Globally convergent inexact quasi-Newton methods for solving nonlinear systems

Ernesto G. Birgin * Nataša Krejić † José Mario Martínez ‡

November 11, 2002

Abstract

Large scale nonlinear systems of equations can be solved by means of inexact quasi-Newton methods. A global convergence theory is introduced that guarantees that, under reasonable assumptions, the algorithmic sequence converges to a solution of the problem. Under additional standard assumptions, superlinear convergence is preserved.

Key words: Nonlinear systems, inexact Newton methods, global convergence, superlinear convergence, quasi-Newton methods.

AMS classification: 65H10.

Short title: Inexact quasi-Newton methods.

1 Introduction

Newton's method is the most widely used algorithm for solving nonlinear systems of equations in real-life applications. Consider the system

$$F(x) = 0, \tag{1}$$

where $F : I\!R^n \to I\!R^n$ has continuous partial derivatives. We denote by $J(x)$ the Jacobian matrix of F for all $x \in I\!R^n$. The Newton direction d_k^N is defined, at each iteration, by

$$J(x_k) d_k^N = -F(x_k). \tag{2}$$

*Department of Computer Science IME-USP, University of São Paulo, Rua do Matão 1010, Cidade Universitária, 05508-090, São Paulo SP, Brazil. This author was supported by PRONEX-Optimization 76.79.1008-00 and FAPESP (Grants 99/08029-9 and 01/04597-4). e-mail: egbirgin@ime.usp.br

†Institute of Mathematics, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad, Yugoslavia. This author was supported by a visiting professor grant of FAPESP (Grant 99/03102-0). e-mail: natasa@unsim.ns.ac.yu

‡Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This author was supported by PRONEX-Optimization 76.79.1008-00, FAPESP (Grant 01/04597-4), CNPq and FAEP-UNICAMP. e-mail: martinez@ime.unicamp.br

If the Jacobian matrix is Lipschitz-continuous and nonsingular at a solution of the system, the iteration (2) defines a locally and quadratically convergent method which, moreover, is invariant under linear transformations both in the range and in the domain space. In order to generate a globally convergent algorithm, d_k^N is used as search direction and the Newtonian iteration takes the form

$$x_{k+1} = x_k + \alpha_k d_k,$$

where $\alpha_k > 0$ is such that $\|F(x)\|$ is sufficiently reduced. The reduction is possible because d_k^N is a descent direction of the merit function.

Solving (2) at each iteration can be expensive. On one hand, one has to compute the Jacobian at every iteration and, on the other hand, a linear system must be solved exactly. These drawbacks motivated the development of quasi-Newton methods and inexact Newton methods in the last three decades.

In inexact Newton methods [4, 8] the computation of the Newtonian direction (2) is replaced by

$$\|J(x_k)d_k^{IN} + F(x_k)\| \le \theta_k \|F(x_k)\| \tag{3}$$

where $0 \le \theta_k < \theta < 1$ for all $k = 0, 1, 2, \ldots$. Usually, d_k^{IN} is obtained by applying some iterative linear solver to the system $J(x_k)d = -F(x_k)$.

In quasi-Newton methods the direction is computed by solving

$$B_k d_k^{QN} = -F(x_k), \tag{4}$$

where, in general, B_k is not the Jacobian. In many quasi-Newton methods the matrices B_k are generated in such a way that the linear algebra involved in (4) is minimal. See [6, 11, 18] and references therein. The most straightforward way of doing this is to keep the same Jacobian during some iterations, so that no new factorizations are needed during some steps. This is generally called the Shamanski Method [24].

The combination of the ideas (3) and (4) leads to inexact quasi-Newton methods. See, for example, [22]. In this case, the search directions are computed by

$$\|B_k d_k^{IQN} + F(x_k)\| \le \theta_k \|F(x_k)\| \tag{5}$$

where $0 \le \theta_k < \theta < 1$ for all $k = 0, 1, 2, \ldots$.

Global convergence of methods for solving nonlinear equations is usually obtained using the residual norm as merit function. See [8, 12] and many others. Unfortunately, in (4) and (5) it is not possible to ensure that the generated directions are descent directions for any norm.

In this paper we suggest how that difficulty can be overcome. On one hand, the inexact Newton condition (3) is imposed sufficiently often. On the other hand, we use a nonmonotone technique which is similar to the one introduced by Li and Fukushima [13] for proving global convergence of Broyden's method and for proving convergence of algorithms for some nonsmooth problems. The main difference is that the amount of reduction required at each iteration is proportional to the residual norm in our method, which is more adequate than the use of the squared norm of the increment for scaling reasons.

The paper is organized as follows. In Section 2 we describe the general algorithm and we give global convergence results. In Section 3 we show that local superlinear convergence arises from updatings with bounded deterioration. We draw some conclusions in Section 4.

2 Model algorithm and convergence

Assume that $F : I\!R^n \to I\!R^n$, $F \in C^1(I\!R^n)$. From now on, $\|\cdot\|$ denotes an arbitrary norm. Assume that $\theta \in [0, 1)$, $\sigma \in (0, 1)$, $0 < \tau_{min} < \tau_{max} < 1$ and $\{\eta_k\}$ is a sequence such that $\eta_k > 0$ for all $k = 0, 1, 2, \ldots$ and $\sum_{k=0}^{\infty} \eta_k = \eta < \infty$. Finally, let $x_0 \in I\!R^n$ be an initial approximation for the solution of $F(x) = 0$.

Given $x_k \in I\!R^n$, the $k-$th iterate of the algorithm, the steps for obtaining x_{k+1} are given in Algorithm 1.

Algorithm 1. (Model Algorithm)

Step 1. (Compute the search direction)
Compute $d_k \in I\!R^n$.

Step 2. (Backtracking)
Step 2.1. Set $\alpha \leftarrow 1$.
Step 2.2. If

$$\|F(x_k + \alpha d_k)\| \le [1 + \alpha\sigma(\theta - 1)]\|F(x_k)\| + \eta_k, \tag{6}$$

set $\alpha_k = \alpha$ and

$$x_{k+1} = x_k + \alpha_k d_k. \tag{7}$$

If (6) does not hold, compute $\alpha_{new} \in [\tau_{min}\alpha, \tau_{max}\alpha]$, set $\alpha \leftarrow \alpha_{new}$ and repeat Step 2.2.

Remark. Since $\eta_k > 0$ the condition (6) is satisfied for $\alpha > 0$ sufficiently small. So, the backtracking process at Step 2 is necessarily completed at every iteration independently of the choice of d_k. Therefore, the iteration is always well defined.

Lemma 1. *Assume that $\{x_k\}$ is a sequence generated by Algorithm 1. If, for some sequence of indices $K_0 \subset \{0, 1, 2, \ldots\}$, $\lim_{k\in K_0} F(x_k) = 0$, then*

$$\lim_{k\to\infty} F(x_k) = 0. \tag{8}$$

In particular, if x_* is a limit point of $\{x_k\}$ such that $F(x_*) = 0$ then every limit point of the sequence is a solution.

Proof. Let $\varepsilon > 0$ be arbitrary. Let $k \in K_0$ be such that $\|F(x_k)\| \le \varepsilon/2$ and

$$\sum_{\ell=k}^{\infty} \eta_\ell \le \varepsilon/2.$$

Observe that $0 \leq 1 + \alpha\sigma(\theta - 1) < 1$. So, by (6),

$$\|F(x_j)\| \leq \|F(x_{j-1})\| + \eta_{j-1} \ \ \forall\, j \geq 1.$$

Therefore, if $j > k$ we have that

$$\|F(x_j)\| \leq \|F(x_k)\| + \eta_k + \ldots + \eta_{j-1} \leq \|F(x_k)\| + \sum_{\ell=k}^{\infty} \eta_\ell \leq \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon.$$

For the second part of the proof observe that if x_* is a limit point such that $F(x_*) = 0$ then the sequence K_0 satisfying $\lim_{k \in K_0} F(x_k) = 0$ necessarily exists. Thus, the proof follows from the first part. $\square$

The following lemma will be used in the proof of Theorem 3.

Lemma 2. *Assume that $\{x_k\}$ is a sequence generated by Algorithm 1 and that all the limit points of the sequence $\{x_k\}$ are solutions of (1). Assume, further, that x_* is a limit point such that $J(x_*)$ is nonsingular and*

$$\lim_{k \to \infty} \|x_{k+1} - x_k\| = 0. \tag{9}$$

Then, the whole sequence converges to x_*.

Proof. Since $J(x_*)$ is nonsingular, the inverse function theorem guarantees that there exists $\delta > 0$ such that $\|F(x)\| > 0$ whenever $0 < \|x - x_*\| \leq \delta$.

Let $\varepsilon \in (0, \delta)$ be arbitrary.

The set $\{x \in I\!R^n \ \mid \ \varepsilon \leq \|x - x_*\| \leq \delta\}$ does not contain any solution of the system. Therefore, it does not contain any limit point and so, it can contain only a finite number of iterates. So, there exists k_0 such that, for all $k \geq k_0$,

$$x_k \notin \{x \in I\!R^n \ \mid \ \varepsilon \leq \|x - x_*\| \leq \delta\}. \tag{10}$$

Let $k_1 \geq k_0$ be such that

$$\|x_{k+1} - x_k\| < \delta - \varepsilon$$

for all $k \geq k_1$.

Since x_* is a limit point, there exists $k \geq k_1$ such that

$$\|x_k - x_*\| < \varepsilon.$$

So,

$$\|x_{k+1} - x_*\| \leq \|x_{k+1} - x_k\| + \|x_k - x_*\| < \delta.$$

But, since $k \geq k_0$, by (10) we have that

$$\|x_{k+1} - x_*\| < \varepsilon.$$

Repeating this argument, we obtain that

$$\|x_{k+j} - x_*\| \leq \varepsilon$$

for all $j = 1, 2, 3, \ldots$. Since ε was arbitrary, this implies that the sequence converges to x_*. $\square$

Theorem 1. *Assume that $\{x_k\}$ is generated by Algorithm 1 and there exists $M > 0$ such that, for an infinite sequence of indices $K_1 \subset \{0, 1, 2, \ldots\}$,*

$$\|J(x_k)d_k + F(x_k)\| \leq \theta \|F(x_k)\| \tag{11}$$

and

$$\|d_k\| \leq M. \tag{12}$$

Then any limit point of the subsequence $\{x_k\}_{k \in K_1}$ is a solution of the system. Moreover, if a limit point of $\{x_k\}_{k \in K_1}$ exists, then $F(x_k) \to 0$ and every limit point of $\{x_k\}$ is a solution of (1).

Proof. Suppose $K_2 \subset K_1$ is a sequence of indices such that

$$\lim_{k \in K_2} x_k = x_*. \tag{13}$$

Let us consider first the case in which $\{\alpha_k\}_{k \in K_2}$ does not tend to 0. So, there exists a sequence of indices $K_3 \subset K_2$ and $\bar{\alpha} > 0$ such that

$$\alpha_k \geq \bar{\alpha} > 0$$

for all $k \in K_3$. Therefore, by (6), for all $k \in K_3$,

$$\|F(x_{k+1})\| \leq \|F(x_k)\| + \bar{\alpha}\sigma(\theta - 1)\|F(x_k)\| + \eta_k.$$

But, for all $k \notin K_3$, $\|F(x_{k+1})\| \leq \|F(x_k)\| + \eta_k$. Adding all these inequalities, we obtain that

$$\sigma\bar{\alpha}(1 - \theta) \sum_{k \in K_3} \|F(x_k)\| \leq \|F(x_0)\| + \sum_{k=0}^{\infty} \eta_k = \|F(x_0)\| + \eta.$$

Therefore, $\lim_{k \in K_3} F(x_k) = 0$. So, $F(x_*) = 0$.

Now, let us consider the case in which

$$\lim_{k \in K_2} \alpha_k = 0. \tag{14}$$

By the choice of α_{new}, for $k \in K_2$ large enough, there exists $\alpha'_k > \alpha_k$, $\alpha'_k \in [\alpha_k/\tau_{max}, \alpha_k/\tau_{min}]$ such that

$$\lim_{k \in K_2} \alpha'_k = 0 \tag{15}$$

and

$$\|F(x_k + \alpha'_k d_k)\| > \|F(x_k)\| + \alpha'_k \sigma(\theta - 1)\|F(x_k)\| + \eta_k.$$

So,

$$\|F(x_k + \alpha_k' d_k)\| > [1 + \alpha_k' \sigma(\theta - 1)]\|F(x_k)\|. \tag{16}$$

Therefore,

$$\|F(x_k + \alpha_k' d_k) - [F(x_k) + J(x_k)\alpha_k' d_k]\|$$
$$+\|F(x_k) + J(x_k)\alpha_k' d_k\| > [1 + \alpha_k' \sigma(\theta - 1)]\|F(x_k)\|.$$

So,

$$\|F(x_k + \alpha_k' d_k) - F(x_k) - J(x_k)\alpha_k' d_k]\|$$
$$+\|\alpha_k'[F(x_k) + J(x_k)d_k]\| + (1 - \alpha_k')\|F(x_k)\| > [1 + \alpha_k' \sigma(\theta - 1)]\|F(x_k)\|.$$

Then, by (11),

$$\|F(x_k + \alpha_k' d_k) - F(x_k) - J(x_k)\alpha_k' d_k]\|$$
$$+\alpha_k' \theta\|F(x_k)\| + (1 - \alpha_k')\|F(x_k)\| > [1 + \alpha_k' \sigma(\theta - 1)]\|F(x_k)\|.$$

This implies, after some algebraic manipulation, that

$$\alpha_k'\|F(x_k)\|(1 - \sigma)(1 - \theta) < \|F(x_k + \alpha_k' d_k) - F(x_k) - J(x_k)\alpha_k' d_k]\|. \tag{17}$$

So,

$$\|F(x_k)\|(1 - \sigma)(1 - \theta) < \frac{\|F(x_k + \alpha_k' d_k) - F(x_k) - J(x_k)\alpha_k' d_k]\|}{\alpha_k'}.$$

Since $\|d_k\|$ is bounded and α_k' tends to zero, the continuity of the derivatives of F imply that the right-hand side of the above inequality tends to zero for $k \in K_2$. Therefore, $\lim_{k \in K_2} \|F(x_k)\| = 0$ and, so, $F(x_*) = 0$.

The second part of the proof follows from Lemma 1. □

3 Bounded-deterioration updates

In this section we want to analyze algorithms in which the search direction satisfies the requirement

$$\|B_k d_k + F(x_k)\| \le \theta\|F(x_k)\|, \tag{18}$$

where, periodically, $B_k = J(x_k)$. The matrices B_k will satisfy a "weak bounded deterioration"(WBD) property. Useful quasi-Newton methods that satisfy the WBD property have been introduced in the literature [10, 14, 17, 21] and a recent study by Lukšan and Vlcék [15] seems to indicate that one of them [21] is the most efficient method for many large-scale problems. A plausible conjecture is that these methods do not possess "local convergence without restarts" although a counterexample has not yet been published. Therefore, it is natural to analyze their globalization under a Jacobian restart strategy.

From now on we denote $s_k = x_{k+1} - x_k$. The WBD property for a sequence of Jacobian approximations B_k states that there exist $c_1, c_2 > 0$ (independent of k) such that, for all $k = 0, 1, 2, \ldots$,

$$\|B_{k+1} - J(x_{k+1})\| \le c_1 \|B_k - J(x_k)\| + c_2 \|s_k\|. \tag{19}$$

Many quasi-Newton updates satisfy this property. For example, consider quasi-Newton methods of rank-one secant type that obey the formula

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) v_k^T}{v_k^T s_k},$$

where $|v_k^T s_k| \ge \gamma \|v_k\| \|s_k\|$ for some $\gamma > 0$ and $y_k = F(x_{k+1}) - F(x_k)$ for all k. Multipoint secant methods [3, 9, 16] usually satisfy similar recurrence equations. If $J(x)$ satisfies a Lipschitz condition

$$\|J(z) - J(x)\| \le L\|z - x\| \text{ for all } \quad x, z \in I\!R^n, \tag{20}$$

it can be proved (see [6], p. 75) that

$$\|y_k - J(x_k) s_k\| \le \frac{L}{2} \|s_k\|^2 \text{ for all } \ k = 0, 1, 2, \ldots. \tag{21}$$

Therefore,

$$B_{k+1} - J(x_{k+1}) = B_k - J(x_k) - J(x_{k+1}) + J(x_k) + \frac{(y_k - B_k s_k) v_k^T}{v_k^T s_k}$$

$$= J(x_k) - J(x_{k+1}) + B_k - J(x_k) + \frac{(J(x_k) s_k - B_k s_k) v_k^T}{v_k^T s_k} + \frac{(y_k - J(x_k) s_k) v_k^T}{v_k^T s_k}$$

$$= J(x_k) - J(x_{k+1}) + (B_k - J(x_k))(I - \frac{s_k v_k^T}{v_k^T s_k}) + \frac{(y_k - J(x_k) s_k) v_k^T}{v_k^T s_k}.$$

So,

$$\|B_{k+1} - J(x_{k+1})\| \le L\|s_k\| + \|B_k - J(x_k)\|(1 + \frac{\|s_k\| \|v_k\|}{|v_k^T s_k|}) + \frac{L\|s_k\|^2 \|v_k\|}{2|v_k^T s_k|}$$

$$\le (1 + 1/\gamma)\|B_k - J(x_k)\| + (L + L/(2\gamma))\|s_k\|.$$

Therefore, (19) holds.

Among the rank-one quasi-Newton formulae that satisfy the bounded deterioration condition (19) we can cite Broyden's good and bad methods [2, 5, 20], the Column-Updating method [10, 17] and the Inverse Column-Updating method [14, 21]. Quasi-Newton methods that are not of rank-one type can be found in [1, 23].

Theorem 2. *Assume that*

1. *F is uniformly continuous in $I\!R^n$;*
2. *$\{x_k\}$ is generated by Algorithm 1;*

3. $K = \{k_1, k_2, \ldots\} \subset \{0, 1, 2, \ldots\}$ *is such that* $k_j < k_{j+1} \le k_j + m$ *for all* $j \in \{0, 1, 2, \ldots\}$*;*

4. $B_k = J(x_k)$ *for all* $k \in K$*;*

5. $J(x)$ *is nonsingular and* $\|J(x)^{-1}\| \le M$ *for all* $x \in I\!R^n$*.*

6. The search direction d_k *satisfies the inexact quasi-Newton equation*

$$\|B_k d_k + F(x_k)\| \le \theta \|F(x_k)\|, \tag{22}$$

whenever B_k *is nonsingular (If* B_k *is singular and (22) does not hold, we take* $d_k = 0$*);*

7. The matrices B_k *satisfy the WBD property for fixed constants* $c_1, c_2 > 0$*.*

Then, either $\lim_{k \in K} \|x_k\| \to \infty$ *or there exists* $x_* \in I\!R^n$ *such that* $x_k \to x_*$*,* $B_k \to J(x_*)$*,* $F(x_*) = 0$ *and the search direction satisfies (22) for* k *sufficiently large.*

Proof. If $\|x_k\|$ does not tend to ∞ for $k \in K$ then $\{x_k\}_{k \in K}$ admits a bounded subsequence. So, $\{x_k\}_{k \in K}$ admits a limit point $x_* \in I\!R^n$.

Observe that, if B_k is nonsingular, then (22) implies that

$$\begin{aligned} \|d_k\| = \|B_k^{-1} B_k d_k\| &= \|B_k^{-1}[B_k d_k + F(x_k)] - B_k^{-1} F(x_k)\| \\ &\le \|B_k^{-1}\| \theta \|F(x_k)\| + \|B_k^{-1}\| \|F(x_k)\| \\ &= \|B_k^{-1}\| (\theta + 1) \|F(x_k)\|. \end{aligned} \tag{23}$$

Therefore, for all $k \in K$, $\|d_k\| \le 2M \|F(x_k)\|$. Thus, the assumptions of Theorem 1 hold and, consequently,

$$\lim_{k \to \infty} \|F(x_k)\| = 0$$

and every limit point is a solution. Let us now prove by induction on j that

$$\lim_{k \in K} \|B_{k+j} - J(x_{k+j})\| = \lim_{k \in K} \|B_{k+j} - J(x_k)\| = 0. \tag{24}$$

and

$$\lim_{k \in K} \|x_{k+j+1} - x_{k+j}\| = 0. \tag{25}$$

For $j = 0$, (24) follows from the definition of K. Since $\|J(x_k)^{-1}\|$ is bounded for $k \in K$, (25) follows from (23).

Assume that (24) and (25) hold for some j. Then, by the WBD property,

$$\|B_{k+j+1} - J(x_{k+j+1})\| \le c_1 \|B_{k+j} - J(x_{k+j})\| + c_2 \|x_{k+j+1} - x_{k+j}\|.$$

Therefore, by the inductive hypothesis,

$$\lim_{k \in K} \|B_{k+j+1} - J(x_{k+j+1})\| = 0.$$

But, by (20) and (25),

$$\lim_{k\in K} \|J(x_{k+j+1}) - J(x_{k+j})\| = 0,$$

therefore,

$$\lim_{k\in K} \|B_{k+j+1} - J(x_{k+j})\| = 0. \tag{26}$$

Finally,

$$\|B_{k+j+1} - J(x_k)\| \le \|B_{k+j+1} - J(x_{k+j})\| + \|B_{k+j} - J(x_{k+j})\| + \|B_{k+j} - J(x_k)\|,$$

so, by (26) and the inductive hypothesis,

$$\lim_{k\in K} \|B_{k+j+1} - J(x_k)\| = 0$$

as we wanted to prove. This implies that $\{\|B_{k+j+1}^{-1}\|\}$ is bounded for k large enough. So, by (23),

$$\lim_{k\in K} \|x_{k+j+2} - x_{k+j+1}\| = 0.$$

Therefore, (24) and (25) are proved for all $j = 0, 1, 2, \ldots$.

Let us prove now that

$$\lim_{k\to\infty} \|x_{k+1} - x_k\| = 0. \tag{27}$$

Suppose, by contradiction, that this is not true. Then, there exists a sequence of indices $K_4 \subset \{0, 1, 2, \ldots\}$ such that $\|x_{k+1} - x_k\|$ is bounded away from zero for $k \in K_4$. Now, by the definition of K, each $k \in K_4$ is of the form $k = \ell(k) + j(k)$ for some $\ell(k) \in K$ and $j(k) \le m$. Therefore, for infinitely many indices in the sequence K_4, $j(k)$ is equal to the same integer $\bar{j} \le m$.

So, $\|x_{\ell+\bar{j}+1} - x_{\ell+\bar{j}}\|$ is bounded away from zero for infinitely many indices ℓ contained in K. This contradicts the fact that (25) holds for each fixed j.

By (27), we are under the hypothesis of Lemma 2. Therefore, the whole sequence converges as we wanted to prove. The fact that B_k converges to $J(x_*)$ follows from (24). This implies that (22) holds for k large enough. $\square$

Theorem 3. *Assume that the hypotheses of Theorem 2 hold and that $\|x_k\|$ does not tend to ∞. Assume that*

$$\|B_k d_k + F(x_k)\| \le \theta_k \|F(x_k)\| \tag{28}$$

for all k large enough, where $\lim_{k\to\infty} \theta_k = 0$.

Then, $\{x_k\}$ converges superlinearly to a solution of $F(x) = 0$.

Proof. By Theorem 2, the sequence converges to a solution x_* and $B_k \to J(x_*)$. By (23) we have that, for k large enough,

$$\|d_k\| \le 2\|J(x_*)^{-1}\|\|F(x_k)\|. \tag{29}$$

By the uniform continuity of $J(x)$ we have that

$$\|F(x_k + d_k) - F(x_k) - J(x_k)d_k\| \le o(\|d_k\|)$$

for all $k = 0, 1, 2, \ldots$. Therefore, by (28) and (29),

$$\|F(x_k + d_k)\| \le \|F(x_k) + J(x_k)d_k\| + o(\|d_k\|)$$

$$\le \|F(x_k) + B_k d_k\| + \|B_k - J(x_k)\|\|d_k\| + o(\|d_k\|)$$

$$\le \theta_k \|F(x_k)\| + 2\|B_k - J(x_k)\|\|J(x_*)^{-1}\|\|F(x_k)\| + o(\|F(x_k)\|).$$

Since $B_k \to J(x_*)$ and $\|F(x_k)\| \to 0$, this implies that

$$\lim_{k\to\infty} \frac{\|F(x_k + d_k)\|}{\|F(x_k)\|} = 0. \tag{30}$$

So, for k large enough,

$$\|F(x_k + d_k)\| \le (1 - \sigma)\|F(x_k)\|.$$

Therefore, (6) holds with $\alpha = 1$. So, for k large enough, $x_{k+1} = x_k + d_k$. Then, by (30),

$$\lim_{k\to\infty} \frac{\|F(x_{k+1})\|}{\|F(x_k)\|} = 0. \tag{31}$$

By the nonsingularity of $J(x_*)$ there exist constants $c_{small}, c_{big} > 0$ such that

$$c_{small}\|x - x_*\| \le \|F(x)\| \le c_{big}\|x - x_*\|$$

for all x in a neighborhood of x_*. Then, by (31),

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

as we wanted to prove. □

4 Conclusions

Quasi-Newton methods are important tools for solving nonlinear simultaneous equations. They are especially useful when the evaluation of the Jacobian matrix (or a part of it) is difficult, cumbersome or impossible, at least analytically. When the system has many variables, direct resolution of the linear system that arises at each iteration can be impractical and, so, its inexact resolution using iterative linear methods is usually preferred. These are the main motivations for the development of inexact quasi-Newton methods.

The main difficulty for proving global convergence of quasi-Newton methods is that the directions generated are not, in general, descent directions for the norm of the system. Up to our knowledge, only for Broyden's method a global convergence theory has been given where this difficulty is satisfactory circumvented. See [13].

However, there are many other quasi-Newton methods which are potentially useful in practical situations. For example, structured quasi-Newton methods are used when some entries of the Jacobian matrix are easy to evaluate but others are not. The local convergence of these methods has been exhaustively analyzed in [6] (Cap. 11), [7, 19]. It does not seem to be possible to provide global versions of these methods that do not use Jacobian information in some sense.

Structured quasi-Newton updates can be very useful and are usually suggested by the structure of the system. Sometimes it is not possible to prove that these structured updates satisfy least change variational principles as defined, for example, in [6] (Cap. 11). However, weak bounded deterioration is usually satisfied. For example, in nonlinear programming problems the Hessian of the augmented Lagrangian has a structure that inspires the approximation of a term of the form $\sum h_i(x)\nabla^2 h_i(x)$ by a symmetric matrix that is close to the null matrix when $|h_i(x)|$ is small for all i. Several useful possibilities that satisfy WBD can be suggested.

Superlinear convergence is preserved if one uses updates that satisfy a weak bounded deterioration property. Many practical algorithms, including most structured quasi-Newton methods, satisfy this property, so that suitable globally and locally convergent algorithms can be easily devised.

Acknowledgements.

The authors are indebted to two anonymous referees for their careful reading of this paper. Due to the suggestion of one of them, we eliminated an unnecessary periodicity hypothesis in Theorem 1 and we improved other proofs, leading to a considerable simplification of the material.

References

[1] I. D. L. Bogle and J. D. Perkins, *A new sparsity-preserving quasi-Newton update for solving nonlinear equations*, SIAM Journal on Scientific and Statistical Computing 11 (1990), pp. 621–630.

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

[3] O. Burdakov, *On superlinear convergence of some stable variants of the secant method*, ZAMM 66 (1986), pp. 615–622.

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

[5] J. E. Dennis Jr. and J. J. Moré, *A characterization of superlinear convergence and its applications to quasi-Newton methods*, SIAM Review 19 (1977), pp. 46–89.

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

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

[8] S. C. Eisenstat and H. F. Walker, *Globally convergent inexact Newton methods*, SIAM Journal on Optimization 4 (1994), pp. 392-422.

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

[10] M. A. Gomes-Ruggiero and J. M. Martínez, *The column-updating method for solving nonlinear equations in Hilbert space*, RAIRO Mathematical Modelling and Numerical Analysis 26 (1992), pp. 309–330.

[11] C. T. Kelley, *Iterative methods for linear and nonlinear equations*, SIAM Publications, Philadelphia, 1995.

[12] N. Krejić and J. M. Martínez, *A globally convergent inexact-Newton method for solving reducible nonlinear systems of equations*, Optimization methods and Software 13 (2000), pp. 11-34.

[13] Li, Dong-Hui and M. Fukushima, *Derivative-Free Line Search and Global Convergence of Broyden-Like Method for Nonlinear Equations*, Optimization Methods and Software 13 (2000), pp. 181–201.

[14] V. L. R. Lopes and J. M. Martínez, *Convergence properties of the inverse Column-Updating method*, Optimization Methods and Software 6 (1995), pp. 127–144.

[15] L. Lukšan, and J. Vlček, *Computational experience with globally convergent descent methods for large sparse systems of nonlinear equations*, Optimization Methods and Software 8 (1998), pp. 185–199.

[16] J. M. Martínez, *Three new algorithms based on the sequential secant method*, BIT 19 (1979), pp. 236–243.

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

[18] J. M. Martínez, *Algorithms for solving nonlinear systems of equations*, in "Continuous Optimization. The state of art", edited by E. Spedicato, Kluwer Academic Publishers (1994), pp. 81-108.

[19] J. M. Martínez, *Local convergence theory of inexact Newton methods based on structured least change updates* Mathematics of Computation 55 (1990), pp. 143-167.

[20] J. M. Martínez, *Practical quasi-Newton methods for solving nonlinear systems*, Journal of Computational and Applied Mathematics 124 (2000), pp. 97-122.

[21] J. M. Martínez and M. C. Zambaldi, *An inverse column-updating method for solving large scale nonlinear systems of equations*, Optimization Methods and Software 1 (1992), pp. 129–140.

[22] I. Moret, *On the convergence of inexact quasi-Newton methods*, International Journal of Computer Mathematics 28 (1989), pp. 117-137.

[23] L. K. Schubert, *Modification of a quasi-Newton method for nonlinear equations with a sparse Jacobian*, Mathematics of Computation 24 (1970) pp. 27-30.

[24] V. E. Shamanski, *A modification of Newton's method*, Ukrain Mat. Z. 19 (1967), pp. 133–138.