Frank, P.D. [1984]: Tensor methods for nonlinear equations, *SIAM Journal on Numerical Analysis* 21, pp. 815 - 843.

Shamanskii, V.E. [1967]: A modification of Newton's method,*Ukrain Mat. Z.* 19, pp. 133-138.

Schwandt, H. [1984]: An interval arithmetic approach for the construction of an almost globally convergent method for the solution of the nonlinear Poisson equation on the unit square, *SIAM Journal on Scientific and Statistical Computing* 5, pp. 427 - 452.

Schwetlick, H. [1978]: *Numerische Losung Nichtlinearer Gleichungen*, Deutscher Verlag der Wissenschaften, Berlin.

Spedicato, E. [1991] (editor): *Computer Algorithms for Solving Linear Algebraic Equations. The State of Art*, NATO ASI Series, Series F: Computer and Systems Sciences, Vol. 77, Springer Verlag, Berlin.

Spedicato, E.; Chen, Z.; Deng, N. [1992]: A class of difference ABS - type algorithms for a nonlinear system of equations, Technical Report, Department of Mathematics, University of Bergamo.

Tewarson, R.P. [1988]: A new Quasi-Newton algorithm, *Applied Mathematics Letters* 1, pp. 101 - 104.

Tewarson, R.P.; Zhang, Y. [1987]: Sparse Quasi-Newton LDU update, *International Journal on Numerical Methods in Engineering* 24, pp. 1093-1100.

Tompkins, C. [1955]: Projection methods in calculation, *Proceedings Second Symposium on Linear Programming*, Washington D.C., pp. 425 - 448.

Walker, H.F.; Watson, L.T. [1989]: Least - Change Update Methods for underdetermined systems, Research Report, Department of Mathematics, Utah State University.

Watson, L.T. [1979]: An algorithm that is globally convergent with probability one for a class of nonlinear two–point boundary value problems, *SIAM Journal on Numerical Analysis* 16, pp. 394–401.

Watson, L.T. [1980]: Solving finite difference approximations to nonlinear two-point boundary value problems by a homotopy method,*SIAM Journal on Scientific and Statistical Computing* 1, pp. 467-480.

Toint, Ph.L. [1986]: Numerical solution of large sets of algebraic nonlinear equations, *Mathematics of Computation* 16, pp. 175 - 189.

Watson, L.T. [1983]: Engineering applications of the Chow–Yorke algorithm, in *Homotopy Methods and Global Convergence* (B.C. Eaves and M.J. Todd eds.), Plenum, New York.

Watson, L.T.; Billups, S.C.; Morgan, A.P. [1987]: Algorithm 652: HOMPACK: A suite of codes for globally convergent homotopy algorithms, *ACM Trans. Math. Software* 13, pp. 281–310.

Watson, L.T.; Scott, M.R. [1987]: Solving spline-collocation approximations to nonlinear two-point boundary value problems by a homotopy method, *Applied Mathematics and Computation* 24, pp. 333-357.

Watson, L.T.; Wang, C.Y. [1981]: A homotopy method applied to elastica problems, *International Journal on Solid Structures* 17, pp. 29-37.

Wolfe, P. [1959]: The secant method for solving nonlinear equations ,*Communications ACM* 12, pp. 12 - 13.

Zambaldi, M.C. [1990]: *Estruturas estáticas e dinâmicas para resolver sistemas não lineares esparsos*, Tese de Mestrado, Departamento de Matemática Aplicada, Universidade Estadual de Campinas, Campinas, Brazil.

Zlatev, Z.; Wasniewski, J.; Schaumburg, K. [1981]: *Y12M. Solution of large and sparse systems of linear algebraic equations*, Lecture Notes in Computer Science 121, Springer-Verlag, New York, Berlin, Heidelberg and Tokyo.