AN EXTENSION OF THE THEORY OF SECANT PRECONDITIONERS [(*)]

José Mario Martínez[(**)]

Abstract. A theory of inexact Newton methods with secant preconditioners for solving large nonlinear systems of equations has been developed recently by Martínez (Mathematics of Computation, 60, pp. 681-698, 1993). According to this theory, local and superlinear convergence with bounded work per iteration of the inexact Newton method is obtained if the first trial increment at each iteration is a suitable quasi-Newton step computed using least-change secant-update procedures. The Jacobian approximation is interpreted as a preconditioner of the iterative linear method. In this paper, we extend the theory in two ways. One one hand, since in many iterative methods the true residual is not computed but the preconditioned residual is, we show how to stop the linear iteration using the preconditioned residual instead of the original one. On the other hand, we introduce damping parameters that modify the usual unitary secant step. Two natural damping parameters are introduced, one of them tries do reduce the true residual and the other one tries to reduce the preconditioned residual. We prove that the main results of the theory of secant preconditioners hold under these modifications.

Key words: Nonlinear systems, inexact Newton method, secant methods, Quasi-Newton methods.

AMS: 65H10

Date: February 7, 1994 (revision).

1. Introduction.

Newton's method is the most popular algorithm for solving differentiable systems of nonlinear equations

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

[0(*)]Work supported by FAPESP (Grants 90-3724-6), CNPq and FAEP-UNICAMP. This paper was presented at the International Meeting on Linear-Nonlinear Iterative methods held at Matsuyama, Japan, in July 1993 and was published in *Journal of Computational and Applied Mathematics* 60 (1995) 115-125.

[0(**)]Department of Applied Mathematics, University of Campinas, IMECC-UNICAMP, CP 6065, 13081 Campinas, SP Brazil. E-MAIL: MARTINEZ@CCVAX.UNICAMP.BR

where $F : \mathbb{R}^n \to \mathbb{R}^n$. We denote $J(x) \equiv F'(x)$. See Ortega and Rheinboldt [1970], Dennis and Schnabel [1983], Ostrowski [1973], Schwetlick [1978], etc. At each iteration of this method we solve the linear system

$$J(x_k)s_k = -F(x_k) \tag{1.2}$$

and we define

$$x_{k+1} = x_k + s_k. \tag{1.3}$$

When n is large and $J(x_k)$ is sparse we can use sparse LU techniques for solving (1.2). See Duff, Erisman and Reid [1986], George and Ng [1987], Gomes-Ruggiero, Martínez and Moretti [1992]. However, when n is very large and the nonzero structure of $J(x_k)$ does not help, LU techniques produce a large amount of fill-in and, so, they cannot be used for practical computations. In these cases, a good alternative is to use an iterative method for solving the linear system (1.2). Conjugate-gradient (CG) type methods are generally preferred for solving this linear system. See Hestenes and Stiefel [1952], Saad and Schultz [1986], etc.

Dembo, Eisenstat and Steihaug [1982] suggested a criterion for stopping the iterative linear method that has nice practical and theoretical properties. An increment s_k is accepted as an approximate solution of (1.2) if

$$\|J(x_k)s_k + F(x_k)\| \leq \theta_k \|F(x_k)\| \tag{1.4}$$

where $0 < \theta_k \leq \theta < 1$ for all $k \in \mathbb{N}$. Under suitable local assumptions, the "inexact Newton" method based on (1.3) and (1.4) has local linear convergence. If $\theta_k \to 0$ the convergence is superlinear.

The performance of CG methods can be improved using preconditioning techniques. See Golub and Van Loan [1989]. In this case, instead of applying the CG method to (1.2), we apply it to the equivalent preconditioned system

$$B_k^{-1} J(x_k)s = -B_k^{-1} F(x_k), \tag{1.5}$$

where B_k^{-1}, the inverse of the preconditioner, should be easy to compute and $B_k \approx J(x_k)$. Some authors (Nazareth and Nocedal [1978], Nash [1984, 1985]) use the BFGS formula (see Dennis and Schnabel [1983]) for preconditioning the Newton linear system in unconstrained minimization problems.

Martínez [1993a] proposed to use LCSU (least-change secant-update) procedures to generate the preconditioners B_k at each iteration of the inexact Newton method. See Dennis and Schnabel [1979], Dennis and Walker [1981], Martínez [1990, 1992]. Given a sequence $\theta_k \to 0$ such that $0 < \theta_k \leq \theta < 1$ for all $k \in \mathbb{N}$, Martínez's procedure consists of defining, at each iteration,

$$d_k = -B_k^{-1} F(x_k), \tag{1.6}$$

and to accept $s_k = d_k$ if

$$\|J(x_k)d_k + F(x_k)\| \leq \theta \|F(x_k)\|. \tag{1.7}$$

If d_k does not satisfy (1.7), Martínez's algorithm chooses a different s_k satisfying (1.4), using any iterative linear method.

Kozakevich, Martínez and Zambaldi [1993] used, in practical computations, a first extension of Martínez's algorithm, which consists of replacing (1.4) by

$$\|B_k^{-1}[J(x_k)s_k + F(x_k)\| \leq \theta_k \|B_k^{-1}F(x_k)\|. \tag{1.8}$$

In this paper we suggest an additional modification. We define $s_k^0 = \lambda_k d_k$ where

$$\lambda_k = \lambda_k^1 = \operatorname*{Argmin}_{\lambda \in I\!R} \|J(x_k)\lambda d_k + F(x_k)\|^2 \tag{1.9}$$

or

$$\lambda_k = \lambda_k^2 \equiv \operatorname*{Argmin}_{\lambda \in I\!R} \|B_k^{-1}[J(x_k)\lambda d_k + F(x_k)]\|^2, \tag{1.10}$$

where $\|.\| = \|.\|_2$. Consequently, the test (1.7) is replaced by

$$\|J(x_k)s_k^0 + F(x_k)\| \leq \theta \|F(x_k)\|.$$

Clearly, $\|J(x_k)\lambda_k^1 d_k + F(x_k)\| \leq \|F(x_k)\|$ and $\|B_k^{-1}[J(x_k)\lambda_k^2 d_k + F(x_k)]\| \leq \|B_k^{-1}F(x_k)\|$. So, from a practical point of view, the choice (1.9) for λ_k is associated to the stopping criterion (1.4), while (1.10) is associated to (1.8). However, the convergence analysis is not affected by these natural associations.

In Section 2 of this paper we define the new inexact Newton method with secant preconditioners and we prove the main local convergence results. In Section 3 we give some examples of potentially useful secant preconditioners. Conclusions and lines for future research are given in Section 4.

2. Main Algorithm.

Let $\{\theta_k\}$ be a sequence that converges to 0, $0 < \theta_k \leq \theta < 1$ for all $k \in I\!N$. 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 a nonsingular matrix, the steps for obtaining $x_{k+1}, k = 0, 1, 2, \ldots$ are the following:

Step 1. Compute d_k by (1.6) and $s_k^0 = \lambda_k d_k$ where $\lambda_k \in \{1, \lambda_k^1, \lambda_k^2\}$ and λ_k^1, λ_k^2 are given by (1.9) and (1.10) respectively.

Step 2. If

$$\|J(x_k)s_k^0 + F(x_k)\| \leq \theta \|F(x_k)\| \tag{2.1}$$

define $s_k = s_k^0$. Else, find an increment s_k satisfying (1.8).

Step 3. Compute $x_{k+1} = x_k + s_k$.

Local assumptions. Assume that $F : \Omega \subset I\!R^n \to I\!R^n$, Ω an open and convex set, $F \in C^1(\Omega)$, $x_* \in \Omega$, $J(x_*)$ nonsingular and $F(x_*) = 0$. Assume that there exists $L \geq 0$ such that, for all $x \in \Omega$,

$$\|J(x) - J(x_*)\| \leq L\|x - x_*\|. \tag{2.2}$$

This implies (see Broyden, Dennis and Moré [1973]) that, for all $x, z \in \Omega$,

$$\|F(z) - F(x) - J(x_*)(z - x)\| \leq L\|z - x\|\sigma(x, z) \tag{2.3}$$

where

$$\sigma(x, z) = \max\{\|x - x_*\|, \|z - x_*\|\}. \tag{2.4}$$

The first theorem shows that the basic local linear convergence theorem holds if one uses the stopping criterion (1.8), based on the preconditioned residual.

Theorem 2.1. *Suppose that the local assumptions are satisfied and that $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded. Let $\theta \in (0, 1)$. There exists $\varepsilon > 0$ such that, if $\|x_0 - x_*\| \leq \varepsilon$, the sequence defined by the Main Algorithm converges to x_* and satisfies*

$$\|x_{k+1} - x_*\|_* \leq \theta\|x_k - x_*\|_* \tag{2.5}$$

for all $k \in I\!N$, where $\|z\|_* = \|J(x_*)z\|$.

Proof. Let M be a bound of $\|B_k\|.\|B_k^{-1}\|$. Define $c_k = \theta_k M$. Clearly, $\lim_{k\to\infty} c_k = 0$. If (1.8) holds, we have that

$$\begin{aligned}
\|J(x_k)s_k + F(x_k\| &\leq \|B_k\| \, \|B_k^{-1}[J(x_k)s_k + F(x_k)]\| \\
&\leq \|B_k\|\theta_k\|B_k^{-1}F(x_k)]\| \\
&\leq \theta_k\|B_k\| \, \|B_k^{-1}\| \, \|F(x_k)\| \leq c_k\|F(x_k)\|.
\end{aligned}$$

Let $\varepsilon_1 > 0$ be such that $B(x_*, \varepsilon_1) \equiv \{x \in I\!R^n \mid \|x - x_*\| \leq \varepsilon_1\} \subset \Omega$ and $\|J(x)^{-1}\| \leq 2\|J(x_*)^{-1}\|$ for all $x \in B(x_*, \varepsilon_1)$.

Let $k_0 \in I\!N$ be such that $c_k \leq \theta$ for all $k \geq k_0$. Let us define $c = \max\{\theta, c_0, c_1, \ldots, c_{k_0}\}$. From Theorem 2.3 of Dembo, Eisenstat and Steihaug [1982] we know that there exists $\varepsilon(\theta) \in (0, \varepsilon_1)$ such that the desired result holds if $\|x_{k_0} - x_*\| \leq \varepsilon(\theta)$.

Let us prove that there exists $\varepsilon > 0$ such that $\|x_{k_0} - x_*\| \leq \varepsilon(\theta)$ holds whenever $\|x_0 - x_*\| \leq \varepsilon$. In fact, from $\|J(x_k)s_k + F(x_k)\| \leq c_k\|F(x_k)\|$, (2.1) and (2.3) we deduce that

$$\begin{aligned} \|s_k\| &\leq \|J(x_k)^{-1}\|(1 + max\{c_k, \theta\})\|F(x_k)\| \\ &\leq 2\|J(x_*)^{-1}\|(1+c)[\|J(x_*)(x_k - x_*)\| + L\|x_k - x_*\|^2]. \end{aligned}$$

So, since $\|x_{k+1} - x_*\| \leq \|s_k\| + \|x_k - x_*\|$, we have that

$$\|x_{k+1} - x_*\| \leq \beta\|x_k - x_*\|$$

where

$$\beta = 1 + 2\|J(x_*)^{-1}\|(1+c)(\|J(x_*)\| + L\varepsilon_1).$$

Therefore, if we choose $\varepsilon = \varepsilon(\theta)/\beta^{k_0}$, we obtain $\|x_{k_0} - x_*\| \leq \varepsilon(\theta)$. This completes the proof. $\square$

In the proof of Theorem 2.1 we strongly used the boundedness of the condition number of $\|B_k\|$. This hypothesis is necessary since we are not assuming that B_k is a good approximation of $J(x_k)$. (B_k is an arbitrary matrix here, rather than a preconditioner.) When B_k approximates $J(x_k)$, the criterion (1.8) is close to the affine invariant stopping criterion of Ypma [1984], for which a convergence theory, parallel to the one of Dembo, Eisenstat and Steihaug, exists. If, instead of Theorem 2.3 of Dembo, Eisenstat and Steihaug [1982], we invoke the convergence theorem of Ypma, we see that the size of the convergence neighborhood and the speed of convergence depend on the condition number of $B_k^{-1}J(x_k)$. From a technical point of view, let us observe that, under very weak assumptions, the use of LCSU preconditioners guarantees boundedness of the condition numbers (Martínez [1993a]).

In the following theorem, we show that superlinear convergence takes place if the preconditioners satisfy a Dennis-Moré condition. See Dennis and Moré [1974].

Theorem 2.2. *Suppose that F satisfies the local assumptions, $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded, the sequence $\{x_k\}$ generated by the Main Algorithm converges to x_* and*

$$\lim_{k\to\infty} \frac{\|[B_k - J(x_*)]s_k\|}{\|s_k\|} = 0. \tag{2.6}$$

Then

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0. \tag{2.7}$$

Proof. We consider two possibilities:

(i) There exists $k_0 \in I\!N$ such that for all $k \geq k_0$, the increment s_k is computed by (1.8).

(ii) For all $k_0 \in I\!N$, there exists $k > k_0$ such that $s_k = s_k^0$.

If (i) holds, we have that

$$\begin{aligned} \|J(x_k)s_k + F(x_k)\| &\leq \|B_k\| \, \|B_k^{-1}[J(x_k)s_k + F(x_k)]\| \\ &\leq \|B_k\|\theta_k\|B_k^{-1}F(x_k)\| \\ &\leq \|B_k\| \, \|B_k^{-1}\| \, \theta_k\|F(x_k)\|. \end{aligned}$$

Thus, the conditions for the superlinear convergence of the inexact Newton method of Dembo, Eisenstat and Steihaug hold. So, (2.7) is proved in this case.

Let us assume now that (ii) is true. That is, there exists an infinite set of indices K_1 such that $s_k = s_k^0$ for all $k \in K_1$. The fact that $\lim\limits_{k \notin K_1} \dfrac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0$ follows as in the case (i). So, we only need to prove that $\lim\limits_{k \in K_1} \dfrac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0$. Let us prove first that $\lim\limits_{k \in K_1} \lambda_k^1 = \lim\limits_{k \in K_1} \lambda_k^2 = 1$. Observe that $\lambda_k^1 = -\dfrac{\langle J(x_k)d_k, F(x_k)\rangle}{\|J(x_k)d_k\|^2}$. So,

$$\begin{aligned} \lambda_k^1 &= \frac{\langle J(x_k)d_k, B_k d_k\rangle}{\|J(x_k)d_k\|^2} = \frac{\langle J(x_k)s_k, B_k s_k\rangle}{\|J(x_k)s_k\|^2} \\ &= 1 + \left\langle \frac{(B_k - J(x_k))s_k}{\|J(x_k)s_k\|}, \frac{J(x_k)s_k}{\|J(x_k)s_k\|} \right\rangle. \end{aligned} \tag{2.8}$$

Now, $\|J(x_k)s_k\| \geq \|s_k\|/\|J(x_k)^{-1}\|$. Since $J(x_*)$ is nonsingular, the convergence of $\{x_k\}$ implies that there exist $\beta > 0, k_0 \in I\!N$ such that $\|J(x_k)s_k\| \geq \beta\|s_k\|$ for all $k \geq k_0$. So, by (2.6), the second term of the right hand side of (2.8) tends to zero. Therefore $\lambda_k^1 \to 1$.

Let us analyze the behavior of λ_k^2. By the boundedness of $\|B_k^{-1}\|$ we have that

$$\lim_{k\to\infty} \|[I - B_k^{-1}J(x_k)]\frac{s_k}{\|s_k\|}\| \leq \lim_{k\to\infty} \|B_k^{-1}\| \, \|(B_k - J(x_k))\frac{s_k}{\|s_k\|}\| = 0. \tag{2.9}$$

Therefore,

$$\lim_{k\to\infty} \frac{\|B_k^{-1}J(x_k)s_k\|}{\|s_k\|} = 1. \tag{2.10}$$

Now, observe that

$$\lambda_k^2 = \frac{\langle B_k^{-1} J(x_k) s_k / \|s_k\|, s_k / \|s_k\| \rangle}{\| B_k^{-1} J(x_k) s_k / \|s_k\| \, \|^2}. \tag{2.11}$$

By (2.10), the denominator of (2.11) tends to 1. Let us prove that the numerator also tends to 1. In fact, by (2.9),

$$\left| \left\langle B_k^{-1} J(x_k) \frac{s_k}{\|s_k\|}, \frac{s_k}{\|s_k\|} \right\rangle - 1 \right| = \left| \left\langle \frac{B_k^{-1} J(x_k) s_k - s_k}{\|s_k\|}, \frac{s_k}{\|s_k\|} \right\rangle \right| \leq \| [I - B_k^{-1} J(x_k)] s_k / \|s_k\| \, \|$$

and, by (2.9), the last quantity tends to zero. This completes the proof that $\lim_{k\to\infty} \lambda_k^1 = \lim_{k\to\infty} \lambda_k^2 = 1$. To prove that $\lim_{k \in K_1} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0$ we repeat the steps of the proof of Theorem 2.2 and Corollary 2.3 of Dennis and Moré [1974], using that $\lim_{k\to\infty} \lambda_k = 1$. □

In the following theorem we prove that, when the preconditioners satisfy a Dennis-Moré condition, not only the convergence is superlinear, but also we can guarantee that, for k large enough, the quasi-Newton step $-\lambda_k B_k^{-1} F(x_k)$ will satisfy the preconditioned Dembo-Eisenstat-Steihaug criterion. So, increasing accuracy in the resolution of the linear system will not be necessary.

Theorem 2.3. *Suppose that F satisfies the local assumptions and that the sequence $\{x_k\}$ generated by the Main Algorithm converges to x_*. Assume that the Dennis-Moré condition (2.6) is satisfied and that $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded. Then, there exists $k_0 \in I\!N$ such that $s_k = s_k^0$ for all $k \geq k_0$ and the convergence is superlinear.*

Proof. Using the same arguments of Theorem 2.3 of Martínez [1993a] we prove that

$$\lim_{k\to\infty} \frac{\|J(x_k) d_k + F(x_k)\|}{\|F(x_k)\|} = 0$$

Since $\lambda_k \to 1$, this implies, by (2.2) and the nonsingularity of $J(x_*)$, that

$$\lim_{k\to\infty} \frac{\|J(x_k) s_k^0 + F(x_k)\|}{\|F(x_k)\|} = 0.$$

Therefore, there exists $k_0 \in I\!N$ such that the test (2.1) is satisfied for all $k \geq k_0$. This completes the proof. □

Since the Dennis-Moré condition is a theoretical assumption, it needs to be replaced by more practical conditions. This is made in the two following theorems. Briefly speaking, we prove that local superlinear convergence and bounded work per iteration takes place if the preconditioners satisfy the classical secant equation and the difference between

two consecutive preconditioners tend to zero.

Theorem 2.4. *Suppose that F satisfies the local assumptions, $\{x_k\}$ converges to x_*, $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded, $\lim_{k\to\infty}\|B_{k+1}-B_k\|=0$ and*

$$\lim_{k\to\infty}\frac{\|[B_{k+1}-J(x_*)]s_k\|}{\|s_k\|}=0. \tag{2.12}$$

Then, the convergence is superlinear and there exists $k_0\in I\!N$ such that $s_k=s_k^0$ for all $k\geq k_0$.

Proof. It follows from (2.12) and $\|B_{k+1}-B_k\|\to 0$, using Theorems 2.1 and 2.3. □

Theorem 2.5. *Assume that F satisfies the local assumptions and that $\|B_k\|$ and $\|B_k^{-1}\|$ are bounded. Then there exists $\varepsilon>0$ such that $\{x_k\}$ converges to x_* whenever $\|x_0-x_*\|\leq\varepsilon$. If, in addition, $\lim_{k\to\infty}\|B_{k+1}-B_k\|=0$ and the secant equation*

$$B_{k+1}s_k=y_k\equiv F(x_{k+1})-F(x_k) \tag{2.13}$$

holds for all $k\in I\!N$, the convergence is superlinear and there exists $k_0\in I\!N$ such that $s_k=s_k^0$ for all $k\geq k_0$.

Proof. The convergence of $\{x_k\}$ follows from Theorem 2.1. From $x_k\to x_*$, (2.3) and (2.13), we deduce (2.12). So, the desired result is a consequence of Theorem 2.4. □

3. Some particular methods.

Martínez [1993a] proved that the conditions on the boundedness of $\|B_k\|$ and $\|B_k^{-1}\|$ and $\|B_{k+1}-B_k\|\to 0$ can be achieved if we generate the $B_k's$ using least-change secant update formulae. Our preference is to choose

$$B_k=C(x_k)+A_k \tag{3.1}$$

or

$$B_k^{-1}=C(x_k)^{-1}+A_k \tag{3.2}$$

in such that a way that $C(x)$ is a natural preconditioner of $J(x)$ and A_k is a low rank matrix such that (2.13) is satisfied. We will mention here some formulae of type (3.1) and (3.2). The first two are LCSU formulae and the last two are not.

1. Structured "Broyden-good" formula. Let us define

$$y_k^{\#}=y_k-C(x_{k+1})s_k. \tag{3.3}$$

We define

$$A_{k+1} = A_k + \frac{(y_k^{\#} - A_k s_k) s_k^T}{s_k^T s_k} \tag{3.4}$$

and

$$B_{k+1} = C(x_{k+1}) + A_{k+1} \tag{3.5}$$

for all $k \in I\!N$. By (3.3) - (3.5), the secant equation is satisfied. So, these are secant preconditioners of type (3.1). Defining

$$u_k = (y_k^{\#} - A_k s_k)/s_k^T s_k \tag{3.6}$$

we have that $A_{k+1} = u_0 s_0^T + \cdots + u_k s_k^T$ for all $k \in I\!N$. So,

$$B_{k+1} = C(x_{k+1}) + u_0 s_0^T + \cdots + u_k s_k^T.$$

Define $U_k = (u_0, \ldots, u_k),\ S_k = (s_0, \ldots, s_k)$. So,

$$B_{k+1} = C(x_{k+1}) + U_k S_k^T. \tag{3.7}$$

By the Sherman-Morrison-Woodbury formula (Golub and Van Loan [1989, p. 3]), we have that

$$B_{k+1}^{-1} = C(x_{k+1})^{-1} - C(x_{k+1})^{-1} U_k (I + S_k^T C(x_{k+1})^{-1} S_k)^{-1} S_k^T C(x_{k+1})^{-1} \tag{3.8}$$

By (3.8), the products $B_{k+1}^{-1} v$ are cheap to compute, if k is not large. However, since memory requirements grow with k, a restart procedure is necessary. A natural idea should be to restart with $B_{k+1} = C(x_{k+1})$ if $k+1$ multiple of a fixed integer m. However, in this way the secant equation would not hold at this iteration. Therefore, when $k+1 \equiv 0 \pmod{m}$ we prefer to re-define $A_k = 0$ and to compute A_{k+1}, B_{k+1} using (3.3)-(3.7). Similar procedure is adopted for restarts when using the following formulae.

2. Structured "Broyden-bad" formula. We define

$$s_k^{\#} = s_k - C(x_{k+1})^{-1} y_k, \tag{3.9}$$

$$A_{k+1} = A_k + \frac{(s_k^{\#} - A_k y_k) y_k^T}{y_k^T y_k} \tag{3.10}$$

and

$$B_{k+1}^{-1} = C(x_{k+1})^{-1} + A_{k+1} \tag{3.11}$$

for all $k \in I\!N$. Then, defining $u_k = \dfrac{(s_k^{\#} - A_k y_k)}{y_k^T y_k}$, $U_k = (u_0, \ldots, u_k)$, $Y_k = (y_0, \ldots, y_k)$, we have that

$$B_{k+1}^{-1} = C(x_{k+1})^{-1} + U_k Y_k^T. \tag{3.12}$$

3. Structured "Column-Updating" formula. This formula was motivated by the nice properties exhibited by the "Column-Updating" method in practical computations. See Martínez [1984, 1993b], Gomes-Ruggiero and Martínez [1992]. The idea is that A_{k+1} should differ from A_k in only one column. So, we define, instead of (3.4),

$$A_{k+1} = A_k + \frac{(y_k^{\#} - A_k s_k) e_{jk}^T}{e_{jk}^T s_k}, \tag{3.13}$$

where $|e_{jk}^T s_k| = \|s_k\|_\infty$. The implementation of this formula follows as in (3.5) - (3.7).

4. Structured "Inverse Column - Updating" formula. This formula is motivated by the Inverse Column-Updating method introduced by Martínez and Zambaldi [1991]. Instead of (3.10), we define

$$A_{k+1} = A_k + \frac{(s_k^{\#} - A_k y_k) e_{jk}^T}{e_{jk}^T y_k}$$

where $|e_{jk}^T y_k| = \|y_k\|_\infty$. Then B_{k+1}^{-1} is defined as in (3.11) - (3.12).

4. Final remarks.

In the implementation of inexact Newton methods for solving nonlinear systems we need to choose an iterative linear solver for (1.2). An appealing idea, when one uses secant preconditioners, is to use the secant method associated to the preconditioner as iterative linear method. Since the "first iteration" associated to (1.9) and (1.10) is a damped linear iteration, the analysis of damped secant methods for linear systems is important. Deuflhard, Freund and Walter [1990] have analyzed the behavior of damped "Broyden good" and "Broyden bad" methods for solving systems of linear equations. See Broyden [1965]. In particular, they showed that Broyden's good method, with the damping parameter that corresponds to formula (1.10) of this paper, is many times superior to GMRES, when applied to the resolution of systems that come from fine discretizations of partial differential equations. The reason for this superiority is that the preconditioned residual tends to approach the error more closely than the residual norm. Deuflhard [1991] also used Broyden's good method as iterative linear solver in his implementation of an inexact Newton method with affine invariancy properties. Their results suggest that, when we use the structured Broyden good formula to generate the preconditioners, and (1.10) as damping parameter, it is natural to use Broyden's good

method to solve the newtonian linear systems, instead of a CG method. A different implementation and convergence analysis of damped secant methods for linear systems was given by Martínez and Qi [1993]. Needless to say, the structured Broyden's good method is equivalent to Broyden's good method for linear systems. Ypma [1984] also looked at the inexact Newton method from the affine invariant point of view.

The results presented in this paper are of local nature. Global modifications of the inexact Newton method, based on the minimization of the residual norm, have been analyzed by Eisenstat and Walker [1993]. They show that any step satisfying (1.4) is a descent direction for $\|F(x)\|^2$, and that a backtracking procedure along this direction gives rise to a globally convergent algorithm. We can easily adapt their procedure to our case, when we use (1.4) and (1.9). In particular, we think that, when we globalize the method using (1.9), the probability of accepting the first step (without backtracking) increases. The situation is not so clear when we use (1.8) or (1.10).

Acknowledgements. This paper was presented at the Workshop on Linear/Nonlinear iterative methods held at Matsuyama in July 1993. Thanks are given to Prof. T. Yamamoto and to the committee who organized the meeting for hospitality and support. I am also indebted to many colleagues who participated of that meeting for interesting discussions. Finally, I acknowledge two anonymous referees for interesting comments.

References

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

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

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

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

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

DENNIS, J.E., JR.; SCHNABEL, R.B. (1983): *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*, Prentice Hall, Englewood Cliffs, N.J.

DENNIS, J.E., JR.; 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.

DEUFLHARD, P.; FREUND, R.; WALTER, A. (1990): Fast secant methods for the iterative solution of large nonsymmetric linear systems, *Impact of Computing in Science and Engineering* **2**, pp. 244-276.

DUFF, I.S.; ERISMAN, A.M.; REID, J.K. (1986): *Direct Methods for Sparse Matrices*, Clarendon Press, Oxford.

EISENSTAT, S.C.; WALKER, H.F. (1993): Globally convergent inexact Newton methods, to appear in *SIAM Journal on Optimization.*

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, *RAIRO 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.

HESTENES, M.R.; STIEFEL, E. (1952): Methods of conjugate gradients for solving linear systems, *Journal of Research of the National Bureau of Standards.* B **49**, pp. 409-436.

KOZAKEVICH, D.N.; MARTÍNEZ, J.M.; ZAMBALDI, M.C. (1993) *Experiments using secant preconditioners*, in preparation.

MARTÍNEZ, J.M. (1984): A quasi-Newton method with modification of one column per iteration, *Computing* **33**, pp. 353-362.

MARTÍNEZ, J.M. (1990): Local Convergence Theory of Inexact Newton Methods based on structured least change updates, *Mathematics of Computation* **55**, pp. 143-168.

MARTÍNEZ, J.M. (1992): On the relation between two local convergence theories of least change secant update methods, *Mathematics of Computation*, **59**, pp. 457-481.

MARTÍNEZ, J.M. (1993a): A Theory of Secant Preconditioners, *Mathematics of Computation*, **60**, pp. 681 - 698.

MARTÍNEZ, J.M. (1993b): On the convergence of the Column - Updating method, *Matemática Aplicada e Computacional* **12**, pp. 83-94.

MARTÍNEZ, J.M.; QI, L. (1993): Inexact Newton methods for solving nonsmooth equations, Relatório de Pesquisa 67/93, Instituto de Matemática, Universidade Estadual de Campinas, Brazil.

MARTÍNEZ, J.M.; ZAMBALDI, M.C. (1991): An Inverse Column-Updating method for solving large-scale nonlinear systems of equations, *Optimization Methods and Software*, 1, pp. 129–140.

NASH, S.G. (1984): Newton-type minimization via the Lanczos method, *SIAM Journal on Numerical Analysis*, **21**, pp. 770-787.

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

NAZARETH, L.; NOCEDAL, J. (1978): A study of conjugate-gradient methods, Techical Report SOL 78-29, Department of Operations Research, Stanford University.

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

OSTROWSKI, A.M. (1973): *Solution of equations in Euclidean and Banach spaces*, Academic Press, New York.

SAAD, Y.; SCHULTZ, M.H. (1986): GMRES: A generalized minimal residual algorithm for solving nonsymmetric linear systems, *SIAM Journal on Scientific and Statistical Computing*, **7**, pp. 856-869.

SCHWETLICK, H. (1978): *Numerische Lösung nichtlinearer Gleichungen*, Deutscher Verlag der Wissenschaften, Berlin.

YPMA, T.J. (1984): Local convergence of inexact Newton methods, *SIAM Journal on Numerical Analysis*, **21**, pp. 583-590.