Practical quasi-Newton methods for solving nonlinear systems

José Mario Martínez*

January 24, 2000

Abstract

Practical quasi-Newton methods for solving nonlinear systems are surveyed. The definition of quasi-Newton methods that includes Newton's method as a particular case is adopted. However, especial emphasis is given to the methods that satisfy the secant equation at every iteration, which are called here, as usually, secant methods. The least-change secant update (LCSU) theory is revisited and convergence results of methods that do not belong to the LCSU family are discussed. The family of methods reviewed in this survey includes Broyden's methods, structured quasi-Newton methods, methods with direct updates of factorizations, row-scaling methods and column-updating methods. Some implementation features are commented. The survey includes a discussion on global convergence tools and linear-system implementations of Broyden's methods. In the final section, practical and theoretical perspectives of this area are discussed.

1 Introduction

In this survey we consider nonlinear systems of equations

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

where $F : \mathbb{R}^n \to \mathbb{R}^n$ has continuous first partial derivatives. We denote $F = (f_1, \dots, f_n)^T$ and $J(x) = F'(x)$ for all $x \in \mathbb{R}^n$.

*Institute of Mathematics, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil. This work was supported by PRONEX-Optimization Project, FAPESP (grant 90-3724-6), FINEP, CNPq, FAEP-UNICAMP.

Problem (1) is a particular case of the problem of minimizing $\|F(x)\|_2^2$. However, special methods are far more efficient than minimization and nonlinear least-squares methods for solving this problem, especially when n is large.

All practical algorithms for solving (1) are iterative. Given an initial approximation $x_0 \in I\!R^n$, a sequence of iterates $x_k, k = 0, 1, 2, \ldots$, is generated in such a way that, hopefully, the approximation to some solution is progressively improved. Newton's is the most widely used method in applications. See [27, 55, 81, 94]. The Newtonian iteration is defined whenever $J(x_k)$ is nonsingular. In this case, the iterate that follows x_k is given by

$$x_{k+1} = x_k - J(x_k)^{-1}F(x_k). \tag{2}$$

The Jacobian inverse $J(x_k)^{-1}$ does not need to be calculated. Instead, $s_k \in I\!R^n$ results from solving

$$J(x_k)s_k = -F(x_k) \tag{3}$$

and the new iterate is defined by

$$x_{k+1} = x_k + s_k. \tag{4}$$

Newton's method has very attractive theoretical and practical properties: if x_* is a solution of (1) at which $J(x_*)$ is nonsingular and x_0 is close enough to x_*, then x_k converges superlinearly to x_*. This means that, given an arbitrary norm $\|\cdot\|$ in $I\!R^n$,

$$\lim_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} = 0. \tag{5}$$

Moreover, if $J(x)$ satisfies the Lipschitz condition

$$\|J(x) - J(x_*)\| \le L\|x - x_*\| \tag{6}$$

for all x close enough to x_*, the convergence is quadratic, so the error at iteration $k+1$ is proportional to the square of the error at iteration k. In other words, the number of correct digits of the approximation x_{k+1} tends to double the number of correct digits of x_k.

Another remarkable property of Newton's method is its invariancy with respect to linear transformations both in the range-space and in the domain space. Invariancy in the range-space means that, given any nonsingular matrix A, the iterates of the method applied to

$$AF(x) = 0$$

coincide with the iterates of the method applied to (1). Domain-space invariancy means that the iterates of the method applied to

$$F(Ay) = 0$$

are given by $A^{-1}x_k$, provided that $y_0 = A^{-1}x_0$, where $\{x_k\}$ is the sequence generated by (2). The main consequence of invariancy is that bad scaling of the variables or the components of the system cannot affect the performance of the method, if rounding errors (which can affect the quality of the solution of (3)) are disregarded.

The Newton iteration can be costly, since partial derivatives must be computed and the linear system (3) must be solved at every iteration. This fact motivated the development of quasi-Newton methods, which are defined as the generalizations of (2) given by

$$x_{k+1} = x_k - B_k^{-1}F(x_k). \tag{7}$$

In quasi-Newton methods, the matrices B_k are intended to be approximations of $J(x_k)$. In many methods, the computation of (7) does not involve computing derivatives at all. Moreover, in many particular methods, B_{k+1}^{-1} is obtained from B_k^{-1} using simple procedures thanks to which the linear algebra cost involved in (7) is much less than the one involved in (3).

According to definition (7), Newton's method is a quasi-Newton method. So is the *stationary Newton* method, where $B_k = J(x_0)$ for all $k = 0, 1, 2, \ldots$ and Newton's method "with p refinements", in which $B_k = J(x_k)$ when k is a multiple of $p + 1$, whereas $B_k = B_{k-1}$ otherwise. The "discrete Newton" method is a quasi-Newton method too. It consists in defining

$$B_k = \left(\frac{F(x_k + h_{k,1}e_1) - F(x_k)}{h_{k,1}}, \ldots, \frac{F(x_k + h_{k,n}e_n) - F(x_k)}{h_{k,n}} \right) \tag{8}$$

where $\{e_1, \ldots, e_n\}$ is the canonical basis of $I\!R^n$ and $h_{k,j} \neq 0$ is a discretization parameter. This parameter must be small enough so that the difference approximation to the derivatives is reliable but large enough so that rounding errors in the differences (8) are not important.

In many problems, $J(x)$ is a sparse matrix, whose sparsity pattern is known. In this case, a procedure given in [20] and refined in [18] (see also [17, 58]) allows one to compute a finite difference approximation to $J(x)$ using less than n auxiliary functional evaluations. When the Jacobian matrix is dense, the discrete Newton method is not competitive with the cheap-linear-algebra versions of (7). But, in many large sparse problems, discrete Newton implementations are quite effective. In these cases, the finite difference technique allows one to compute the approximate Jacobian using a small number of functional evaluations and the matrix structure is such that factorization is not expensive.

In the sixties it was common to justify the existence of most quasi-Newton methods saying that the task of computing derivatives is prone to human errors. However, automatic differentiation techniques have been developed in the last 20 years that, in practice, eliminates the possibility of error. See [31, 45, 50, 87, 88, 89] and many others. Moreover, in most cases, the computation of derivatives using automatic differentiation is not expensive. This implies that, in modern practice, the most interesting quasi-Newton methods are those in which the Jacobian approximations are defined in such a way that much linear algebra is saved per iteration. It must be warned that there are many *minimization* problems in which automatic differentiation techniques cannot be applied to compute gradients [19, 86] but this is not frequent in *nonlinear systems* coming from practical applications.

Usually, in large and sparse problems, the resolution of (3) using direct methods [32, 37, 104] is expensive but not prohibitive. (When it is prohibitive it is probably better to use inexact-Newton methods [7, 22, 55].) In these cases, to use $B_0 = J(x_0)$ generating $B_k, k \geq 1$ using cheap-linear-algebra quasi-Newton techniques is worthwhile.

The name "quasi-Newton" was used after 1965 to describe methods of the form (7) such that the equation

$$B_{k+1} s_k = y_k \equiv F(x_{k+1}) - F(x_k) \tag{9}$$

was satisfied for all $k = 0, 1, 2, \ldots$ See [9]. Equation (9) was called "the fundamental equation of quasi-Newton methods". Following the Dennis-Schnabel book [27], most authors call quasi-Newton to all the methods of the form (7), whereas the class of methods that satisfy (9) are called "secant methods". Accordingly, (9) is called "secant equation".

The iteration (7) admits an interesting and pedagogical interpretation.

Assume that, for all $k = 0, 1, 2, \ldots$ we approximate $F(x)$ by a "linear model"

$$F(x) \approx L_k(x) \equiv F(x_k) + B_k(x - x_k). \tag{10}$$

Then, x_{k+1} is the unique solution of the simpler problem $L_k(x) = 0$. By (10) we also have that

$$L_k(x_k) = F(x_k) \quad \text{for all} \quad k = 0, 1, 2, \ldots \tag{11}$$

It is easy to see that (9) implies that

$$L_k(x_{k-1}) = F(x_{k-1}) \quad \text{for all} \quad k = 1, 2, \ldots \tag{12}$$

Therefore, the affine function $L_k(x)$ interpolates $F(x)$ at x_k and x_{k-1}. "Multipoint" secant methods can be defined satisfying

$$L_k(x_j) = F(x_j) \quad \text{for all} \quad j \in I_k, \tag{13}$$

where $\{k-1, k\} \subset I_k$ for all $k = 1, 2, \ldots$ See [4, 5, 12, 13, 36, 42, 51, 59, 66, 67, 81, 92, 94, 103].

This survey is organized as follows. In Section 2 we sketch a local convergence theory that applies to most secant methods introduced after 1965. In Section 3 we give the most used examples of least-change secant-update methods. In Section 4 we introduce interesting quasi-Newton methods that cannot be justified by the theory of Section 2. In Section 5 we discuss large-scale implementations. In Section 6 we show how to deal with possible singularity of the matrices B_k. In Section 7 we discuss procedures used for obtaining global convergence. In Section 8 we study the behavior of some quasi-Newton methods for linear systems. In Section 9 we survey a few numerical studies on large-scale problems. Finally, in Section 10, we discuss the prospective of the area and we formulate some open problems.

2 Least-change update theory

Most practical quasi-Newton methods can be analyzed under the framework of a general theory introduced in [72]. See, also, [73, 75]. This framework can be useful to understand practical methods. However, this section can be skipped at a first reading of this paper, without risk of missing the main algorithmic ideas presented in the remaining sections.

In our analysis, we will use a finite dimensional linear space $\mathcal{E}$ with a scalar product $\langle \cdot, \cdot \rangle_{x,z}$ determined by each pair $x, z \in I\!R^n$. Denote $|E|_{x,z}^2 =$

$\langle E, E\rangle_{x,z}$, where $E \in \mathcal{E}$. Let $V(x,z) \subset \mathcal{E}$ denote an affine subspace determined by any fixed pair $x, z \in I\!R^n$.

The general algorithm analyzed in this section is defined by (7), where

$$B_k = \varphi(x_k, E_k), \tag{14}$$

where $\varphi : I\!R^n \times \mathcal{E} \to I\!R^{n\times n}$. The initial approximation $x_0 \in I\!R^n$ and the initial parameter $E_0 \in \mathcal{E}$ are arbitrary. Moreover, the parameters are generated by

$$E_{k+1} = P_k(E_k), \tag{15}$$

where $P_k \equiv P_{x_k,x_{k+1}}$ is the projection operator on $V(x_k, x_{k+1})$, with respect to the norm $|\cdot|_{x_k,x_{k+1}}$. Therefore, E_{k+1} is the parameter in $V(x_k, x_{k+1})$ which is closest to E_k. This justifies the term "least-change" in the definition of these methods.

The most simple example of (14,15) is Broyden's "good" method (BGM) [8], which is defined by

$$\mathcal{E} = I\!R^{n\times n}, \tag{16}$$

$$|\cdot|_{x,z} = \|\cdot\|_F = \text{ the Frobenius norm for all } x, z \in I\!R^n, \tag{17}$$

$$\varphi(x, E) = E \quad \text{for all } x \in I\!R^n,\ E \in \mathcal{E} \tag{18}$$

and

$$V(x,z) = \{B \in I\!R^{n\times n} \mid B(z-x) = F(z) - F(x)\}. \tag{19}$$

Broyden's sparse (or Schubert's) method [10, 93] is defined by (7,16,17,18) and

$$V(x,z) = \{B \in \mathcal{S} \mid B(z-x) = F(z) - F(x)\}, \tag{20}$$

where $\mathcal{S} \subset I\!R^{n\times n}$ is the set of matrices that have the sparsity pattern of $J(x)$. See [6] for a variation of this method.

Broyden's "bad" method (BBM) is defined by (7,16,17),

$$\varphi(x, E) = E^{-1} \quad \text{for all } x \in I\!R^n,\ E \in \mathcal{E}, E \text{ nonsingular} \tag{21}$$

and

$$V(x,z) = \{H \in I\!R^{n\times n} \mid H[F(z) - F(x)] = z - x\}. \tag{22}$$

Many other examples are given in [72, 73]. In most cases $|\cdot|_{x,z}$ does not depend on x and z. However, situations where $|\cdot|_{x,z}$ changes appear when one analyzes quasi-Newton methods with symmetric Jacobian. This is the case of function minimization. The analysis of the popular DFP and BFGS methods for unconstrained optimization require explicit dependence of the norm with respect to x, z. See [27]. In Section 3.3 we will define least-change methods where φ explicitly depends of x.

Under standard assumptions, which we will consider below, methods defined by (7,14,15) are locally (and "quickly") convergent. The first two are assumptions on the functional F and the remaining ones are assumptions on the method. A convergence analysis for Broyden's method in a situation where the first assumption is violated can be found in [21]. In the rest of this section, $\|\cdot\|$ denotes an arbitrary norm in $\mathbb{R}^n$ as well as its subordinate norm in $\mathbb{R}^{n\times n}$. Moreover, $\langle\cdot,\cdot\rangle$ will denote a scalar product in $\mathcal{E}$ and $|\cdot|$ will be the associated norm. (So, $|E|^2 = \langle E, E\rangle$ for all $E \in \mathcal{E}$.)

Assumption 1. *There exists $x_* \in \mathbb{R}^n$ such that $F(x_*) = 0$ and $J(x_*)$ is nonsingular.*

Assumption 2. *There exists $L > 0$ such that*

$$\|J(x) - J(x_*)\| \leq L\|x - x_*\| \tag{23}$$

if x belongs to some neighborhood of x_*.

The following assumption says that there exists an ideal parameter E_* which is associated to the solution x_* in the sense that $\varphi(x_*, E_*)^{-1}J(x_*)$ is close to the identity matrix. From now on, we write $B_* = \varphi(x_*, E_*)$. In many algorithms, $B_* = J(x_*)$.

Assumption 3. *There exist $E_* \in \mathcal{E}$ and $r_* \in [0, 1)$ such that φ is well defined and continuous in a neighborhood of (x_*, E_*). Moreover, $\varphi(x_*, E_*)$ is nonsingular and*

$$\|I - \varphi(x_*, E_*)^{-1}J(x_*)\| \leq r_*. \tag{24}$$

Assumption 3 implies that we could define an ideal iteration, given by

$$x_{k+1} = x_k - B_*^{-1}F(x_k), \tag{25}$$

satisfying

$$\lim_{k\to\infty} x_k = x_* \quad \text{and} \quad \limsup_{k\to\infty} \frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} \leq r_* \tag{26}$$

if x_0 is close enough to x_*. Of course, the ideal method defined by (25) cannot be implemented in practice because we do not know the solution x_*. However, the least-change update theory consists in showing that some implementable methods enjoy the property (26). Observe that, in the case $r_* = 0$, (26) means superlinear convergence.

Let us define, for all $x, z \in \mathbb{R}^n$,

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

Assumption 4. *In addition to Assumption 3, for all x, z close enough to x_* there exists $E \in V(x, z)$, $c_1 > 0$ such that*

$$|E - E_*| \leq c_1 \sigma(x, z). \tag{27}$$

In the description of the algorithm, we saw that E_{k+1} is a projection of E_k on $V(x_k, x_{k+1})$. Assumption 4 says that the distance between E_* and this affine subspace is of the same order as the maximum distance between $\{x_k, x_{k+1}\}$ and x_*. In other words, we are projecting on manifolds that are not far from the ideal parameter E_*. An algorithm where projections are performed on the intersection of manifolds with boxes can be found in [15]. The relation between the different norms used in the projections is given by Assumption 5.

Assumption 5. *There exists $c_2 > 0$ such that, for all x, z close enough to x_*, $E \in \mathcal{E}$,*

$$|E|_{x,z} \leq [1 + c_2 \sigma(x, z)]|E| \quad \text{and} \quad |E| \leq [1 + c_2 \sigma(x, z)]|E|_{x,z}. \tag{28}$$

Assumption 5 says that the different norms tend to be the same when x and z are close to x_*. Assumptions 4 and 5 do not guarantee that the approximation of E_k to E_* improves through consecutive iterations. (This improvement certainly occurs if $E_* \in V(x_k, x_{k+1})$.) In fact, E_{k+1} might be a worse approximation to E_* than E_k. However, using these assumptions one can prove that the deterioration of E_{k+1} as an approximation to E_* is bounded in such a way that the "error" $|E_{k+1} - E_*|$ is less than the error $|E_k - E_*|$ plus a term which is proportional to the error $\|x_k - x_*\|$. This is a typical "bounded deterioration principle", as introduced in [11]. See, also, [25, 26, 29, 91] and many other papers. By bounded deterioration, the parameters E_k cannot escape from a neighborhood of E_* for which it can be guaranteed that local convergence holds. Therefore, Assumptions 1 to 5 are sufficient to prove the following theorem.

Theorem 1. *Suppose that Assumptions 1 to 5 hold and let $r \in (r_*, 1)$. If $\{x_k\}$ is generated by (7,14,15), there exist $\varepsilon > 0, \delta > 0$ such that, if $\|x_0 - x_*\| \le \varepsilon$ and $|E_0 - E_*| \le \delta$, the sequence is well-defined, converges to x_* and satisfies*

$$\|x_{k+1} - x_*\| \le r\|x_k - x_*\| \tag{29}$$

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

Moreover,

$$\lim_{k\to\infty} |E_{k+1} - E_k| = \lim_{k\to\infty} \|B_{k+1} - B_k\| = 0. \tag{30}$$

At a first sight, the result (29) is disappointing because the same result can be obtained (with $r_* = 0$) if one uses (7) with $B_k = J(x_0)$ for all $k = 0, 1, 2, \ldots$ It could be argued that there is no reason for modifying B_k at every iteration if one can obtain the same result not modifying this Jacobian approximation at all. Obviously, (30) also holds for this stationary-Newton choice of B_k.

Fortunately, some additional results help us to prove that, under some conditions, the ideal speed of convergence (26) can be reached. From a well-known theorem of Dennis and Walker [29] the following result can be obtained.

Theorem 2. *In addition to the hypotheses of Theorem 1, suppose that*

$$\lim_{k\to\infty} \frac{\|[B_k - B_*](x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0. \tag{31}$$

Then, (26) holds.

Theorem 2 corresponds, in the case $r_* = 0$, to the well-known Dennis-Moré condition [24], which characterizes the superlinear convergence of sequences generated by (7). Now, by (30), Theorem 2 implies the following more practical result.

Theorem 3. *Assume the hypotheses of Theorem 1, and*

$$\lim_{k\to\infty} \frac{\|[B_{k+1} - B_*](x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0. \tag{32}$$

Then, (26) holds.

Theorem 2 says that (26) holds if $B_k v_k \approx B_* v_k$, where v_k is the normalized increment. Since the increment is computed *after* B_k, it is not evident that many methods satisfy this condition. On the other hand, Theorem 3 says that (26) holds if $B_{k+1} v_k \approx B_* v_k$. Observe that the increment is computed *before* B_{k+1}. Since, in general, we know how to approximate $B_*(x_{k+1} - x_k)$ (for example, if $B_* = J(x_*)$, we have that $B_*(x_{k+1} - x_k) \approx F(x_{k+1}) - F(x_k)$) the task of computing B_{k+1} satisfying (32) is not so difficult. The most popular situation corresponds to the case $B_* = J(x_*)$ and consists in defining $V(x, z)$ in such a way that B_{k+1} satisfies the secant equation (9). In this case, (32) is equivalent to

$$\lim_{k\to\infty} \frac{\|F(x_{k+1}) - F(x_k) - J(x_*)(x_{k+1} - x_k)\|}{\|x_{k+1} - x_k\|} = 0$$

and this identity holds, if $x_k \to x_*$, due to the assumption (23).

The most important consequence of Theorem 3 is that superlinear convergence of the sequence $\{x_k\}$ takes place when $B_* = J(x_*)$.

None of the theorems above imply that, even when $r_* = 0$, E_k converges to E_*. Simple counter-examples can be shown where this is not true. Moreover, nothing guarantees that E_k is convergent at all. Even in the case of BGM, the best studied least-change secant-update method, it is not known if, under the conditions that are sufficient to prove local-superlinear convergence, the sequence of matrices B_k is convergent.

3 Some least-change secant-update methods

3.1 Broyden's methods

Broyden's "good" method is defined by (7) and (14–19). A simple quadratic programming exercise shows that, for this method,

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k}. \tag{33}$$

Moreover, the relation between the inverses of B_k and B_{k+1} is, in this case,

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1} y_k) s_k^T B_k^{-1}}{s_k^T B_k^{-1} y_k}. \tag{34}$$

This formula shows that the iteration (7) can be computed without solving a linear system at each iteration. For computing B_{k+1}^{-1} we only need to perform $O(n^2)$ operations, whereas $O(n^3)$ operations are necessary for solving a (dense) linear system. It is generally believed that the most stable way in which BGM can be implemented (when the number of variables is small) requires to store the QR factorization of B_k. Since B_{k+1} differs from B_k by a rank-one matrix, the factorization of B_{k+1} can be obtained using $O(n)$ plane rotations. See [80].

Broyden's "bad" method is given by (7), (16), (17), (21) and (22). As in the case of BGM, after some linear algebra the calculations can be organized so that the definition of the method becomes:

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1} y_k) y_k^T}{y_k^T y_k} \tag{35}$$

for all $k = 0, 1, 2, \ldots$ Moreover, according to (35) we have:

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) y_k^T B_k}{y_k^T B_k s_k}. \tag{36}$$

From (34) it is easy to deduce that, if the proper choices are made on the initial point and the initial Jacobian approximation, Broyden's "good" method is invariant under linear transformations in the range space. From (35) we see that Broyden's-"bad" has the same property in the domain space. Therefore, if rounding errors are not considered and the behavior of Broyden's-"good" for $F(x) = 0$ is satisfactory, it must also be satisfactory for solving $AF(x) = 0$. On the other hand, if Broyden's "bad" method works well on $F(x) = 0$, it will also work on $F(Ax) = 0$.

The reasons why BGM is good and BBM is bad are not well understood. Moreover, it is not clear that, in practice, BGM is really better than BBM. In [76] it was observed that, for BGM, since $B_k s_{k-1} = y_{k-1}$, we have, if $k \geq 1$,

$$B_{k+1} s_{k-1} - y_{k-1} = \frac{(y_k - B_k s_k) s_k^T s_{k-1}}{s_k^T s_k}.$$

Analogously, for BBM,

$$B_{k+1}s_{k-1} - y_{k-1} = \frac{(y_k - B_k s_k) y_k^T y_{k-1}}{y_k^T B_k s_k}.$$

Therefore, the "secant error" $B_{k+1}s_{k-1} - y_{k-1}$ is, in both cases, a multiple of $y_k - B_k s_k$. It is natural to conjecture that the BGM iteration will be better than the BBM iteration when

$$\frac{|s_k^T s_{k-1}|}{s_k^T s_k} < \frac{|y_k^T y_{k-1}|}{|y_k^T B_k s_k|}. \tag{37}$$

An analogous reasoning involving $B_{k+1}^{-1} y_{k-1} - s_{k-1}$ leads to conjecture that BGM is better than BBM when

$$\frac{|s_k^T s_{k-1}|}{|s_k^T (B_k)^{-1} y_k|} < \frac{|y_k^T y_{k-1}|}{y_k^T y_k}. \tag{38}$$

In [76] a combined method was implemented that chooses BGM or BBM according to the test (38). This method was tested using a set of small problems and turned out to be superior to both BGM and BBM. By (37-38) BGM tends to be better than BBM if B_k underestimates the true Jacobian. This means that, if B_0 is arbitrarily chosen and the true Jacobian is "larger than B_0", Broyden's "good" method tends to be better than Broyden's "bad". This is also confirmed by small numerical experiments.

3.2 Direct updates of factorizations

Suppose that, for all $x \in I\!R^n$, $J(x)$ can be factorized in the form

$$J(x) = M(x)^{-1} N(x), \tag{39}$$

where $N(x) \in \mathcal{S}_1$, $M(x) \in \mathcal{S}_2$ for all $x \in I\!R^n$, and $\mathcal{S}_1$, $\mathcal{S}_2$ are affine subspaces of $I\!R^{n \times n}$. A least-change secant update method associated to the factorization (39) can be defined by

$$x_{k+1} = x_k - N_k^{-1} M_k F(x_k). \tag{40}$$

In this method, (N_{k+1}, M_{k+1}) is the row-by-row orthogonal projection of (N_k, M_k) on the affine subspace of $I\!R^{n \times n} \times I\!R^{n \times n}$ defined by

$$V = \{(N, M) \in \mathcal{S}_1 \times \mathcal{S}_2 \mid N s_k = M y_k\}. \tag{41}$$

If, in a neighborhood of a solution x_*, $M(x)$ and $N(x)$ are continuous, the theory of Section 2 can be applied to this family of methods to prove that they are locally and superlinearly convergent. See [71]. If (39) represents the LU factorization, we obtain the method introduced in [52]. If we take into account possible sparsity of L^{-1} and U we obtain a method introduced in [16]. Orthogonal factorizations and structured situations were considered in [71]. In this paper it was also shown that the Dennis-Marwil method [23] is a limit method in the family (40-41). By this we mean that, although Dennis-Marwil is not a least-change superlinear convergent method, each Dennis-Marwil iteration can be arbitrarily approximated by iterations of the least-change family. Finally, it is easy to show that Broyden's "good" and "bad" methods are also particular cases of (40-41). Nontrivial methods based on (40-41) can be useful when the system

$$N_k s_k = -M_k F(x_k) \tag{42}$$

is easy to solve.

3.3 Structured methods

Suppose that $J(x) = C(x) + D(x)$ for all $x \in I\!R^n$, where $C(x)$ is easy to compute whereas $D(x)$ is not. In this case, it is natural to introduce the quasi-Newton iteration:

$$x_{k+1} = x_k - [C(x_k) + D_k]^{-1} F(x_k), \tag{43}$$

where, for each $k = 0, 1, 2, \ldots$, D_{k+1} is a projection of D_k on the affine subspace

$$V_{full} = \{D \in I\!R^{n\times n} \mid D s_k = y_k - C(x_{k+1}) s_k\}. \tag{44}$$

Writing $\bar{y}_k = y_k - C(x_{k+1}) s_k$ and considering the Frobenius projection, we see that

$$D_{k+1} = D_k + \frac{(\bar{y}_k - D_k s_k) s_k^T}{s_k^T s_k}. \tag{45}$$

If $C(x_k)^{-1}$ is easy to compute (perhaps because $C(x)$ has a nice sparsity structure) and k is small, some linear algebra can be saved in the computation of $[C(x_k) + D_k]^{-1} F(x_k)$ using the techniques that will be explained in Section 5.

Sometimes one also knows that $D(x)$ belongs to some fixed affine subspace $\mathcal{S}$ for all $x \in I\!R^n$. In this case, we can define D_{k+1} as the projection of D_k on

$$V_{structured} = \{D \in \mathcal{S} \mid Ds_k = \bar{y}_k \equiv y_k - C(x_{k+1})s_k\}, \tag{46}$$

but the formula (45) is not valid anymore, even for Frobenius projections. Moreover, the affine subspace given by (46) can be empty so that the method only makes sense if this definition is conveniently modified. Let us redefine:

$$V_{minimizers} = \{ \text{ Minimizers of } \|Ds_k - \bar{y}_k\|_2 \text{ subject to } D \in \mathcal{S}\}. \tag{47}$$

The affine subspace given by (47) is obviously nonempty and, so, it is possible to project on it. Algorithms for computing this projection were given in [26]. Defining $s = z - x$, $y = F(z) - F(x)$, $\bar{y} = y - C(z)s$ and

$$V(x, z) = \{ \text{ Minimizers of } \|Ds - \bar{y}\|_2 \text{ subject to } D \in \mathcal{S}\} \tag{48}$$

we can apply the theory of Section 2 so that the resulting method turns out to be locally and superlinearly convergent. In principle, Assumption 4 is necessary for proving superlinear convergence. See, also, [29]. However, it can be conjectured that this assumption can be deduced, in this case, from the definitions (46) and (47).

Examples of applied structured quasi-Newton methods can be found, among others, in [3, 46, 47, 56, 61, 62].

4 Other secant methods

The Column-Updating method (COLUM) was introduced in [69] with the aim of reducing the computational cost of BGM. The idea is that, at each iteration, only the j-th column of B_k is changed, where j is defined by $\|s_k\|_\infty = |[s_k]_j|$. So, COLUM is defined by (7) and

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)e_{j_k}^T}{e_{j_k}^T s_k} \tag{49}$$

where $\{e_1, \ldots, e_n\}$ is the canonical basis of $I\!R^n$ and $|e_{j_k}^T s_k| = \|s_k\|_\infty$. The QR and the LU factorizations of B_{k+1} can be obtained from the corresponding factorizations of B_k using classical linear-programming updating techniques. By (49), we have that

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1}y_k)e_{j_k}^T B_k^{-1}}{e_{j_k}^T B_k^{-1} y_k}. \tag{50}$$

Partial convergence results for COLUM were given in [39, 69, 74]. It has been proved that COLUM enjoys local and superlinear convergence if the method is restarted (taking $B_k = J(x_k)$) every m iterations, where m is an arbitrary positive integer. Moreover, when the method (with or without restarts) converges, the convergence is $r-$superlinear and quadratic every $2n$ iterations. Finally, COLUM (without restarts) is superlinearly convergent if $n = 2$.

The Inverse Column-Updating method (ICUM), introduced in [78], is given by (7) and

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1} y_k) e_{j_k}^T}{e_{j_k}^T y_k}, \tag{51}$$

where $|e_{j_k}^T y_k| = \|y_k\|_\infty$. Therefore, B_{k+1}^{-1} is identical to B_k^{-1} except on the j_k-th column. So,

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) e_{j_k}^T B_k}{e_{j_k}^T B_k s_k}. \tag{52}$$

Similar local convergence results to those of COLUM were given in [60, 78].

It is easy to see that COLUM and ICUM have the invariancy properties of BGM and BBM respectively. Probably, combined methods in the sense of [76] can also be efficient. See the rationale preceding formula (38) in Subsection 3.1 of this survey.

The discussion that leads to (38) suggests the introduction of quasi-Newton methods of the form

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k) v_k^T}{v_k^T s_k}, \tag{53}$$

where $v_k \perp s_{k-1}$, or

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1} y_k) w_k^T}{w_k^T y_k}, \tag{54}$$

where $w_k \perp y_{k-1}$. These methods are close to the multipoint secant methods studied in [4, 5, 12, 13, 36, 42, 51, 59, 66, 67, 81, 92, 94, 103] in the sense that they satisfy an additional interpolatory condition. Their convergence analysis using the techniques of the above cited papers must be easy, but their practical efficiency does not seem to have been studied. Some authors [49, 98] choose the parameter w_k in (54) with the aim of maintaining well-conditioning properties of the matrix B_k.

The first quasi-Newton method with direct updates of factorizations was introduced by Dennis and Marwil in [23]. We did not talk about this method in Subsection 3.2 because this is not a least-change method in the sense of Section 2. The Dennis-Marwil algorithm modifies the upper-triangular factor of the LU factorization of B_k at each iteration, so that the secant equation is always satisfied (with some stability safeguards). See, also, [83, 99]. Convergence results for the Dennis-Marwil method are even weaker than the ones that can be proved for the Stationary Newton method commented in Section 1. The work [23] inspired the introduction of other methods with direct updates of factorizations with stronger convergence results. We have already mentioned the least-change secant update methods introduced in [71], which enjoy local and superlinear convergence. Other methods, having the same theoretical convergence properties as the Stationary Newton method, were introduced in [41, 68, 70]. The Row-Scaling method (see [41]) is particularly simple and, sometimes, quite effective. It consists in the updating $B_{k+1} = D_k B_k$, where D_k is diagonal and it is chosen so that the secant equation is satisfied, when this is possible. The good numerical properties of the Row-Scaling method are quite surprising. Unfortunately, this updating technique cannot be used in function minimization because it does not preserve possible symmetry of the Jacobian approximations.

We finish this section mentioning the quasi-Newton method introduced by Thomas [100], which is given by

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1} y_k) d_k^T B_k^{-1}}{d_k^T B_k^{-1} y_k}, \tag{55}$$

where

$$d_k = [R_k + (\|s_k\|_2/2) I] s_k$$

and

$$R_{k+1} = (1 + \|s_k\|_2)(\|s_k\|_2 I + R_k - \frac{d_k d_k^T}{d_k^T s_k}).$$

The properties of this method are not yet well understood. However, in spite of its larger cost per iteration, very good numerical results have been reported in several works. See, for example [48].

5 Large-scale implementations

The best known general-purpose modern implementations of quasi-Newton methods for solving large nonlinear systems are based on rank-one correction formulae like BGM, BBM, COLUM and ICUM. See [39, 41, 64]. Unfortunately, the methods based on direct updates of factorizations which have pleasant convergence properties [71] need sparsity of the L^{-1} factor in the LU decomposition of the true Jacobian, a property that holds only in very structured problems.

A crucial decision involves the choice of the initial Jacobian approximation B_0. The most favorable situation occurs when one is able to compute a good approximation of $J(x_0)$ (perhaps using automatic differentiation or the techniques given in [18, 20]) and the LU factorization of this approximation is sparse. In this case, after possible permutations of rows and columns, we compute

$$B_0 = LU \tag{56}$$

and we use this sparse factorization throughout the calculations.

If a sparse factorization of a suitable approximation of $J(x_0)$ is not available, it is sensible to use

$$B_0^{-1} = \left(\frac{\nabla f_1(x_0)}{\|\nabla f_1(x_0)\|_2^2}, \ldots, \frac{\nabla f_n(x_0)}{\|\nabla f_n(x_0)\|_2^2} \right). \tag{57}$$

In (57), B_0 approximates $J(x_0)$ in the sense that the $J(x_0)B_0^{-1}$ has only 1's on the diagonal. With this choice the initial iteration is scale-invariant and the vector s_0 is a descent direction for $\|F(x)\|_2^2$ (see Section 7). Of course, some alternative choice must be employed for a column of B_0^{-1} if the involved gradient is null.

In BGM and COLUM, we have

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)v_k^T}{v_k^T s_k}, \tag{58}$$

and, consequently,

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1} y_k)v_k^T}{v_k^T B_k^{-1} y_k} B_k^{-1}. \tag{59}$$

Therefore,

$$B_{k+1}^{-1} = (I + u_k v_k^T) B_k^{-1}, \tag{60}$$

where $u_k = (s_k - B_k^{-1} y_k)/v_k^T B_k^{-1} y_k$. Thus,

$$B_k^{-1} = (I + u_{k-1} v_{k-1}^T) \ldots (I + u_0 v_0^T)\, B_0^{-1}. \tag{61}$$

Formula (61) shows that methods of the form (58) can be implemented associated to (56) or (57) adding $O(n)$ operations and storage positions per iteration. By (61), for computing $B_k^{-1}F(x_k)$ one needs to store $u_j, v_j, j = 0, 1, \ldots, k-1$ and the factorization (or the inverse) of B_0. Moreover, this computation involves the solution of a linear system whose matrix is B_0 plus a sequence of k operations consisting in a scalar product, a scalar-vector product and the sum of two vectors. The whole procedure can be quite economic if k is small but becomes prohibitive if k is large. In the case of COLUM, it is obvious that only one additional vector is needed per iteration. For BGM, a clever trick given in [30] allows one to implement (7, 61) storing only one additional vector per iteration. See [39, 41, 43, 79].

In BBM and ICUM we have

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1}y_k)w_k^T}{w_k^T y_k}. \tag{62}$$

Therefore, defining $u_k = (s_k - B_k^{-1}y_k)/w_k^T y_k$, we obtain:

$$B_{k+1}^{-1} = B_0^{-1} + u_0 w_0^T + \ldots + u_k w_k^T. \tag{63}$$

This formula suggests straightforward associations of (62) with (56) or (57). A recent numerical study by Lukšan and Vlček [64] indicates that ICUM could be the most effective secant method for large-scale problems with the initial choice (56).

6 Dealing with singularity

The quasi-Newton iteration (7) is well defined only if B_k is nonsingular. Local convergence theories usually assume that $J(x_*)$ is nonsingular and that B_0 is close to $J(x_*)$. Under these conditions it can usually be proved that B_k is nonsingular for all k. However, in practice, the initial choice of B_0 could be singular and, moreover, B_{k+1} could be singular even when B_k is not.

Singularity of B_0 might occur when one chooses $B_0 = J(x_0)$ (or some very good approximation of the Jacobian). Since the (nonsingular) Newton step minimizes $\|J(x_0)s + F(x_0)\|$, it is natural, in the singular case, to choose s_0 as any minimizer of $\|J(x_0)s + F(x_0)\|_2^2$. Choosing the minimum-norm minimizer, we obtain:

$$s_0^\dagger = -J(x_0)^\dagger F(x_0), \tag{64}$$

where $J(x_0)^\dagger$ is the Moore-Penrose pseudoinverse of the initial Jacobian (see [38]). Using a well-known approximation of the pseudoinverse, we can also compute, for some $\mu > 0$,

$$s_0(\mu) = -(J(x_0)^T J(x_0) + \mu I)^{-1} J(x_0)^T F(x_0). \tag{65}$$

When $\mu \to 0$, $s_0(\mu)$ tends to $s_0^\dagger$. The step $s_0(\mu)$ can be interpreted as the minimizer of $\|J(x_0)s + F(x_0)\|_2$ on a ball whose radius is smaller than $\|s_0^\dagger\|_2$.

In practical computations, singularity of $J(x_0)$ is detected during the LU factorization of this matrix: at some stage of the LU algorithm it is impossible to choose a safe nonnull pivot. When the problem is large, and $J(x_0)$ is possibly sparse, computing (65) is expensive and, so, this device is seldom used. It is usually preferred to continue the LU factorization replacing the null or very small pivot by some suitable nonnull quantity that takes into account the scaling of the matrix. See [41]. There is no strong justification for this procedure except that, perhaps, it is not necessary to choose carefully B_0 when x_0 is far from the solution. (Even this statement can be argued.)

On the other hand, a singular B_{k+1} can appear even if B_k is nonsingular.

When B_{k+1} is obtained from B_k by means of a secant rank-one correction,

$$B_{k+1} = B_k + \frac{(y_k - B_k s_k)v_k^T}{v_k^T s_k}, \tag{66}$$

as in the case of BGM and COLUM, we have

$$\det(B_{k+1}) = \frac{v_k^T B_k^{-1} y_k}{v_k^T s_k} \det(B_k). \tag{67}$$

If $|\frac{v_k^T B_k^{-1} y_k}{v_k^T s_k}|$ is very small or very large then, either the scaling of B_{k+1} is very different from that of B_k or their stability characteristics are very different. Conservative small-variation arguments recommend us to impose

$$\sigma |\det(B_k)| \le |\det(B_{k+1})| \le \frac{1}{\sigma} |\det(B_k)|, \tag{68}$$

where $\sigma \in (0,1)$ is small (say, $\sigma \approx 0.1$). By (67), if $|\frac{v_k^T B_k^{-1} y_k}{v_k^T s_k}| \notin [\sigma, 1/\sigma]$, the inequalities (68) do not hold and B_{k+1} must be modified. Following [80], we can replace (66) by

$$B_{k+1} = B_k + \eta_k \frac{(y_k - B_k s_k)v_k^T}{v_k^T s_k}, \tag{69}$$

where $\eta_k \in [0,1]$. Clearly, (68) is satisfied if $\eta_k = 0$, but $\eta_k = 1$ is the best choice in the sense that B_{k+1} satisfies the secant equation. Therefore, it is natural to choose η_k as the maximum $\eta \in [0,1]$ such that (68) is satisfied. This motivates the definition

$$\eta_k = \max\left\{\eta \in [0,1] \mid \sigma \leq |(1-\eta) + \eta \frac{v_k^T B_k^{-1} y_k}{v_k^T s_k}| \leq \frac{1}{\sigma}\right\}. \tag{70}$$

In "inverse" rank-one correction methods like BBM and ICUM, it is easier to write directly

$$B_{k+1}^{-1} = B_k^{-1} + \frac{(s_k - B_k^{-1} y_k) w_k^T}{w_k^T y_k}. \tag{71}$$

An analogous reasoning to the one used to choose (70) leads us to the modification

$$B_{k+1}^{-1} = B_k^{-1} + \eta_k \frac{(s_k - B_k^{-1} y_k) w_k^T}{w_k^T y_k} \tag{72}$$

and, consequently, to the choice

$$\eta_k = \max\left\{\eta \in [0,1] \mid \sigma \leq |(1-\eta) + \eta \frac{w_k^T B_k s_k}{w_k^T y_k}| \leq \frac{1}{\sigma}\right\}. \tag{73}$$

Both in the initial iteration as in the updated ones a close-to-singular matrix B_k usually generates a very large increment s_k. Very simple step-control procedures are always associated to the implementation of quasi-Newton methods. In practical problems, it has been verified that opportunistic ways of controlling the step-length may prevent many divergence situations.

7 Global convergence tools

The results presented in this paper are local, in the sense that convergence to a solution can be guaranteed if the solution is assumed to exist and both the initial point and the initial Jacobian approximation are close enough to the solution and its Jacobian respectively.

It is of maximal practical importance to analyze what happens with sequences generated by quasi-Newton methods when no restrictions are made on the initial approximations. Unfortunately, almost nothing positive can be said about sequences generated by pure formulae like (7), unless strong assumptions are made on F. Newtonian sequences can oscillate between neighborhoods of two or more non-solutions or tend to infinity, even in problems where a unique solution exists. So, if we want to devise algorithms with global convergence properties, the basic iteration (7) must be modified.

Usually, modifications of the basic iteration make use of some merit function. Almost always, some norm of $F(x)$ is used. The squared 2-norm $\|F(x)\|_2^2$ is frequently preferred because of its differentiability properties. We will call $f(x)$ the (continuous and nonnegative) merit function, whose main property is that $f(x) = 0$ if, and only if, $F(x) = 0$. Therefore, the problem of solving $F(x) = 0$ turns out to be equivalent to the problem of finding a global minimizer of $f(x)$. If, at a global minimizer, $f(x)$ does not vanish, the original system has no solution at all. For simplicity, assume that

$$f(x) = \frac{1}{2}\|F(x)\|_2^2, \tag{74}$$

so

$$\nabla f(x) = J(x)^T F(x). \tag{75}$$

From (75), we see that the Newton direction fits well with the necessity of decreasing $f(x)$. Computing the directional derivative, we obtain:

$$\langle -J(x)^{-1}F(x), \nabla f(x)\rangle = -2f(x) < 0,$$

So, it is always possible to decrease $f(x)$ along the Newton direction, if this direction is well defined and $f(x) \neq 0$. Many algorithms can be interpreted as adaptations of unconstrained optimization techniques (see [27]) to the minimization of $f(x)$. In particular, the iteration

$$x_{k+1} = x_k - \alpha_k J(x_k)^{-1} F(x_k) \tag{76}$$

has been exhaustively analyzed. See [33, 65] and references therein. If $\alpha_k > 0$ is chosen is such a way that $f(x_{k+1})$ is sufficiently smaller than $f(x_k)$, then every limit point of the sequence generated by (76) either is a solution or a point where the Jacobian is singular. So, if the Jacobian is nonsingular for all $x \in I\!R^n$ and $f(x)$ has bounded level sets, (76) necessarily finds a solution. Finally, in a vicinity of such a solution it can be proved that $\alpha_k \equiv 1$ satisfies the sufficient decrease requirements, therefore the method (76) coincides, ultimately, with (2) and the convergence is quadratic.

The merit function (74) brings difficulties in connection with nonsymmetric quasi-Newton methods because the direction $-B_k^{-1}F(x_k)$ is not, in general, a descent direction for f. This is one of the reasons why it is important to use good initial Jacobian approximations in this context, whereas diagonal initial Hessian approximations are usually efficient in function minimization. Griewank [44] has proved that Broyden's "good" method, with a suitable line search, also has global convergence properties assuming uniform nonsingularity of the Jacobians. Li and Fukushina [57] introduced a line search for BGM that ensures global and superlinear convergence, if the merit function has bounded level sets and the Jacobians are nonsingular.

Other attempts for globalization of quasi-Newton methods (without nonsingularity assumptions) rely on the exploration of the good descent properties of Newton. Among these we can cite:

1. Hybrid strategies [80, 85], in which Broyden's iteration are combined with special iterations which are, essentially, discretizations of Newton iterations.

2. Nonmonotone strategies [34]: here "ordinary" quasi-Newton iterations are accepted, even if the merit function is increased during some iterations, but the algorithm switches to a Newton iteration if a given tolerance is violated.

3. A strategy due to Bonnans and Burdakov [14]: if the sufficient decrease condition is violated the step-length is reduced, but, at the same time, the Jacobian approximation is updated using a secant formula. As a result, the search direction changes during the current iteration and tends to the Newton direction. An antecedent of this idea can be found in [84].

A common drawback of all the globalization strategies based on decreasing a norm is that local-nonglobal minimizers of $f(x)$ are strong attractors of the iterative process. Other norm-minimization related techniques can be found in [53, 96, 97]. Therefore, globalized algorithms can converge to points in which the Jacobian is singular. Unfortunately, such points are completely useless from the point of view of finding solutions of the nonlinear system. It is easy to see that all the observations related to the Newton direction made in this section, except the ones related to rapid local convergence, are valid for the choice (57) of the Jacobian approximation.

A completely different source of globalization procedures is the homotopic approach, by means of which a sequence of slightly modified problems are solved, in such a way that the first one is trivial and the last one is (1). For example, the "regularizing homotopy", used in [101, 102] is

$$H(x,t) = tF(x) + (1-t)(x - x_0). \tag{77}$$

The solution of $H(x,0) = 0$ is, obviously, x_0 and the solution of $H(x,1) = 0$ is the one required in (1). Many methods for tracing the homotopy path are described in the literature. Locally convergent quasi-Newton methods are useful tools in this case since strategies like (77) deal with several nonlinear systems for which good initial estimates are available. See, also, [1, 2, 90].

8 Results for linear systems

In this section we assume that $F(x) = Ax - b$, $A \in \mathbb{R}^{n\times n}$, $b \in \mathbb{R}^n$. To study the behavior of quasi-Newton methods for linear systems is important under different points of view. On one hand, real-life problems can be linear or nearly linear. On the other hand, the properties of a method in the linear case usually determine the local convergence behavior of the method in the nonlinear case. In a neighborhood of a solution where the Jacobian is nonsingular, the linear approximation of F is dominant and, so, the generated sequence tends to behave as in the linear case. For example, if $F(x) = Ax - b$ and A is nonsingular, Newton's method is well defined and converges in just one iteration. This is the main reason why the local convergence of this method is quadratic.

Until 1979 it was believed that Broyden's methods did not enjoy finite convergence when applied to linear systems. However, in [35] it was proved that Broyden's method and many other methods of the form (66) or (71) also converge in a finite number of steps.

Let us consider the method defined by (7) and (66). Gay's theorem [35] says that, if A and B_0 are nonsingular and $x_0 \in \mathbb{R}^n$ is arbitrary, then $F(x_k) = 0$ for some $k \leq 2n$. The convergence of x_k to $x_* \equiv A^{-1}b$ is far from being monotone in any sense.

The local convergence consequences of Gay's theorem for general nonlinear systems are that, under the usual assumptions that guarantee local convergence, methods like BGM, BBM, COLUM and ICUM enjoy $2n$-step quadratic convergence. Therefore, $\|x_{k+2n} - x_*\| / \|x_k - x_*\|^2$ is asymptotically bounded above. This property implies r-superlinear convergence. See [81].

The finite convergence theorem [35] sheds light on theoretical properties of rank-one secant methods but is of little importance for practical large-scale linear problems. The intermediate iterations (x_k with $k < 2n$) are, usually, very poor approximations of the solution so that the full cycle of $2n$ steps is necessary for obtaining a reasonable approximation of x_*. When n is large, a sequence of $2n$ iterations is not affordable for the methods considered in Section 5, since the cost of the k-th iteration is proportional to kn, both in terms of time and storage. Therefore, practical implementations of rank-one secant methods for linear systems need modifications of the basic iteration (7). See the discussion in [30].

Some authors [30, 77, 80] studied variations of BGM for linear systems. Here we survey the results presented in [77], correcting, by the way, some arithmetic typos of that paper. Given $x_0 \in I\!R^n$ and $B_0 \in I\!R^{n\times n}$ nonsingular, the linear Broyden method is defined by

$$x_{k+1} = x_k - \lambda_k B_k^{-1} F(x_k), \tag{78}$$

where $\lambda_k \neq 0$ and

$$B_{k+1} = B_k + \eta_k \frac{(y_k - B_k s_k) s_k^T}{s_k^T s_k}. \tag{79}$$

The coefficient $\eta_k \in [0.9, 1.1]$ is such that

$$|\det(B_{k+1})| \geq 0.1\ |\det(B_k)|. \tag{80}$$

Moré and Trangenstein [80] proved that (80) holds with $\eta_k \in [0.9, 1.1]$ defining $\gamma_k = s_k^T B_k^{-1} y_k / s_k^T s_k$, with $\eta_k = 1$ if $|\gamma_k| \geq 0.1$, and $\eta_k = (1 - 0.1 sign(\gamma_k))/(1 - \gamma_k)$ if $|\gamma_k| < 0.1$, where $sign(0) = 1$. This choice of η_k provides the number closest to unity such that (80) is satisfied. See [80] and Section 6 of this paper.

For the method defined by (78), (79) and (80) it can be proved that

$$\|B_k - A\|_F \leq \|B_0 - A\|_F$$

for all $k = 0, 1, 2, \ldots$ and

$$\|B_{k+1} - A\|_F^2 \leq \|B_k - A\|_F^2 - 0.891\ \|B_{k+1} - B_k\|_F^2$$

for all $k = 0, 1, 2, \ldots$. It follows that the series $\sum \|B_{k+1} - B_k\|_F^2$ is convergent. So, $\|B_{k+1} - B_k\|$ tends to 0.

It can also be proved that the sequence generated by (78), (79) and (80) satisfies

$$\frac{\|x_{k+1} - x_*\|}{\|x_k - x_*\|} \leq \frac{\varepsilon_k + |\lambda_k - 1|}{1 - \varepsilon_k} \tag{81}$$

for all $k = 0, 1, 2, \ldots$, where $\{\varepsilon_k\}$ is a sequence that tends to zero.

Formula (81) explains the behavior of the error $x_k - x_*$ independently of the convergence of the sequence. In particular, it shows that the sequence is superlinearly convergent if $\lambda_k \to 1$, and that convergence at a linear rate takes place if, eventually, $\lambda_k \in [\sigma, 2 - \sigma]$ for some $\sigma > 0$.

Finally, in [77] it has been proved that $\lambda_k \to 1$ holds when one chooses λ_k as the (nonnull) minimizer of $\|A(x_k + \lambda d_k) - b\|_2^2$ along the direction $d_k \equiv -B_k^{-1}F(x_k)$. If this λ_k is null, we replace it by 1. However, this possible replacement is not necessary for k large enough.

As a result, we have a global and superlinearly convergent BGM-like method for solving linear nonsingular systems. The proposed choice of λ_k has an advantage over the choice $\lambda_k = 1$ in the large-scale case. When λ_k is the one-dimensional minimizer proposed above, the residual norm at the iterate x_{k+1} is smaller than the norm of $Ax_k - b$. Therefore, in terms of the residual norm, the quality of the approximation is improved at every iteration, and an acceptable final approximation can be (perhaps) obtained for $k << 2n$. An alternative choice with similar theoretical properties that, in some sense, minimizes a norm of the error, has been considered in [30] and [77].

The effectivity of Broyden-like methods for solving large-scale linear systems is associated to the availability of good preconditioners. If the initial matrix B_0 is defined as the available preconditioner, a small number of iterations can be expected, at least when one uses clever choices of the steplength. In [30] it has been claimed that these alternatives are competitive with standard Krylov-subspace methods for solving linear systems. However, much research is necessary on this subject both form the theoretical and the practical point of view.

9 Numerical studies

In this section we comment some numerical studies involving the application of quasi-Newton methods for solving large-scale nonlinear systems of equations.

The study [41] involves 7 variably dimensioned nonlinear systems. Six of them are "toy problems" and have been designed with the aim of testing numerical algorithms. The seventh is the discretization of a Poisson equation. The algorithms are Newton's method, the Stationary Newton method, Broyden's "good" method, Broyden's sparse (Schubert) method, the Dennis-Marwil method and three direct-update methods that includes the row-scaling method mentioned in Section 4. Matrix factorizations use the algorithm of George and Ng [37] and a nonmonotone globalization procedure is incorporated.

The study [40] uses 3 discretizations of two-dimensional boundary-value problems with known solutions: Poisson, Bratu and Convection-Diffusion. The three of them depend on a parameter λ according to which the problem is more or less difficult. If $\lambda = 0$ the problems are linear. If $\lambda << 0$, noncoercivity is severe and the discretized problems are very hard. The tested algorithms are Newton, Stationary Newton, BGM and COLUM. All the algorithms have the option of using backtracking to improve global convergence.

The study [34] solves a set of problems given in [63] having similar characteristics to the set of problems of [41]. In addition, a discretization of the driven cavity flow problem is also considered, which has a parameter λ, the Reynolds number, that controls nonlinearity. Finally, the study includes a Convection-Diffusion problem and a set of artificial problems where Newton's method (without step control) do not converge.

The study [64] includes 30 problems. 16 of them are of the type considered in [41] with some superposition with that set. In addition, the study has countercurrent reactor problems, second-order boundary value problems (including Poisson and Convection-Diffusion), problems of flow in a channel, swirling flow problems, porous medium problems, a nonlinear biharmonic problem and the driven cavity problem. The objective of this study is to introduce a globalization procedure. The underlying quasi-Newton methods are the discrete Newton method, the Stationary Newton method, the sparse Broyden (Schubert) method, the variation due to Bogle and Perkins [6], Li's method [58], a combination of Li with Schubert, the row-scaling method [41], Broyden's "good" method, COLUM and ICUM.

None of the above cited studies contradicts the common belief that Newton's method is the most robust algorithm for solving nonlinear systems. Concerning globalization procedures, experiments recommend to be cautious, because in many problems the attempts to reduce the sum of squares lead to convergence to local-nonglobal minimizers. As a matter of fact, the simple stabilization procedure that consists in not letting the step-length to be too large (see Section 6) is, frequently, very effective to turn a divergent algorithmic sequence into a convergent one.

When convergence is maintained, quasi-Newton corrections usually improve substantially the performance of Newton's method. The amount of this improvement depends of the Jacobian structure. In the problems considered in the above cited studies, methods that do not save linear algebra, like Broyden-sparse, must be discarded, since its computational cost per iteration is roughly the same as Newton's. Practically all quasi-Newton corrections are more effective than the Stationary Newton method. According to [64], ICUM ranks first, but there seems to be little difference between this method and BGM or COLUM. Up to our knowledge there are no published numerical studies for large-scale problems where Broyden's "bad" method is included.

10 Conclusions and perspectives

In recent years, quasi-Newton methods for solving square smooth nonlinear systems have been out of the mainstream of numerical analysis research. A popular scientific journal on Numerical Analysis published 4 papers on the subject before 1970, 10 between 1971 and 1980, 11 in the eighties and none from 1991 to 1999. Sometimes, research in a family of numerical techniques becomes out-of-fashion after its incorporation to ordinary practice of problem solvers in Physics, Chemistry, Engineering and Industry. Other times, promising algorithms are completely forgotten, both in research and applications.

The situation of the area surveyed in this paper is perhaps intermediate. The classical paper [25] is cited in most works concerning quasi-Newton methods for nonlinear systems. While this survey was being written it had been cited 361 times in indexed scientific journals. The last 100 citations go from 1992 to the present days. 42 of these citations come from non-mathematical journals. It must be warned that, frequently, the Dennis-Moré paper [25] is cited in connection to quasi-Newton methods for minimization problems, and not for nonlinear systems. Since the everyday practice in Physics, Chemistry and Engineering includes the resolution of nonlinear systems using Newton's method, we are tempted to conclude that the penetration of the quasi-Newton technology in applications, although existing, has not been as intense at the potentiality of the technique deserves.

In the Introduction of most quasi-Newton papers, it was stressed that the main motivation was to avoid computation of cumbersome derivatives. However, even before the boom of automatic differentiation, practitioners found that, for many of their problems, computing derivatives was not as difficult or costly as stated in the quasi-Newton literature. They also verified that beginning a quasi-Newton process with $B_0 = I$, or some other arbitrary matrix, very often causes disastrous results and, so, the computation of an initial Jacobian is almost always necessary. Moreover, the programming effort of computing the initial Jacobian is the same as the one necessary for computing all the Jacobians, so the tendency of many practitioners has been to use Newton's method or its stationary variation with refinements.

In practical problems in which the Jacobian can be computed but its structure is too bad for factorization, the modern tendency is to use the inexact-Newton approach [22], in which an iterative linear solver is used for solving the Newtonian linear equation $J(x_k)s = -F(x_k)$ up to some precision which is sufficient to guarantee convergence of the nonlinear solver. Moreover, the inexact-Newton technology fits well with global convergence requirements. Probably, many users felt disappointed when they tried to globalize quasi-Newton methods by the mere introduction of a damping parameter and backtracking procedures.

However, a reasonable scope of problems exists, for which quasi-Newton methods that save linear algebra are quite effective and, probably, outperform inexact-Newton algorithms. This is the case of large-scale problems in which the Jacobian can be computed, its factorization is affordable but it is very costly in comparison to the single updating procedures of rank-one methods. The recipe for those cases is to begin with a Newtonian iteration, and to continue with some cheap rank-one method as far as this is effective. Unfortunately, a code like that must be prepared to return to Newtonian iterations, a disappointing fact for those who hoped that quasi-Newton techniques could always replace Newton.

Quasi-Newton methods for solving large-scale nonlinear systems will be largely used in applications when both numerical analysts and potential users be conscious about their real advantages and limitations. Our point of view is that rank-one algorithms provide, in many problems, efficient and economic ways to refine a basic (first) Newtonian iteration. If we are right, questions often neglected in the quasi-Newton literature, as "when should one restart?" must be answered, in spite of its poor theoretical appeal.

We finish this survey stating 10 open problems, some of which were incidentally mentioned in the text.

1. It is well known that, under the usual nonsingularity and Lipschitz assumptions, the matrices B_k generated by Broyden's "good" method do not necessarily converge to $J(x_*)$. Does this sequence of matrices always have a limit? What happens with the sequences $\{B_k\}$ corresponding to other methods?

2. Convergence theorems for least-change update and other quasi-Newton methods say that there exist $\varepsilon, \delta > 0$ such that $x_k \to x_*$ superlinearly whenever $\|x_0 - x_*\| \leq \varepsilon$ and $\|B_0 - J(x_*)\| \leq \delta$. Is this superlinear convergence uniform? In other words, for which methods can we prove that "there exist $\varepsilon, \delta > 0$ and a sequence of positive numbers $\varepsilon_k \to 0$ such that whenever $\|x_0 - x_*\| \leq \varepsilon$ and $\|B_0 - J(x_*)\| \leq \delta$, the sequence x_k converges to x_* and $\|x_{k+1} - x_*\| \leq \varepsilon_k \|x_k - x_*\|$ for all k"?

3. Is it possible to prove local convergence without restarts of methods like COLUM and ICUM? What about superlinear convergence?

4. Are there reasonable sufficient conditions under which the convergence of Broyden-like methods for linear systems takes place in less than $2n$ iterations?

5. It is generally accepted that the Dennis-Marwil method (and some other similar direct factorization algorithms) enjoys local convergence only if periodic Jacobian restarts are performed. However, no counter-example showing that local convergence without the restarting condition might not hold is known. Does a counter-example exist in the linear case?

6. Does there exist a cheap and theoretically justified procedure for modifying the LU factorization of B_0 when a null or very small pivot is found?

7. Is it possible to prove that Assumption 4 necessarily holds for the choice (48) of $V(x, z)$?

8. Which are the properties of direct-secant-update and structured quasi-Newton methods when applied to linear systems?

9. The order of convergence of Newton's method with p refinements (the Jacobian is repeated during p consecutive iterations) is $2 + p$. See

[81, 82, 95]. This means that $\|x_{k+p+1} - x_*\| / \|x_k - x_*\|^{2+p}$ is asymptotically bounded. Can something better be expected when, instead of repeating the previous Jacobian, we update it with a secant formula?

10. Many methods in the flourishing interior point field for mathematical programming can be interpreted as clever damped Newton iterations on an homotopic basis. Can they be improved by suitable quasi-Newton updates? (Up to our knowledge, no attempt has been made in this sense, except the one in [28].)

Acknowledgements

The author is indebted to two anonymous referees for their very careful reading of the first draft of this versions and for many suggestions that led to its improvement.

References

[1] Aluffi-Pentini, F.; Parisi, V.; Zirilli, F. [1984]: A differential equations algorithm for nonlinear equations, *ACM Transactions on Mathematical Software* 10, pp. 299-316.

[2] Aluffi-Pentini, F.; Parisi, V.; Zirilli, F. [1984]: DAFNE: differential-equations algorithm for nonlinear equations, *ACM Transactions on Mathematical Software* 10, pp. 317-324.

[3] Ávila, J. H. ; Concus, P. [1979]: Update methods for highly structured systems of nonlinear equations, *SIAM Journal on Numerical Analysis* 16, pp. 260-269.

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

[5] Bittner, L. [1959]: Eine Verallgemeinerung des Sekantenverfahrens zur näherungsweisen Berechnung der Nullstellen eines nichtlinearen Gleichngssystems, *Will. Z. Techn. Univ. Dresden* 9, pp. 325-329.

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

[7] Brown, P. N.; Saad, Y. [1994]: Convergence theory of nonlinear Newton-Krylov Algorithms, *SIAM Journal on Optimization* 4, pp. 297-330.

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

[9] Broyden, C. G. [1967]: Quasi-Newton methods and their applications to function minimization, *Mathematics of Computation* 21, pp. 368-381.

[10] Broyden, C. G. [1971]: The convergence of an algorithm for solving sparse nonlinear systems, *Mathematics of Computation* 19, pp. 577-593.

[11] 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.

[12] Burdakov, O. [1983]: Stable versions of the secant method for solving systems of equations, *U.S.S.R. Comput. Math. and Math. Phys.* 23, pp. 1–10.

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

[14] Burdakov, O. [1998]: Private communication.

[15] Calamai, P. H. ; Moré, J. J. [1987]: Quasi-Newton updates with bounds, *SIAM Journal on Numerical Analysis* 24, pp. 1434-1441.

[16] 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.

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

[18] 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.

[19] Conn, A. R.; Scheinberg, K.; Toint, Ph. L. [1997]: Recent progress in unconstrained nonlinear optimization without derivatives, *Mathematical Programming* 79, pp. 397-414.

[20] 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.

[21] Decker, D. W. ; Kelley, C. T. [1985]: Broyden's method for a class of problems having singular Jacobian at the root, *SIAM Journal on Numerical Analysis* 22, pp. 563-574.

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

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

[24] 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.

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

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

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

[28] Dennis Jr., J. E. ; Morshedi, M.; Turner, K. [1987]: A variable metric variant of the Karmarkar algorithm for linear programming, *Mathematical Programming* 39, pp. 1-20.

[29] 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.

[30] 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.

[31] Dixon, L. C. W. [1987]: Automatic Differentiation and Parallel Processing in Optimisation, TR N$^{\underline{\text{o}}}$ 180, The Hatfield Polytechnique, Hatfield, UK.

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

[33] Eisenstat, S. C. ; Walker, H. F. [1994]: Globally convergent inexact Newton methods, *SIAM Journal on Optimization* 4, pp. 393-422.

[34] Friedlander, A. , Gomes-Ruggiero, M. A. ; Kozakevich, D. N. ; Martínez, J. M. ; Santos, S. A. [1997]: Solving nonlinear systems of equations by means of quasi-Newton methods with a nonmonotone strategy, *Optimization Methods and Software* 8, pp. 25-51.

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

[36] 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.

[37] 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.

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

[39] 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.

[40] Gomes–Ruggiero, M. A. ; Kozakevich, D. N.; Martínez, J. M. [1996]: A numerical study on large-scale nonlinear solvers, *Computers and Mathematics with Applications* 32, pp. 1-13.

[41] 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.

[42] 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.

[43] 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.

[44] Griewank, A. [1986]: The 'global' convergence of Broyden-like methods with a suitable line search, *Journal of the Australian Mathematical Society Ser. B* 28, pp. 75-92.

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

[46] Hart, W. E.; Soesianto, F. [1992]: On the solution of highly structured nonlinear equations, *Journal of Computational and Applied Mathematics* 40, pp. 285-296.

[47] 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.

[48] Huang, Z. ; Spedicato, E. [1994]: Numerical testing of quasi-Newton and some other related methods for nonlinear systems, Quaderni del Dipartimento di Matematica, Statistica, Informatica e Applicazioni 4, Universitá degli Studi di Bergamo, Bergamo, Italy.

[49] Ip, C. M. ; Todd, M. J. [1988]: Optimal conditioning and convergence in rank one quasi-Newton updates, *SIAM Journal on Numerical Analysis* 25, pp. 206-221.

[50] 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.

[51] Jankowska, J. [1979]: Theory of multivariate secant methods, *SIAM Journal on Numerical Analysis* 16, pp. 547-562.

[52] 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.

[53] Kaporin, I. E.; Axelsson, O. [1995]: On a class of nonlinear equation solvers based on the residual norm over a sequence of affine subspaces, *SIAM Journal on Scientific Computing* 16, pp. 228-249.

[54] Kedem, G. [1980]: Automatic Differentiation of Computer Programs, *ACM Transactions on Mathematical Software* 6, 150-165.

[55] Kelley, C. T. [1995]: *Iterative methods for linear and nonlinear equations*, SIAM, Philadelphia, Pennsylvania.

[56] 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.

[57] Li, D. H.; Fukushima, M. [1999]: A derivative-free line search and global convergence of Broyden-like method for nonlinear equations, Technical Report 99004, Department of Applied Mathematics and Physics, Kyoto University. To appear in *Optimization Methods and Software.*

[58] Li, G. [1989]: Successive column correction algorithms for solving sparse nonlinear systems of equations, *Mathematical Programming* 43, pp. 187-207.

[59] 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.

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

[61] Lucia, A. [1985]: Partial molar excess properties, null spaces and a new update for the hybrid method of chemical process design, *AICHE Journal* 31, pp. 558-566.

[62] Lucia, A.; Miller, D. C.; Kumar, A. [1985]: Thermodinamically consistent quasi-Newton formulas, *AICHE Journal* 31, pp. 1381-1388.

[63] Lukšan, L. [1994]: Inexact trust region method for large sparse systems of nonlinear equations, *Journal of Optimization Theory and Applications* 81, pp. 569-590.

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

[65] Lužanin, Z.; Krejić, N.; Herceg, D. [1997]: Parameter selection for Inexact Newton method, *Nonlinear Analysis. Theory, Methods and Applications* 30, pp. 17-24.

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

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

[68] 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.

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

[70] Martínez, J. M. [1987]: Quasi–Newton methods with factorization scaling for solving sparse nonlinear systems of equations, *Computing* 38, pp. 133–141.

[71] Martínez, J. M. [1990]: 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.

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

[73] 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.

[74] Martínez, J. M. [1992]: On the convergence of the Column-Updating method, *Computational and Applied Mathematics* 12, pp. 83-94.

[75] Martínez, J. M. [1992]: Fixed-point quasi-Newton methods, *SIAM Journal on Numerical Analysis* 29, pp. 1413–1434.

[76] Martínez, J. M. ; Ochi, L. S. [1982]: Sobre dois métodos de Broyden, *Matemática Aplicada e Computacional* 1, pp. 135-141.

[77] Martínez, J. M. ; Qi, L. [1995]: Inexact Newton methods for solving nonsmooth equations, *Journal of Computational and Applied Mathematics* 60, pp. 127-145.

[78] Martínez, J. M. ; Zambaldi, M. C. [1992]: An inverse Column-Updating method for solving Large-Scale Nonlinear Systems of Equations, *Optimization Methods and Software* 1, pp. 129-140.

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

[80] Moré, J. J. ; Trangenstein, J. A. [1976]: On the global convergence of Broyden's method, *Mathematics of Computation* 30, pp. 523-540.

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

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

[83] Paloschi, J. R.; Perkins, J. D. [1986]: The updating of LU-factors in quasi-Newton methods, *Computers and Chemical Engineering* 10, pp. 241-247.

[84] Polak, E. [1974]: A globally converging secant method with applications to boundary value problems, *SIAM Journal on Numerical Analysis* 11, pp. 529-537.

[85] Powell, M. J. D. [1970]: A hybrid method for nonlinear equations, in *Numerical Methods for Nonlinear Algebraic Equations*, P. Rabinowitz, ed. , Gordon and Breach, London, pp. 87-114.

[86] Powell, M. J. D. [1998]: Direct search algorithms for optimization calculations, *Acta Numerica 1998*, pp. 287-336.

[87] Rall, L. B. [1981]: *Automatic Differentiation - Techniques and Applications*, Springer Lecture Notes in Computer Science, Vol 120.

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

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

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

[91] Rheinboldt, W. C. ; Vandergraft, J. S. [1974]: On the local convergence of update methods, *SIAM Journal on Numerical Analysis* 11, pp. 1069-1085.

[92] Schnabel, R. B. [1983]: Quasi-Newton methods using multiple secant equations, TR CU-CS-247-83, Department of Computer Science, University of Colorado at Boulder.

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

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

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

[96] Shi, Y. X. [1996]: Solving nonlinear systems using a global-local procedure, *ZAMM* 76, pp. 539-540.

[97] Shi, Y. X. [1996]: A globalization procedure for solving nonlinear systems of equations, *Numerical Algorithms* 12, pp. 273-286.

[98] Spedicato, E. ; Greenstadt, J. [1978]: On some classes of variationally derived Quasi-Newton algorithms for systems of nonlinear algebraic equations, *Numerische Mathematik* 29, pp. 363-380.

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

[100] Thomas, S. W. [1975]: Sequential estimation techniques for Quasi-Newton algorithms, Technical Report TR 75-227, Cornell University.

[101] Watson, L. T. [1979]: A globally convergent algorithm for computing points of C^2 maps, *Applied Mathematics and Computation* 5, pp. 297-311.

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

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

[104] 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.