Convergence properties of the inverse Column-Updating method

Vera L. R. Lopes * José Mario Martínez †

February 24, 1995

Abstract

The inverse Column-Updating method is a secant algorithm for solving nonlinear systems of equations introduced recently by Martínez and Zambaldi (*Optimization Methods and Software* 1(1992), pp. 1 - 140). This method is one of the less expensive reliable quasi - Newton methods for solving nonlinear simultaneous equations, in terms of linear algebra work. Since it does not belong to the well-known LCSU (least-change secant-update) class, special arguments are used for proving local convergence. In this paper we prove that, if convergence is assumed, then R-superlinear convergence takes place. Moreover, we prove local convergence for a version of the method with (not necessarily Jacobian) restarts. Finally, we prove that local and R-superlinear convergence holds without restarts in the two-dimensional case. From a practical point of view, we show that, in some cases, the numerical performance of the inverse Column-Updating method is very good.

Key words. Nonlinear systems, quasi-Newton methods, inverse Column-Updating method.

*Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil (vlopes@ime.unicamp.br). This author was supported by FAPESP (Grant 90-3724-6), FINEP and FAEP-UNICAMP.

†Department of Applied Mathematics, IMECC-UNICAMP, University of Campinas, CP 6065, 13081-970 Campinas SP, Brazil (martinez@ccvax.unicamp.br). This author was supported by FAPESP (Grant 90-3724-6), CNPq and FAEP-UNICAMP.

1 Introduction

In this paper, we address the problem of solving

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

where $F : I\!R^n \to I\!R^n$ is differentiable. We denote $J(x) \equiv F'(x)$.

Practical methods for solving (1.1) are iterative. In particular, quasi-Newton methods proceed as follows: given x_k, the $k-th$ approximation to the solution, an affine model $L_k(x)$ of $F(x)$ is considered, where

$$L_k(x) = F(x_k) + B_k(x - x_k).$$

Then, x_{k+1} is defined as a solution of $L_k(x) = 0$. See Dennis and Schnabel [1983]. When B_k is nonsingular, the quasi-Newton iteration takes the form

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

Secant methods are characterized by the interpolatory condition

$$L_k(x_{k-1}) = F(x_{k-1}), \quad k = 1, 2, \ldots.$$

This condition is equivalent to the *secant equation*

$$B_{k+1}(x_{k+1} - x_k) = F(x_{k+1}) - F(x_k), \quad k = 0, 1, 2, \ldots$$

For $n > 1$, there exist infinite many matrices B that satisfy the secant equation. See Dennis and Moré [1977]. In general, secant methods do not need computation of derivatives and the resolution of the system

$$B_k s_k = -F(x_k)$$

is not expensive, due to suitable updating procedures.

The best known class of secant methods is the least change secant update (LCSU) family (Dennis and Schnabel [1979, 1983]). For this class of methods, under suitable assumptions, Q-superlinear convergence can be proved, provided that x_0 and B_0 are good approximations of the solution x_* and the Jacobian $J(x_*)$ respectively. See Dennis and Walker [1981], Martínez [1990, 1992]. Some authors have introduced quasi-Newton methods that do not belong to the LCSU family, but seem to be useful in practice. Among them there is the Column-Updating method (CUM), in which B_{k+1} is obtained from B_k by changing only one column of B_k, in such a way that the secant

equation is satisfied. See Martínez [1984], Gomes-Ruggiero and Martínez [1992].

The inverse Column-Updating method (ICUM) is analogous to the Column - Updating method, but in ICUM we update one column of H_k, the approximation of $J(x_k)^{-1}$, per iteration. See Martínez and Zambaldi [1992]. This is the method analyzed in this work.

This paper is organized as follows. In Section 2, we prove that, under usual assumptions, linear convergence implies R-superlinear convergence. (See Ortega and Rheinboldt [1970].) In Section 3 we show that, for a restarted version of the algorithm (not necessarily Newton restarts), local, linear, and thus R-superlinear convergence takes place. Moreover, we prove in Section 4 that stronger results are true in the two-dimensional case. Similar results were proved by Martínez [1993] for the Column-Updating method. An essential tool in our proofs is the finite convergence of rank-one secant methods for linear systems, given by Gay [1979]. In Section 5 we show some numerical experiments. Conclusions are given in Section 6.

2 The general algorithm without restarts

We consider the problem of solving (1.1), where $F : \Omega \subset \mathbb{R}^n \to \mathbb{R}^n$, Ω is an open convex set, $F \in C^1(\Omega)$, $x_* \in \Omega$ and $F(x_*) = 0$. From now on $\|\cdot\|$ will be the Euclidean norm on $\mathbb{R}^n$. (However, most results involving $\|\cdot\|$ are true for an arbitrary norm.) We assume that there exists $L > 0$ such that

$$\|J(x) - J(x_*)\| \leq L\|x - x_*\| \tag{2.1}$$

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

$$\|F(z) - F(x) - J(x_*)(z - x)\| \leq L\|z - x\|\sigma(x, z) \tag{2.2}$$

for all $x, z \in \Omega$, where

$$\sigma(x, z) = \max\{\|x - x_*\|,\ \|z - x_*\|\}. \tag{2.3}$$

Moreover, assuming that $J(x_*)$ is nonsingular and defining

$$\|J(x_*)^{-1}\| = \beta, \tag{2.4}$$

we have that (2.2) implies that

$$\|z - x - J(x_*)^{-1}[F(z) - F(x)]\| \leq \beta L\|z - x\|\sigma(x, z). \tag{2.5}$$

The algorithm below is a generalization of the inverse Column-Updating method.

Algorithm 2.1. Assume that $x_0 \in \Omega$ and let H_0 be an arbitrary nonsingular initial matrix, $\alpha \in (0,1)$. For all $k = 0, 1, 2, \ldots$ such that $F(x_k) \neq 0$, perform the following steps:
Step 1. Compute

$$x_{k+1} = x_k - H_k F(x_k) \tag{2.6}$$

$$s_k = x_{k+1} - x_k$$

$$y_k = F(x_{k+1}) - F(x_k). \tag{2.7}$$

If $y_k = 0$, stop. Otherwise,
Step 2. Choose $u_k \in I\!R^n$ such that $\|u_k\| = 1$,

$$|u_k^T y_k| \geq \alpha \|y_k\| \tag{2.8}$$

and

$$u_k^T H_k^{-1} s_k \neq 0. \tag{2.9}$$

Step 3. Define

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k) u_k^T}{u_k^T y_k} \equiv H_k - \frac{H_k F(x_{k+1}) u_k^T}{u_k^T y_k}. \tag{2.10}$$

We remark that in the inverse Column-Updating method u_k is a canonical basis vector of $I\!R^n$.

Applying the Sherman-Morrison formula (see Golub and Van Loan [1989, p. 151]) we verify that if (2.8) and (2.9) are verified then H_{k+1} is well defined and nonsingular.

Since $H_k^{-1} s_k \neq 0$, $\alpha \in (0,1)$ and $y_k \neq 0$ then there exists u_k such that $u_k^T H_k^{-1} s_k \neq 0$ and $y_k^T u_k \neq 0$, so, we may assume that it is always possible to choose u_k satisfying (2.8) and (2.9). The choice $u_k = y_k / \|y_k\|$ defines Broyden's "bad" method. (See Broyden [1965], Dennis and Schnabel [1983].) Obviously, with this choice (2.8) holds even for $\alpha = 1$. The classical local convergence theory of quasi-Newton methods gives sufficient conditions under which (2.9) holds for all k. Below we assume that the sequence defined by Algorithm 2.1 is well-defined, which means that (2.9) is satisfied for all $k = 0, 1, 2, \ldots$.

Assumption A1. Let us assume that an infinite sequence generated by Algorithm 2.1 is well defined and that the following statements are true:
(a)

$$\lim_{k\to\infty} x_k = x_*. \tag{2.11}$$

(b) There exists $r \in (0,1)$ such that

$$\|x_{k+1} - x_*\| \le r\|x_k - x_*\| \tag{2.12}$$

for all $k = 0, 1, 2, \ldots$.
(c) There exists $M > 0$ such that

$$\|H_k\| \le M \ , \quad \|H_k^{-1}\| \le M \tag{2.13}$$

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

It is easy to see that, by (2.1) and the nonsingularity of $J(x_*)$, there exist $\lambda_1 > 0,\ \lambda_2 > 0$ such that

$$\lambda_1\|x_{k+1} - x_k\| \le \|F(x_{k+1}) - F(x_k)\| \le \lambda_2\|x_{k+1} - x_k\| \tag{2.14}$$

if x_k and x_{k+1} are close enough to x_*.

We will prove now that the convergence is R-superlinear.

Theorem 2.1. *Consider Algorithm 2.1 and suppose that Assumption A1 is satisfied. Then,*

$$\lim_{k\to\infty} \frac{\|x_{k+2n} - x_*\|}{\|x_k - x_*\|} = 0, \tag{2.15}$$

which implies that the convergence of $\{x_k\}$ to x_* is R-superlinear.

Proof. Suppose that (2.15) is not true. Then there exist K_1, an infinite subset of $I\!N$, and $c > 0$ such that

$$\|x_{k+2n} - x_*\| \ge c\|x_k - x_*\| \tag{2.16}$$

for all $k \in K_1$.

Assume that $K_1 = \{k_1, k_2, \ldots, \}, \quad k_1 < k_2 < \ldots$. We assume, without loss of generality, that $k_{i+1} \ge k_i + 2n$ for all $i = 1, 2, 3, \ldots$. Define

$$\varepsilon_k = \|x_k - x_*\|, \quad k = 0, 1, 2, \ldots \tag{2.17}$$

By (2.7), (2.8), (2.12), (2.14) and (2.17), we have that

$$|u_k^T y_k| \ge \alpha\|y_k\| = \alpha\|F(x_{k+1}) - F(x_k)\| \ge \alpha\lambda_1\|x_{k+1} - x_k\|$$

$$\geq \alpha\lambda_1[\|x_k - x_*\| - \|x_{k+1} - x_*\|] \geq \alpha\lambda_1[\|x_k - x_*\| - r\|x_k - x_*\|] \geq c_1\varepsilon_k \quad (2.18)$$

for k large enough, where $c_1 = \alpha\lambda_1(1-r)$. Moreover, by (2.12), (2.14), (2.16) and (2.17),

$$\varepsilon_{k+\ell} \geq c\varepsilon_k \quad (2.19)$$

for all $k \in K_1,\ \ell \in \{0, 1, \ldots, 2n\}$.

Let us define, for all $k \in K_1$, $\ell = 0, 1, 2, \ldots, 2n$,

a) $x_{k,0} = x_k$, $H_{k,0} = H_k$, (2.20)

b) $x_{k,\ell+1} = x_k - H_{k,\ell}A(x_{k,\ell} - x_*)$, (2.21)

c) $s_{k,\ell} = x_{k,\ell+1} - x_{k,\ell}$, (2.22)

d) $y_{k,\ell} = A(x_{k,\ell+1} - x_{k,\ell}) = A(x_{k,\ell+1} - x_*) - A(x_{k,\ell} - x_*)$, (2.23)

e) $H_{k,\ell+1} = H_{k,\ell} - \frac{H_{k,\ell}A(x_{k,\ell+1} - x_*)u_{k+\ell}^T}{u_{k+\ell}^T y_{k,\ell}}$, (2.24)

where $A = J(x_*)$. Clearly, $x_{k,\ell+1}$ will be well defined for all $\ell = 0, 1, \ldots, 2n-1$, if $u_{k+\ell}^T y_{k,\ell} \neq 0$. This will be the case if k is large enough. More precisely, we will prove that, for large enough $k \in K_1$, $x_{k,1}, x_{k,2}, \ldots, x_{k,2n}$ are well defined, $H_{k,1}, H_{k,2}, \ldots, H_{k,2n-1}$ are well defined and nonsingular, and

(i) $\|x_{k+\ell} - x_{k,\ell}\| = O(\varepsilon_k^2)$ (2.25)

(ii) $\|H_{k+\ell} - H_{k,\ell}\| = O(\varepsilon_k)$ (2.26)

The proof will be done by induction on ℓ. For $\ell = 0$, (2.25) and (2.26) are trivial. Assume that (2.25) and (2.26) hold for some fixed ℓ. Then, by induction, (2.6) and (2.21),

$$\|x_{k+\ell+1} - x_{k,\ell+1}\| = \|x_{k+\ell} - H_{k+\ell}F(x_{k+\ell}) - x_{k,\ell} + H_{k,\ell}A(x_{k,\ell} - x_*)\|$$

$$\leq \|x_{k+\ell} - x_{k,\ell}\| + \|H_{k,\ell}A(x_{k,\ell} - x_*) - H_{k+\ell}F(x_{k+\ell}) + H_{k+\ell}A(x_{k,\ell} - x_*) - H_{k+\ell}A(x_{k,\ell} - x_*)\|$$

$$\leq O(\varepsilon_k^2) + \|H_{k+\ell}\|\ \|F(x_{k+\ell}) - A(x_{k,\ell} - x_*)\| + \|H_{k+\ell} - H_{k,\ell}\|\ \|A\|\ \|x_{k,\ell} - x_*\|.$$

Now, by induction, (2.2) and (2.12),

$$\|F(x_{k+\ell}) - A(x_{k,\ell} - x_*)\| \leq \|F(x_{k+\ell}) - A(x_{k+\ell} - x_*)\| + \|A(x_{k+\ell} - x_*) - A(x_{k,\ell} - x_*)\|$$

$$\leq L\|x_{k+\ell} - x_*\|^2 + \|A\|\ \|x_{k+\ell} - x_{k,\ell}\| \leq Lr^{2\ell}\varepsilon_k^2 + \|A\|O(\varepsilon_k^2).$$

So, by (2.26), and using that

$$\|x_{k,\ell} - x_*\| \leq \|x_{k,\ell} - x_{k+\ell}\| + \|x_{k+\ell} - x_*\| \leq O(\varepsilon_k^2) + r^\ell\varepsilon_k = O(\varepsilon_k),$$

we obtain:

$$\|x_{k+\ell+1} - x_{k,\ell+1}\| \leq O(\varepsilon_k^2) + \|H_{k+\ell}\|[Lr^{2\ell}\varepsilon_k^2 + \|A\|O(\varepsilon_k^2)] + O(\varepsilon_k)\|A\|\varepsilon_k = O(\varepsilon_k^2). \quad (2.27)$$

Using (2.2), (2.12) and (2.27) we have that

$$\|y_{k+\ell} - y_{k,\ell}\| = \|F(x_{k+\ell+1}) - F(x_{k+\ell}) - A(x_{k,\ell+1} - x_*) + A(x_{k,\ell} - x_*)\|$$

$$\leq \|F(x_{k+\ell+1}) - A(x_{k+\ell+1} - x_*)\| + \|A(x_{k+\ell+1} - x_*) - A(x_{k,\ell+1} - x_*)\|$$

$$+\|F(x_{k+\ell}) - A(x_{k+\ell} - x_*)\| + \|A(x_{k+\ell} - x_*) - A(x_{k,\ell} - x_*)\|$$

$$\leq L\|x_{k+\ell+1} - x_*\|^2 + \|A\| \, \|x_{k+\ell+1} - x_{k,\ell+1}\| + L\|x_{k+\ell} - x_*\|^2 + \|A\| \, \|x_{k+\ell} - x_{k,\ell}\|$$

$$\leq Lr^{2(\ell+1)}\varepsilon_k^2 + \|A\|O(\varepsilon_k^2) + Lr^{2\ell}\varepsilon_k^2 + \|A\|O(\varepsilon_k^2)\| = O(\varepsilon_k^2). \tag{2.28}$$

Now, by (2.18), (2.19) and (2.28),

$$\begin{aligned} |u_{k+\ell}^T y_{k,\ell}| &= |u_{k+\ell}^T y_{k+\ell} + u_{k+\ell}^T (y_{k,\ell} - y_{k+\ell})| \\ &\geq |u_{k+\ell}^T y_{k+\ell}| - |u_{k+\ell}^T (y_{k,\ell} - y_{k+\ell})| \\ &\geq c_1 \varepsilon_{k+\ell} - O(\varepsilon_k^2) \geq c_1 c \varepsilon_k - O(\varepsilon_k^2) \geq c_2 \varepsilon_k > 0 \end{aligned} \tag{2.29}$$

if k is large enough, where $c_2 = c_1 c/2$.

By (2.29), $H_{k,\ell+1}$ is well defined. Moreover, by the inductive hypothesis,

$$\|H_{k,\ell+1} - H_{k+\ell+1}\| = \|H_{k,\ell} - \frac{H_{k,\ell}A(x_{k,\ell+1} - x_*)u_{k+\ell}^T}{u_{k+\ell}^T y_{k,\ell}} - H_{k+\ell} + \frac{H_{k+\ell}F(x_{k+\ell+1})u_{k+\ell}^T}{u_{k+\ell}^T y_{k+\ell}}\|$$

$$\leq O(\varepsilon_k) + \|\frac{H_{k+\ell}F(x_{k+\ell+1})}{u_{k+\ell}^T y_{k+\ell}} - \frac{H_{k,\ell}A(x_{k,\ell+1} - x_*)}{u_{k+\ell}^T y_{k,\ell}}\|$$

$$\leq O(\varepsilon_k) + \|\frac{H_{k+\ell}F(x_{k+\ell+1})}{u_{k+\ell}^T y_{k+\ell}} - \frac{H_{k+\ell}A(x_{k,\ell+1} - x_*)}{u_{k+\ell}^T y_{k+\ell}}\|$$

$$+\|\frac{H_{k,\ell}A(x_{k,\ell+1} - x_*)}{u_{k+\ell}^T y_{k,\ell}} - \frac{H_{k+\ell}A(x_{k,\ell+1} - x_*)}{u_{k+\ell}^T y_{k+\ell}}\|.$$

By (2.2), (2.18), (2.19) and the inductive hypothesis,

$$\|\frac{H_{k+\ell}F(x_{k+\ell+1})}{u_{k+\ell}^T y_{k+\ell}} - \frac{H_{k+\ell}A(x_{k,\ell+1} - x_*)}{u_{k+\ell}^T y_{k+\ell}}\|$$

$$= \frac{1}{|u_{k+\ell}^T y_{k+\ell}|}\|H_{k+\ell}F(x_{k+\ell+1}) - H_{k+\ell}A(x_{k,\ell+1} - x_*)\|$$

$$\leq \frac{1}{c\varepsilon_k}\|H_{k+\ell}\| \, \|F(x_{k+\ell+1} - A(x_{k,\ell+1} - x_*)\|$$

$$\leq \frac{M}{c_1\varepsilon_k}[\|F(x_{k+\ell+1}) - A(x_{k+\ell+1} - x_*)\| + \|A(x_{k+\ell+1} - x_*) - A(x_{k,\ell+1} - x_*)\|]$$

$$\leq \frac{M}{c_1\varepsilon_k} L r^{2(\ell+1)}\varepsilon_k^2 + \frac{M\|A\|}{c_1\varepsilon_k} O(\varepsilon_k^2) = O(\varepsilon_k). \tag{2.30}$$

On the other hand,

$$\|\frac{H_{k,\ell}A(x_{k,\ell+1} - x_*)}{u_{k+\ell}^T y_{k,\ell}} - \frac{H_{k+\ell}A(x_{k,\ell+1} - x_*)}{u_{k+\ell}^T y_{k+\ell}}\|$$

$$= \|\frac{1}{u_{k+\ell}^T y_{k+\ell} u_{k+\ell}^T y_{k,\ell}}[u_{k+\ell}^T y_{k+\ell} H_{k,\ell} A(x_{k,\ell+1} - x_*) - u_{k+\ell}^T y_{k,\ell} H_{k+\ell} A(x_{k,\ell+1} - x_*)]\|$$

$$\leq \frac{1}{|u_{k+\ell}^T y_{k+\ell}|\ |u_{k+\ell}^T y_{k,\ell}|}\|u_{k+\ell}^T y_{k+\ell} H_{k,\ell} - u_{k+\ell}^T y_{k,\ell} H_{k+\ell}\|\ \|A\|\ \|x_{k,\ell+1} - x_*\|. \tag{2.31}$$

But, using (2.12) and (2.27), we see that

$$\|x_{k,\ell+1} - x_*\| \leq \|x_{k,\ell+1} - x_{k+\ell+1}\| + \|x_{k+\ell+1} - x_*\|$$

$$\leq O(\varepsilon_k^2) + r^{\ell+1}\varepsilon_k = O(\varepsilon_k). \tag{2.32}$$

Moreover, by (2.28) and induction,

$$\|u_{k+\ell}^T y_{k+\ell} H_{k,\ell} - u_{k+\ell}^T y_{k,\ell} H_{k+\ell}\| \leq \|u_{k+\ell}^T y_{k+\ell} H_{k,\ell} - u_{k+\ell}^T y_{k+\ell} H_{k+\ell}\|$$

$$+\|u_{k+\ell}^T y_{k+\ell} H_{k+\ell} - u_{k+\ell}^T y_{k,\ell} H_{k+\ell}\|$$

$$\leq |u_{k+\ell}^T y_{k+\ell}|\ \|H_{k,\ell} - H_{k+\ell}\| + |u_{k+\ell}^T y_{k,\ell} - u_{k+\ell}^T y_{k+\ell}|\ \|H_{k+\ell}\|$$

$$\leq |u_{k+\ell}^T y_{k+\ell}| O(\varepsilon_k) + \|u_{k+\ell}^T\|\ \|y_{k,\ell} - y_{k+\ell}\|\ \|H_{k+\ell}\|$$
$$\leq |u_{k+\ell}^T y_{k+\ell}| O(\varepsilon_k) + O(\varepsilon_k^2). \tag{2.33}$$

So, by (2.18), (2.29), (2.32) and (2.33), the right-hand side of (2.31) is dominated by

$$\frac{\|A\| O(\varepsilon_k)}{|u_{k+\ell}^T y_{k,\ell}|}\frac{1}{|u_{k+\ell}^T y_{k+\ell}|}(|u_{k+\ell} y_{k,\ell}| O(\varepsilon_k) + O(\varepsilon_k^2))$$

$$\leq \frac{\|A\| O(\varepsilon_k)}{O(\varepsilon_k)}[O(\varepsilon_k) + O(\varepsilon_k)] = O(\varepsilon_k). \tag{2.34}$$

So, by (2.30) and (2.34),

$$\|H_{k,\ell+1} - H_{k+\ell+1}\| \leq O(\varepsilon_k) + O(\varepsilon_k) + O(\varepsilon_k) = O(\varepsilon_k). \tag{2.35}$$

Since $\|H_k^{-1}\| \leq M$, we obtain, using Banach's perturbation lemma (see Golub and van Loan [1989, pp. 59-60] that H_{k+1}^{-1} exists and, moreover, if k is large enough,

$$\|H_{k,\ell+1}^{-1} - H_{k+\ell+1}^{-1}\| = O(\varepsilon_k).$$

Now, by Theorem 2.2 of Gay [1979], we have that $x_{k,2n} = x_*$ for all $k \in K_1$ such that $x_{k,2n}$ is well defined. So, by (2.25) with $\ell = 2n$, we have that $\varepsilon_{k+2n} = O(\varepsilon_k^2)$ if $k \in K_1$. This contradicts (2.16).

Therefore, the set K_1 cannot exist and, thus, (2.15) is proved. QED.

3 The restarted algorithm

In this section we analyze a variation of Algorithm 2.1 that consists of computing x_{k+1} and H_{k+1} as in Algorithm 2.1, except when $k+1$ is a multiple of a fixed integer m. In this case, we set $H_{k+1} = C(x_{k+1})$ where, for all $x \in \Omega$, $C(x)$ is an approximation of $J(x)^{-1}$.

Algorithm 3.1.

Let $x_0 \in \Omega$, $H_0 = C(x_0)$ a nonsingular matrix, $\alpha \in (0,1)$, m a fixed integer. If $F(x_k) \neq 0$, compute x_{k+1} as in Step 1 of Algorithm 2.1. If $k+1$ is not a multiple of m, compute s_k, y_k, u_k and H_{k+1} as in Algorithm 2.1. If $k+1 \equiv 0 \pmod{m}$ set

$$H_{k+1} = C(x_{k+1}). \tag{3.1}$$

As in the case of Algorithm 2.1, ICUM corresponds to the following choice of u_k:

$$u_k = e_{j_k}$$

where e_{j_k} is a canonical basis vector. In general, we choose

$$j_k = \arg\max\{|e_1^T y_k|, \ldots, |e_n^T y_k|\}. \tag{3.2}$$

Martínez and Zambaldi [1992] gave sufficient conditions under which we can guarantee that the matrices H_k generated using (3.2) are nonsingular. By the Sherman-Morrison formula, this is equivalent to say that (2.9) is satisfied for all k.

Theorem 3.1. *Let $r \in (0,1)$. There exist $\varepsilon, \delta > 0$ such that, if $\|x_0 - x_*\| \leq \varepsilon$ and $\|C(x_k) - J(x_*)^{-1}\| \leq \delta$ for all $k \equiv 0 \pmod m$ then the sequences $\{x_k\}$ and $\{H_k\}$ generated by Algorithm 3.1 are well defined, $\{x_k\}$ converges to x_* and satisfies*

$$\|x_{k+1} - x_*\| \leq r\|x_k - x_*\| \tag{3.3}$$

for all $k = 0, 1, 2, \ldots$. Moreover, $\|H_k\|$ and $\|H_k^{-1}\|$ are bounded.

Proof. See the proof of Theorem 3.1 of Martínez and Zambaldi [1992]. In this theorem we are dealing with a general u_k while in Martínez and Zambaldi [1992] only the particular choice $u_k = e_{j_k}$ is considered. However, the proof does not present additional difficulties in the general case, if we work with suitable matrix norms. QED.

By Theorem 3.1, we know that, under reasonable conditions we can obtain linear convergence and boundedness of $\|H_k\|$ and $\|H_k^{-1}\|$ for Algorithm 3.1. This is not the case for the non-restarted Algorithm 2.1. Let us prove now that the convergence of Algorithm 3.1 is also R-superlinear.

Theorem 3.2. *Consider Algorithm 3.1 with $m \geq 2n$. Assume that a well defined infinite sequence is generated, $\|H_k\|$ and $\|H_k^{-1}\|$ are bounded and (3.3) holds. Then*

$$\lim_{j \to \infty} \frac{\|x_{jm+2n} - x_*\|}{\|x_{jm} - x_*\|} = 0, \tag{3.4}$$

which implies that the convergence of $\{x_k\}$ to x_* is R-superlinear.

Proof. The proof of (3.4) is done by contradiction in the same way we did in Theorem 2.1. QED.

4 The inverse Column-Updating method without restarts when $n = 2$

In this section we consider the ICUM implementation of Algorithm 2.1, where

$$u_k = e_{j_k} \quad k = 0, 1, 2, \ldots, \tag{4.1}$$

with

$$|e_{j_k}^T y_k| = \max \{|e_j^T y_k|, \ j = 1, \ldots, n\} \tag{4.2}$$

and $e_1, \ldots, e_n$ is the canonical basis of $\mathbb{R}^n$. We will prove that in the two-dimensional case, with the choice (4.1) – (4.2), the Assumption A1 holds if x_0 and H_0 are close enough to x_* and $J(x_*)^{-1}$, respectively. This result is essentially based on the following "bounded deterioration" property.

Lemma 4.1. *Assume $n = 2$. Consider Algorithm 2.1 with the choice (4.1) – (4.2). There exist ε_1, $\delta_1 > 0$ such that, if $\|x_k - x_*\| \leq \varepsilon_1$ and $\|H_k^{-1} - J(x_*)^{-1}\| \leq \delta_1$, then*

$$\|H_{k+1} - J(x_*)^{-1}\|_1 \leq \|H_k - J(x_*)^{-1}\|_1 + \frac{n\beta L}{\lambda_1}\sigma(x_k, x_{k+1}). \tag{4.3}$$

Proof. By (2.10),

$$\|H_{k+1} - J(x_*)^{-1}\|_1 = \|H_k + \frac{(s_k - H_k y_k)e_{j_k}^T}{e_{j_k}^T y_k} - J(x_*)^{-1}\|_1$$

$$\leq \|H_k + \frac{[J(x_*)^{-1} - H_k]y_k e_{j_k}^T}{e_{j_k}^T y_k} - J(x_*)^{-1}\|_1 + \|\frac{(s_k - J(x_*)^{-1}y_k)e_{j_k}^T}{e_{j_k}^T y_k}\|_1. \tag{4.4}$$

Now, by (4.2), observing that (2.5) and (2.14) hold for $\|\cdot\|_1$, for small enough $\varepsilon_1, \delta_1 > 0$, we have:

$$\|\frac{(s_k - J(x_*)^{-1}y_k)e_{j_k}^T}{e_{j_k}^T y_k}\|_1 = \|\frac{\{x_{k+1} - x_k - J(x_*)^{-1}[F(x_{k+1}) - F(x_k)]\}e_{j_k}^T}{e_{j_k}^T y_k}\|_1$$

$$\leq \frac{2}{\|y_k\|_1}\beta L\|s_k\|_1\sigma(x_k, x_{k+1}) = \frac{2\beta L}{\lambda_1}\sigma(x_k, x_{k+1}). \tag{4.5}$$

On the other hand,

$$\|\frac{H_k + (J(x_*)^{-1} - H_k)y_k e_{j_k}^T - J(x_*)^{-1}}{e_{j_k}^T y_k}\|_1$$

$$\|H_k + \frac{J(x_*)^{-1}y_k e_{j_k}^T}{e_{j_k}^T y_k} - \frac{H_k y_k e_{j_k}^T}{e_{j_k}^T y_k} - J(x_*)^{-1}\|_1$$

$$\leq \|H_k - J(x_*)^{-1}\|_1\|I - \frac{y_k e_{j_k}^T}{e_{j_k}^T y_k}\|_1. \tag{4.6}$$

Let us define $y_k = \begin{pmatrix} y_k^1 \\ y_k^2 \end{pmatrix}$. By (4.1) and (4.2), we have:

$$\|I - \frac{y_k e_{j_k}^T}{e_{j_k}^T y_k}\|_1 = \|I - \frac{y_k e_1^T}{e_1^T y_k}\|_1 = \| \begin{pmatrix} 0 & 0 \\ -y_k^2/y_k^1 & 1 \end{pmatrix} \|_1 = 1$$

if $j_k = 1$, and

$$\|I - \frac{y_k e_{j_k}^T}{e_{j_k}^T y_k}\|_1 = \|I - \frac{y_k e_2^T}{e_2^T y_k}\|_1 = \| \begin{pmatrix} 1 & -y_k^1/y_k^2 \\ 0 & 0 \end{pmatrix} \|_1 = 1$$

if $j_k = 2$. So (4.3) follows from (4.4), (4.5) and (4.6) and from the above observations. QED.

Theorem 4.1. *Let $r \in (0,1)$, $n = 2$. There exist ε, $\delta > 0$ such that, if $\|x_0 - x_*\| \le \varepsilon, \|H_0 - J(x_*)^{-1}\| \le \delta$, the sequence generated by Algorithm 2.1 with the choice (4.1) – (4.2) is well defined, converges to x_* and satisfies*

$$\|x_{k+1} - x_*\|_1 \le r\|x_k - x_*\|_1 \tag{4.7}$$

for all $k = 0, 1, 2, \ldots$. Moreover, $\|H_k\|$ and $\|H_k^{-1}\|$ are bounded, (2.15) holds and the convergence is R-superlinear.

Proof. Inequality (4.7) and boundedness of $\|H_k\|$ and $\|H_k^{-1}\|$ follow from (4.3) and the general assumptions of Section 2 using a classical inductive proof. See, for example, the proof of Theorem 3.2 of Broyden, Dennis and Moré [1973]. Therefore, by Theorem 2.1, (2.15) holds for $\|\cdot\|_1$ and, since superlinear convergence is independent of the norm, (2.15) follows for any norm. Finally, by Theorem 2.1, the convergence is R-superlinear. QED.

5 Numerical examples

The implementation of ICUM is based on the formula

$$H_k = H_0 + \sum_{\ell=0}^{k-1} v_\ell e_{j_\ell}^T, \tag{5.1}$$

where

$$v_\ell = \frac{(s_\ell - H_\ell y_\ell)}{e_{j_\ell}^T y_\ell}$$

and

$$x_{k+1} = x_k - H_k F(x_k), \quad k = 0, 1, 2, \ldots.$$

Due to the increasing power of automatic differentiation (see Griewank and Corliss [1991], Griewank [1992]), we prefer to justify the use of quasi-Newton method in terms of savings on linear algebra work rather than on the property of avoiding computation of analytic derivatives. In our implementation, we chose H_0 as the diagonal matrix whose elements are the inverses of the diagonal elements of $J(x_0)$. If one of these derivatives is null, we replace it by 1. We also implemented Broyden's "good" method (Broyden [1965], Dennis and Schnabel [1983]) with the same choice for H_0. Broyden's method corresponds to formula (2.10) with $u_k = H_k^T s_k / \|H_k^T s_k\|$. In this section we wish to show some examples where the method analyzed in this paper turned out to be efficient. As we will see later, we do not claim that ICUM is the best possible quasi-Newton method, but that it is (together with many others) one of the algorithms that provide potentially good affine models for the resolution of (1.1). With this purpose, we will see that, for some classical problems taken from Moré, Garbow and Hillstrom [1981] and other collections, the performance of ICUM is particularly good.

In all the tests we used the convergence criterion

$$\|F(x_k)\| \leq 10^{-5} \, \|F(x_0)\|. \tag{5.2}$$

We also stopped the execution when the number of iterations exceeded 100 and when $\|F(x_k)\| \geq 10^4 \, \|F(x_0)\|$.

The test problems are given below:

Problem 1. Rosenbrock ($n = 2$). Function 1 of Moré, Garbow and Hillstrom [1981]. $x_0^T = (-1.2, 1)$.

Problem 2. Extended Rosenbrock ($n = 50$). Function 21 of Moré, Garbow and Hillstrom [1981]. $x_0^T = (-1.2, 1, -1.2, 1, \ldots)$.

Problem 3. Freudenstein - Roth ($n = 2$). Function 2 of Moré, Garbow and Hillstrom [1981]. $x_0^T = (0.5, -2)$.

Problem 4. Linear system ($n = 50$).

$$F(x) = Ax$$

where

$$[A]_{ij} = 1/(i+j-1) \text{ if } i \neq j,$$

$$[A]_{ii} = 10 + 1/(2i-1), \quad i,j = 1,\ldots,n.$$

$x_0^T = (1,-1,1,\ldots)$.
Observe that, by Gay's theorem, both Broyden and ICUM should converge in $2n$ iterations for linear systems. It can be conjectured that this number can be reduced depending on the eigenvalue structure of the matrix, but, up to our knowledge, nothing has been proved yet in this direction.

Problem 5. Complementarity problem ($n = 8$). We consider the linear complementarity problem (LCP): given an $m \times m$ matrix M and a vector $q \in I\!R^m$, find $w,\ z \in I\!R^m$ such that

$$z = Mw + q, \quad w^T z = 0, \quad w,\ z \geq 0.$$

Dropping the positivity constraints, the LCP defines a nonlinear system of equations. We choose M and q as in Hock and Schittkowski [1990] (example 35) and Kanzow [1993]:

$$m = 4, \quad q = (-8,-6,-4,3)^T.$$

Entries of M (columnwise): 2, 1, 1, -1, 1, 2, 0, -1, 1, 0, 1, -2, 1, 1, 2, 0.
$x_0^T \equiv (w_0^T, z_0^T) = (7/3, 5/3, 13/9, 11/9, 1, 1, 1, 1)^T$.

Problem 6. Complementarity problem ($n = 14$). See Kanzow [1993]. We consider the LCP problem defined by

$$m = 7, \quad q = (-1,-3,1,-1,5,4,-1.5)^T.$$

Entries of M (columnwise): 1, 0, -0.5, 0, -1, -3, 0, 0, 0.5, 0, 0, -2, -1, 1, -0.5, 0, 1, 0.5, -1, -2, 4, 0, 0, 0.5, 0.5, -1, 1, 0, 1, 2, 1, 1, 0, 0, 0, 3, 1, 2, -1, 0, 0, 0, 0, -1, -4, 0, 0, 0, 0.
$x_0^T \equiv (w_0^T, z_0^T) = (1.27, 3.09, 1, 1.54, 1.45, 1, 1, 0, \ldots, 0)$.

Problem 7. Complementarity problem ($n = 50$). See Kanzow [1993]. We consider the LCP problem defined by

$$m = 25, \quad q = (-1,\ldots,-1)^T.$$

$$[M]_{ii} = 4(i-1) + 1,$$

$$[M]_{ij} = [M]_{ii} + 1 \text{ if } j > i,$$
$$[M]_{ij} = [M]_{jj} + 1 \text{ if } j < i,$$

$i, j = 1, \ldots, m$.
$x_0^T \equiv (w_0^T, z_0^T) = (1, \ldots, 1)$.

Problem 8. Chandrasekhar H-equation (Moré [1990], Chandrasekhar [1960]), Kelley [1995] ($n = 50$). Chandrasekhar introduced the following integral equation in the context of radiative transfer problems: find $x \in C[0,1]$ such that

$$x(t) = 1 + \frac{c}{2} \int_0^1 \frac{tx(t)x(y)}{t+y} dy$$

As in Kelley [1995], we approximate the integral using the midpoint composite rule with equally spaced nodes $\{t_i\}_{i=1}^n$. So, $t_i = (i - 1/2)n, 1 \leq i \leq n$. This gives

$$x_i = x(t_i) = 1 + \frac{c}{2n} \sum_{j=1}^{n} \frac{t_i x_i x_j}{t_i + t_j}$$

or

$$F(x) = 0\,, \quad F = (f_1 \cdots f_n)^T\,,$$

where

$$f_i(x) = -x_i + 1 + \frac{c}{2n} \sum_{j=1}^{n} \frac{t_i x_i x_j}{t_i + t_j} \quad .$$

$x_0 = (0, \ldots, 0)^T$.

The numerical results are given in Tables 1 (first 7 problems) and 2 (problem 8 for different values of c). In these tables "C, k" means that the method converged using k iterations. In this case, the number of functional evaluations is $k+1$ since we are not using globalization strategies to decrease the value of $\|F(x)\|^2$. "Div" means divergence (a very large value of $\|F(x_k)\|$ was found). Finally, "NC" means that the method did not converge after 100 iterations. For solving the linear systems of Newton's method we used $L - U$ factorizations with partial pivoting. So, Newton uses $n^3/3 + O(n)$ flops per iteration. In these small-scale problems, the linear algebra work associated to ICUM was organized as follows:

Algorithm 5.1.
Step 1. Compute $v = H_k F(x_{k+1})$.

Step 2. Divide v by the j_k-th entry of y_k ($|e_{j_k}^T y_k| = \|y_k\|_\infty$), obtaining w.
Step 3. Add $-w$ to the j_k-th column of H_k, obtaining H_{k+1}.
Step 4. Compute

$$H_{k+1}F(x_{k+1}) = v - \frac{e_{j_k}^T F(x_{k+1})}{e_{j_k}^T y_k} F(x_{k+1}). \tag{5.3}$$

The justification of this procedure comes from formula (2.10). Clearly, the number of flops per iteration with a dense matrix approach for Step 1 is $n^2 + O(n)$. In a similar way, we implemented Broyden's method.

Problem	Newton	Broyden	ICUM
1	C, 2	Div, $\|F(x_3)\| = 0.3E5$	C, 8
2	C, 2	Div, $\|F(x_3)\| = 0.2E6$	C, 8
3	C, 42	Div, $\|F(x_3)\| = 0.2E7$	C, 20
4	C, 1	C, 5	C, 5
5	C, 12	C, 8	C, 9
6	C, 7	C, 14	C, 14
7	C, 1	NC	C, 83

Table 1. Numerical results: First 7 problems.

c	Newton	Broyden	ICUM
0.1	C, 3	C, 4	C, 4
0.5	C, 4	C, 6	C, 7
0.9	C, 5	C, 10	C, 10
0.99	C, 6	C, 13	C, 13
0.999	C, 8	C, 16	C, 16
$1 - 10^{-4}$	C, 9	C, 18	C, 17
$1 - 10^{-5}$	C, 10	C, 26	C, 17
$1 - 10^{-6}$	C, 10	C, 28	C, 17
$1 - 10^{-7}$	C, 11	C, 31	C, 18
$1 - 10^{-8}$	C, 11	C, 28	C, 18
1.	C, 11	C, 33	C, 18

Table 2. Numerical results: Problem 8.

The most economic implementation of Broyden's method, specially suited for large-scale problems, was given by Deuflhard, Freund and Walter [1990] (see also Deuflhard [1995] and Kelley [1995, Algorithm 7.3.1]). The $k-$th iteration can be performed storing only the steps $\{s_j\}_{j=0}^n$. The work per iteration of this implementation consists on computing $H_0F(x_k)$, $k+1$ scalar products and k operations of type $v_1 + \alpha v_2$ (vector + scalar $\times$ vector).

A large-scale implementation of ICUM can be deduced from Algorithm 5.1. In fact assume that, in addition to the information relative to the initial matrix H_0, we store, for all k, the index of the column that is modified at this iteration, j_k, and the corresponding "new column" z_{j_k}. We write $J_0 = \emptyset$ and

$$J_k = \{j_0, \ldots, j_{k-1}\}.$$

Observe that J_k can have less than k elements since some j_ℓ's could be repeated. Call $\bar{H}_k$ the matrix which is identical to H_0 except that the columns $j_0, \ldots, j_{k-1}$ have been replaced by null columns. So, the vector v computed at Step 1 of Algorithm 5.1, is

$$v = \bar{H}_k F(x_{k+1}) + \sum_{j \in J_k} z_j e_j^T F(x_{k+1}). \tag{5.4}$$

The step 3 of Algorithm 5.1 can be rephrased as follows: If $j_k \in J_k$ replace z_{j_k} by $z_{j_k} - w$ and define $J_{k+1} = J_k$, otherwise add $J_{k+1} \leftarrow J_k \cup \{j_k\}$ and define z_{j_k} as $-w$ plus the j_k-th column of H_0. Step 4 stands without modifications, therefore, the computational work of the iteration is concentrated in (5.4). This step uses, at most, k operations of type $v_1 + \alpha v_2$ and a matrix-vector product with $\bar{H}_k$, which is not more complex than H_0. So, roughly speaking, this implementation of ICUM saves k scalar products in relation to the cheapest implementation of Broyden's method.

6 Final remarks

In this paper, we proved new convergence properties of the inverse Column-Updating method:

(a) If linear convergence is assumed, the convergence is R-superlinear.

(b) If the method is restarted periodically (not necessarily with true Jacobians) we obtain local and R-superlinear convergence.

(c) If $n = 2$, local and R-superlinear convergence are obtained without restarts.

Secant methods provide affine models for generating iteratively approximations to the solution of a nonlinear system. Frequently, the "best" model is the one given by Newton's method, which, in general, is expensive. However, in some cases, the model provided by the inverse Column-Updating method is better than Newton's model and than other popular quasi-Newton models. Our present feeling, derived from many numerical experiments, is that there exists a family of *reliable* quasi-Newton methods that can be considered optimal in a "multiobjective" sense. By this we mean that, if we consider a sufficiently large set of test problems, none of these methods is better than none of the others for *all* the problems. So, all the methods in the *optimal set* deserve to be included in a comprehensive library for solving nonlinear systems. We also think that theoretical convergence results like the ones presented in this paper indicate that ICUM may belong to the optimal set. Further research is necessary in order to discover other theoretical properties characterizing this type of optimality.

We have the strong feeling that savings on linear algebra work will be the main reason for using a particular quasi-Newton method provided that some stability properties are satisfied. The theoretical results proved in this paper seem to show that ICUM can be reliable in many situations and the numerical results indicate that these theoretical properties are reflected in practical computations. To the best of our knowledge, ICUM is the less expensive secant method for nonlinear systems (having good convergence properties). This suggests that this method can be one of the first choices when we wish to solve a practical problem.

Acknowledgements. We are indebted to two anonymous referees for their comments on the first draft of this paper and to Peter Deuflhard for his observations on the implementation of Broyden's method.

References

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

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

Chandrasekhar, S. [1960]: *Radiative Transfer*, Dover, New York.

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

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

Dennis, J. E. and Schnabel, R. B. [1983]: *Numerical methods for unconstrained optimization and nonlinear equations*, Prentice-Hall, Englewood Cliffs, N. J..

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

Deuflhard, P. [1995]: *Newton techniques for highly nonlinear problems. Theory and algorithms*, in preparation.

Deuflhard, P., Freund, R. and 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.

Fletcher, R. [1987]: *Practical methods for optimization*, John Wiley and sons, Chichester, New York, Brisbane, Toronto and Singapore.

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

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

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

Griewank, A. [1992]: Achieving logarithmic growth of temporal and spatial complexity in reverse automatic differentiation, *Optimization Methods and*

Software 1, pp. 35-54.

Griewank, A. and Corliss, G. [1991]: *Automatic differentiation of algorithms: theory, implementation and application*, SIAM, Philadelphia.

Hock, W. and Schittkowski, K. [1990]: *Test examples for nonlinear programming codes*, Lecture Notes in Economics and Mathematical Systems 187, Springer Verlag, Berlin.

Kanzow, C. [1993]: Global convergenge properties of some iterative methods for linear complementarity problems, Preprint 72, Institut für Angewandte Mathematik der Universität Hamburg.

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

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

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

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

Martínez, J. M. [1993]: On the convergence of the Column-Updating method, *Matemática Aplicada e Computacional* 12, pp. 83-94.

Martínez, J. M. and 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.

Moré, J.J. [1990]: A collection of nonlinear model problems. In E.L. Allgower and K. Georg, editors, *Computational solution of nonlinear systems of equations*, volume 26 of *Lectures in Applied Mathematics*, pages 723-762. American Mathematical Society, Providence, RI.

Moré, J. J., Garbow, B. S. and Hillstrom, K. E. [1981]: Testing unconstrained optimization software, *ACM Transactions on Mathematical Software* 7, pp. 17-41.

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

Schwetlick, H. [1979]: *Numerische Lösung Nichtlinearer Gleichungen*, Deutscher Verlag der Wissenschaften, Berlin.

6 A carta para Burdakov

Campinas, February 27, 1995

Prof. Dr.
Oleg Burdakov
Editor of *Optimization methods and software*
CERFACS
Parallel Algorithms Team
42, Avenue Gustave Coriolis
31057 Toulouse Cedex
FRANCE

Dear Oleg:

Based on the new report of Referee 1, we wrote a new version of *Convergence properties of the inverse column-updating methdod.* Although I think that we must wait for the report of Referee 2 on the second version, we thought (since Referee 1 was the most critical one about the first version) that this second revised version can be useful.

As I told you in an e-mail, our code of the Chandrasekhar equation used a special ad-hoc device for avoiding problems at the singularity. Since this could be rather artificial, we decided to run again the problem using the rule used by Kelley in the book *Iterative methods for linear and nonlinear equations.* Moreover, to provide more interesting experiments, we used a finer grid. To avoid further misinterpretations, we are sending a copy of the subroutine that evaluates the function and the Jacobian, (without special savings just to make it easily readable).

We referenced the implementation of Broyden's method originated in the paper of Deuflhard, Freund and Walter. Consequently, we described a large-scale implementation of ICUM, which is "k scalar products" more economic than Broyden (the storage is the same).

Best regards,