Practical Quasi-Newton algorithms for singular nonlinear systems

Sandra Buhmiler* Nataša Krejić† Zorana Lužanin‡

January 25, 2010

Abstract

Quasi-Newton methods for solving singular systems of nonlinear equations are considered in this paper. Singular roots cause a number of problems in implementation of iterative methods and in general deteriorate the rate of convergence. We propose two modifications of QN methods based on Newton's and Shamanski's method for singular problems. The proposed algorithms belong to the class of two-step iterations. Influence of iterative rule for matrix updates and the choice of parameters that keep iterative sequence within convergence region are empirically analyzed and some conclusions are obtained.

Key words: nonlinear system of equations, singular system, Quasi-Newton method, local convergence

AMS Classification: 65H10, 47J20

1 Introduction

Consider the problem of solving a system of nonlinear equations

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

where $F : \mathbb{R}^n \to \mathbb{R}^n$ is nonlinear mapping. We assume that there exists a solution $x^* \in \mathbb{R}^n$. If F is continuously differentiable in the neighborhood of x^*, F' is Lipschitz continuous at x^* and $F'(x^*)$ is a nonsingular matrix, the problem (1) is called nonsingular and x^* is a nonsingular root. If $F'(x^*)$ is a singular matrix, then the problem (1) is singular and x^* is a singular root.

*Department of Mathematics, Faculty of Engineering, University of Novi Sad, 21000 Novi Sad, Serbia

†Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad Serbia, e-mail: natasak@uns.ac.rs, Research supported by grant no. 144006, Ministry of Science, Republic of Serbia

‡Department of Mathematics and Informatics, Faculty of Science, University of Novi Sad, Trg Dositeja Obradovića 4, 21000 Novi Sad Serbia, e-mail: zorana@dmi.uns.ac.rs, Research supported by grant no. 144006, Ministry of Science, Republic of Serbia

If (1) is nonsingular then the most famous iterative method is the Newton method

$$x^{k+1} = x^k - F'(x^k)^{-1} F(x^k)$$

which possesses good theoretical characteristics such as quadratic convergence and affine invariance. Despite its obvious qualities, this method has a number of disadvantages in practice. It is primarily the case with the high cost of a Newton iteration, due to necessity to compute all elements of the Jacobian matrix, as well as the need for exact solution of a linear system using new matrix for every iteration. Therefore, many modification of Newton method have been proposed. One group of modifications are Quasi-Newton methods. The basic idea behind any Quasi-Newton method is to eliminate computation of the Jacobian in every iteration. Detailed overviews of Quasi-Newton methods are presented in [18] and [31].

When the problem (1) is singular, the performance of Newton and Quasi-Newton methods deteriorates. A number of proposals for modification of the Newton method for singular systems have been considered in literature. Roughly speaking, we can distinguish between two classes of modifications. The first class are methods with two-step iteration like the ones presented in [13] for Newton's method, [14] for Shamanski's method and in [15] for inexact Newton's method. The second class of modifications are tensor methods introduced in [27], [7], [1]. The basic idea behind the tensor method is to use the Hessian, i.e., the second derivative of F and thus achieve better convergence rate for singular problems. Theoretical analysis of both classes of methods for singular problems is cumbersome and technically demanding. In the last decade, analysis of continuous Newton method is in focus too, see [10],[11]. Singular problems and continuous methods are considered in [12],[20],[23],[24],[25], [26]. In [26] it is shown that the quadratic converegence to singular roots can be recovered by using two-stage integration schemes for the continuous Newton method, which is the extension of the results derived heuristically by Kelley and Suresh [13], for their two-step iteration scheme.

In this paper we consider the class of singular problems that is characterized by $\text{rank}(F'(x^*)) = n - 1$ or the so-called problems with regular singularity. For such problems we propose two possible modifications of QN methods and investigate their properties empirically using a collection of test problems. The considered QN methods - Broyden's method, Martínez's Column Updating Method and Thomas' method are well known for their good behaviour in the case of nonsingular systems, see [29]. Thus their modification for singular problems is an attractive possibility. The application of the Broyden method to singular problems is already analysed in [6]. Based on the theoretical results available for the Newton, Shamanski and Inexact Newton method we conclude that QN methods should behave better if the same kind of two-step iterative procedure is applied. Therefore we propose two possibilities for such modification and analyse their numerical behaviour. All considered QN methods use two consequtive iterations for the matrix update. Since we are dealing with a two-step iterative procedures each iterative step yields two iteration points, mid-iteration v^k and full-iteration x^{k+1}. So two obvious possibilities arise for the update rule of QN matrix - either two consequtive iterations or mid- and full iteration can be used to derive the update of QN matrix. The results presented in this paper strongly favour the update with full iterations only. Conclusions in this paper are based only on numerical results but the presented results suggest that the proposed modification has useful properties. Theoretical analysis will be the subject of future research.

This paper is organized as follows: in Section 2 we consider behaviour of Quasi-Newton methods near singularities, practical acceleration algorithms for a nonlinear system at singular roots are introduced in Section 3, numerical results and comparisons are given in Section 4 and conclusions are presented in the last section. Throughout this paper we will use the Euclidean norm denoted by $\|\cdot\|$.

2 The behaviour of Newton-like method near singularities

The convergence behavior of an iterative sequence depends on the nature of singularity of $F'(x^*)$. Therefore, during our analysis we assume the following singular assumptions, [6].

A1. F is twice Lipschitz continuously differentiable.

A2. $\text{rank}(F'(x^*)) = n-1$.

A3. Let N be the null space of $F'(x^*)$ spanned by $\varphi \in \mathbb{R}^n$ and X the range space such that $\mathbb{R}^n = N \oplus X$. For any projection P_N onto N parallel to X we assume

$$P_N F''(x^*)(\varphi, \varphi) \neq 0.$$

If singular assumptions A1-A3 are satisfied then $F'(x)$ is singular on the manifold $\mathcal{S} = \{x \in \mathbb{R}^n | \det F'(x) = 0\}$ and $x^* \in \mathcal{S}$. Also $\mathcal{S}$ is transversal to φ. All these sets are illustrated at Figure 1 by a simple example.

Assumptions A1-A3 can be modified and generalized. The assumption $F''(x^*)(\phi, \phi) \notin \text{range}(F'(x^*))$ is an alternative statement of assumption A3, but note that this need not assume that the range is transversal to N, see [24],[25].

Example 1. [22]

$$F(x) = \begin{bmatrix} x_1 + x_1 x_2 + x_2^2 \\ x_1^2 - 2x_1 + x_2^2 \end{bmatrix}$$

The singular root is $x^* = (0,0)^T$, null space is $N = \text{span}(0,1)$ and range space is $X = \text{span}(1,-2)$. The Jacobian $F'(x)$ is singular on the hyperbole given by $2x_1 - 2x_1^2 + 6x_2 - 4x_1x_2 + 2x_2^2 = 0$.

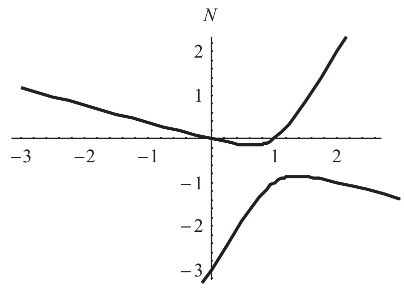


Figure 1

The first and the most important results about Newton method in singular case are presented in [22] and [8], one analysis of the Broyden method is given in [6] and properties of the chord or fixed Newton method for singular problem are analyzed in [5].

It is clear that the proper choice of x^0 is much more difficult in the singular case due to singularity of Jacobian on the manifold $\mathcal{S}$. Therefore the region for initial approximation must be modified. The initial ball from the nonsingular problem $\mathcal{B}(\rho) = \{x \in \mathbb{R}^n | \; ||x^* - x|| \leq \rho\}$ is replaced by a cone defined by

$$W(\rho, \theta, \nu) = \{x \in \mathbb{R}^n \| \; 0 < ||x^* - x|| < \rho, \;\; ||P_X(x - x^*)|| \leq \theta ||P_N(x - x^*)||^\nu\}.$$

For $\nu = 1$ the set $W(\rho, \theta, \nu)$ is a cone which corresponds to folded singularities and it is the main one considered in this paper. The class of methods we considered can be described by the general algorithm presented below.

Algorithm QN: Quasi-Newton

Let $x^0 \in \mathbb{R}^n$ and $B_0 \in \mathbb{R}^{n \times n}$ be given.

For $k = 0, 1, 2, \ldots$

Step 1. Compute s^k from

$$B_k s^k = -F(x^k). \tag{2}$$

Step 2. Define $x^{k+1} = x^k + s^k$.

Step 3. Update the approximation of the Jacobian

$$B_{k+1} = G(B_k, x^{k+1}, x^k),$$

where G is an iterative function for approximation of the Jacobian.

Clearly, each QN method of this class is determined by a particular function G in Step 3. We will consider the following four QN methods.

1. **Fixed Newton method**

$$B_{k+1} = G(B_k) = B_0, \quad k = 0, 1, 2, \ldots.$$

The method is very cheap, but the convergence for nonsingular problems is only linear if B_0 is a good enough approximation of $F'(x^*)$ and x^0 is close enough to x^*.

2. **Broyden's method**, [2]. The update rule is given by

$$B_{k+1} = G(B_k, x^k, x^{k+1}) = B_k + \frac{(y^k - B_k s^k){s^k}^T}{(s^k, s^k)}, \quad k = 0, 1, 2, \dots, \tag{3}$$

where $y^k = F(x^{k+1}) - F(x^k)$ and $s^k = x^{k+1} - x^k$. The Broyden method is the most popular QN method. Under the standard (nonsingular) assumptions there exist $\varepsilon > 0$ and $\delta > 0$ such that if $||B_0 - F'(x^*)|| < \varepsilon$ and $||x^0 - x^*|| < \delta$ then the sequence $\{x^k\}$ obtained by the Broyden method converges superlinearly.

3. **Martínez's Column Updating Method - MCUM**, [17]

   Let $\alpha \in (0, \frac{1}{\sqrt{n}})$ and $j \in \{1, 2, \dots, n\}$ be an index such that $|s_j| > \alpha||s^k||$. Let us denote by $I_j^k \subseteq \{1, \dots, n\}$ the set of indices of elements from jth column which should be modified. Then the matrix B_{k+1} differs from B_k only in the jth column,

$$b_{ij}^{k+1} = \begin{cases} \frac{y_i - \sum_{l \neq j} b_{il}^k s_l^k}{s_j^k}, & i \in I_j^k \\ b_{ij}^k, & i \notin I_j^k \end{cases}, \quad k = 0, 1, 2, \dots \tag{4}$$

   Theoretically, this method is superlinearly convergent only if restarted i.e., $B_{i \cdot m} = F'(x^{i \cdot m})$ for some $m \in \mathbb{N}$, and $i = 1, 2, \dots$. However in practice it converges without a restart. Some modifications of this method are presented in [19].

4. **Thomas' method**, [30]. The update rule is given by

$$B_{k+1} = G(B_k, x^k, x^{k+1}) = B_k + \frac{(y^k - B_k s^k){d^k}^T}{(d^k, s^k)}, \quad k = 0, 1, 2, \dots, \tag{5}$$

   where $d^k = (P_k + \frac{\|s^k\|}{2} I)s^k$, $P_{k+1} = (1 + \|s^k\|)(\|s^k\| I + P_k - \frac{d^k {d^k}^T}{(d^k, s^k)})$, P_0 is given and $y^k = F(x^{k+1}) - F(x^k)$. If the standard assumptions are satisfied then the Thomas method is locally superlinearly convergent, [30], [29].

   Thomas' method belongs to the same class of rank one updates as Broyden's method. It was introduced in [30]. Although it has not seen wide application, probably due to lack of proper understanding, [18], some numerical studies show its good properties, [29]. The method applies geometrical sequential estimation techniques in calculation of the Quasi-Newton matrix B_k as an estimate for the Jacobian matrix. Behavior of Thomas' method for a special class semismooth system is investigated in [3]. Although less popular, the method appears to be comptetitive with Broyden's methods.

The following theorems describe properties of Newton's and Broyden's methods for singular problems.

Theorem 2.1 *[6] Assume that singular assumptions A1-A3 hold. Let $x^0 \in W(\rho, \theta, 1)$. If ρ and θ are sufficiently small then the Newton method is well defined and*

$$\lim_{n\to\infty} \frac{||x^* - x^{k+1}||}{||x^* - x^k||} = \frac{1}{2} \tag{6}$$

$$\lim_{n\to\infty} \frac{||P_X(x^* - x^{k+1})||}{||P_N(x^* - x^k)||^2} = 0. \tag{7}$$

Theorem 2.2 *[6] Assume that singular assumptions A1-A3 hold and that γ, ν are given. Let $x^0 \in W(\rho, \theta, 1)$, and*

$$||(B_0 - F'(x^0))P_X|| \leq \gamma\rho, \qquad ||(B_0 - F'(x^0))P_N|| \leq \gamma\rho^2.$$

Then for ρ and θ sufficiently small the Broyden method is well defined and

$$\lim_{n\to\infty} \frac{||x^* - x^{k+1}||}{||x^* - x^k||} = \frac{\sqrt{5}-1}{2} \tag{8}$$

$$\lim_{n\to\infty} \frac{||P_X(x^* - x^{k+1})||}{||P_N(x^* - x^k)||^2} = 0. \tag{9}$$

Newton method is quadratically convergent for nonsingular problems, but for singular problems the convergence rate is only linear with $1/2$ being the asymptotic linear rate. For a nonsingular case, Broyden's method converges superlinearly, while for a singular case the convergence deteriorates to linear, with $(\sqrt{5}-1)/2$ being the asymptotic linear rate. Relations (7) and (9) show that the rate of convergence is not the same for N and X spaces. Convergence is evidently slower in the null space N. The difference in convergence rate for the considered QN method is illustrated using a set of examples below. We report the results for Example 2 in all details in Tables 1-6. The results for Examples 3-7 indicate essentially the same behavior and thus we present them in a condensed form in Tables 7-8. The last considered Example 7 does not fulfill the regular singularity assumption, i.e $rank(F'(x^*)) = n-2$. Nevertheless the results are similar to the other examples as can be seen in Table 9.

Example 2. Let F be defined in Example 1 and take $x^0 = (0.5, 0.8)$. When applying Martínez Column Updating Method to this simple example we have three posibilities. The first one is denoted as MCUM12 and is the method originally defined by Martínez in [17] and denoted by (4) here. MCUM1 updates only the first column of Jacobian, while MCUM2 updates only the second column of Jacobian. Column 2 corresponds to the null space of $F'(x^*)$. The values of iterative sequences $\{x^k\}$ and rates of convergence are presented in Tables 1-4. The results clearly indicate that the best approach is to use only MCUM2 updates i.e. to update only the component in the null space. MCUM1 updates only the component in space X. The results for this particular example show that MCUM2 is linearly convergent, MCUM1 is sublinear, while the rate for MCUM12 can not be clearly defined.

it	x_1^k	x_2^k	$\frac{\|\|x^*-x^{k+1}\|\|}{\|\|x^*-x^k\|\|}$
0	0.500000	0.800000	–
1	-0.041165	0.530522	0.564043
2	0.031498	0.378183	0.713176
3	0.014447	0.321575	0.848236
4	-0.021966	-0.018204	0.088625
5	-0.016718	-0.300130	10.536739
6	-0.007690	0.108152	0.360701
7	-0.001290	0.213124	1.965666
8	-0.004484	-0.000622	0.021238
9	-0.001712	-0.039564	8.748796
10	0.000760	-0.097064	2.451110
11	-0.001579	0.026270	0.271125
12	-0.001048	0.107980	4.103219
13	-0.000584	-0.000098	0.005487
14	-0.000187	-0.008393	14.170156
15	0.000023	-0.013775	1.640832
...	...	...	...
31	0.000000	-0.000065	1.294385
32	-0.000000	-0.000022	0.346329
33	-0.000000	-0.000008	0.373427
34	-0.000000	0.000002	0.270369

Table 1: MCUM12, $B_0 = F'(x^0)$

it	x_1^k	x_2^k	$\frac{\|\|x^*-x^{k+1}\|\|}{\|\|x^*-x^k\|\|}$
0	0.500000	0.800000	–
1	-1.040000	0.910000	1.464830
2	-0.378924	-1.367040	1.026532
3	0.374792	2.407648	1.717659
4	-0.286696	-3.553462	1.463081
5	0.160703	7.771839	2.180500
6	-0.324485	-6.711689	0.864415
7	-1.672315	-52.588910	7.830236
8	-0.250115	-5.943120	0.113054
9	-0.148141	-5.334998	0.897228
10	0.500350	-2.925186	0.556050
	...	...	...
30	0.000021	-0.004262	0.442657
31	0.000015	-0.000570	0.133711
32	0.000005	0.001629	2.857754
33	-0.000002	0.003839	2.357061
	...	...	...
48	0.000000	0.000031	0.432544
49	0.000000	0.000008	0.257287
50	0.000000	-0.000001	0.150652

Table 2: MCUM12, $B_0 = E$

it	x_1^k	x_2^k	$\frac{\|\|x^*-x^{k+1}\|\|}{\|\|x^*-x^k\|\|}$
0	0.500000	0.800000	–
1	-0.041165	0.530522	0.564043
2	0.031498	0.378183	0.713176
3	0.014447	0.321575	0.848236
4	0.006965	0.251758	0.782400
5	0.003880	0.198532	0.788431
6	0.002217	0.156050	0.785951
7	0.001483	0.128098	0.820852
8	0.000516	0.088208	0.688564
9	0.001040	0.103339	1.171572
10	0.000772	0.090810	0.878744
11	0.000458	0.073349	0.807708
12	0.000443	0.069784	0.951395
13	0.000460	0.070988	1.017262
14	0.000261	0.055398	0.780376
15	0.000258	0.053570	0.966994
	...	...	...
31	0.000097	0.033435	0.900999
32	0.000091	0.032061	0.958920
33	0.000103	0.034093	1.063360
34	0.000040	0.023299	0.683391

Table 3: MCUM1, $B_0 = F'(x^0)$

it	x_1^k	x_2^k	$\frac{\|\|x^*-x^{k+1}\|\|}{\|\|x^*-x^k\|\|}$
0	0.500000	0.800000	–
1	-0.041165	0.530522	0.564043
2	0.200706	0.023434	0.379747
3	0.002728	0.193727	0.958806
4	-0.000608	0.162746	0.840005
5	0.001599	0.060361	0.371016
6	-0.000168	0.051306	0.849695
7	-0.000754	0.094054	1.833239
8	-0.000012	0.033631	0.357563
9	-0.000010	0.025399	0.755226
10	0.000001	0.014356	0.565207
11	-0.000000	0.009293	0.647355
	...	...	...
20	0.000000	0.000120	0.617114
21	-0.000000	0.000074	0.618696
22	0.000000	0.000046	0.617554
23	-0.000000	0.000028	0.618380
24	0.000000	0.000018	0.617784
25	-0.000000	0.000011	0.618214
26	0.000000	0.000007	0.617904

Table 4: MCUM2, $B_0 = F'(x^0)$

As expected, we see that the convergence of Martínez's MCUM depends on the choice of column for modification and the best way is to update the column which corresponds to null space. But detection of null space is impossible when the solution is unknown. Then the best column for modification is generally unknown which is why we shall proceed with the original Martínez method.

The results for Thomas's method are presented in Tables 5 and 6. The ratio of errors in two consecutive iterations indicates behavior similar to Broyden's method, like in nonsingular case, [18] and semismooth case, [3].

it	x_1^k	x_2^k	$\frac{\|\|x^*-x^{k+1}\|\|}{\|\|x^*-x^k\|\|}$
0	0.500000	0.800000	–
1	-0.041165	0.530522	0.564043
2	0.043229	0.353589	0.669443
3	-0.001107	0.261983	0.735456
4	-0.000779	0.126238	0.481859
5	-0.000058	0.084496	0.669329
6	0.000032	0.051488	0.609358
7	0.000007	0.032241	0.626187
8	-0.000001	0.019784	0.613628
9	-0.000001	0.012217	0.617487
10	-0.000000	0.007547	0.617805
11	0.000000	0.004671	0.618853
12	0.000000	0.002888	0.618224
13	-0.000000	0.001784	0.617867
14	-0.000000	0.001102	0.617839
15	-0.000000	0.000681	0.618036

Table 5: Thomas' method, $B_0 = F'(x^0)$,

it	x_1^k	x_2^k	$\frac{\|\|x^*-x^{k+1}\|\|}{\|\|x^*-x^k\|\|}$
0	0.500000	0.800000	–
1	-1.040000	0.910000	1.464830
2	-0.440601	-1.154596	0.894269
3	0.903702	4.726926	3.894244
4	-0.185928	-1.725719	0.360663
5	-0.065560	-2.774219	1.598769
6	-0.171779	-0.945279	0.346221
7	-0.090057	-0.613697	0.645603
8	-0.002694	-0.350650	0.565335
9	0.010065	-0.271554	0.774938
10	0.007278	-0.213237	0.785165
11	0.001028	-0.133014	0.623440
12	-0.001869	-0.053734	0.404207
13	-0.001239	0.001104	0.030866
14	-0.000253	0.032667	19.684711
15	0.000503	0.093199	2.852922

Table 6: Thomas' method, $B_0 = E$

Another remark worth making concerns both Martínez's and Thomas's method. Tables above clearly indicate that the choice of initial approximation B_0 has strong influence on the iterative sequence. That is expected due to the fact that the initial approximation also plays an important role in the nonsingular case. In fact, both Broyden's and Thomas's method belong to the class of QN methods that satisfy the bounded-deterioration principle in the nonsingular case. For such methods we are not able to claim convergence of $\{B_k\}$ to $F'(x^*)$ without additional assumptions, [9]. But the bounded deterioration property is powerful enough to secure the convergence of $\{x^k\}$. Although the initial choice $B_0 = E$ is often successful, better results are obtained with $B_0 = F'(x^0)$. The question of initial approximation B_0 is the subject of many papers particularly in the case of large dimensions. It is evident from this example that proper choice of initial approximations x^0 and B_0 plays an even more important role in the singular case than in the nonsingular case and that $B_0 = F'(x^0)$ is much better. Furthermore, the proof of Theorem 2.2 in [6] suggests that it is essential to take the first iteration as Newton's iteration for the Broyden method. It is quite reasonable to expect the same for MCUM and Thomas' method. From now on we will assume that $B_0 = F'(x^0)$.

Previous conclusions about Martínez's and Thomas' method are checked and confirmed in the next five examples. We also list the solution, null space and initial approximations for all examples.

Example 3. [21]

$$F(x) = \begin{bmatrix} x \\ \frac{10x}{x=0.1} + 2y^2 \end{bmatrix}, \quad x^* = (0,0),\ N = \mathrm{span}(0,1),\ x^0 = (3,1)$$

Example 4.

$$F(x) = \begin{bmatrix} x^2 - y \\ x^2 + y^2 \end{bmatrix}, \quad x^* = (0,0)\ N = \mathrm{span}(1,0),\ x^0 = (3,1)$$

Example 5. [28]

$$F(x) = \begin{bmatrix} x + y - 2 \\ x^2 + y^2 - 2 \end{bmatrix}, \quad x^* = (1,1)\ N = \mathrm{span}(1,-1),\ x^0 = (3,2)$$

Example 6. [4]

$$F(x) = \begin{bmatrix} x + y^2 \\ \frac{3}{2}xy + y^2 + y^3 \end{bmatrix}, \quad x^* = (0,0)\ N = \mathrm{span}(0,1),\ x^0 = (0.5,0.5)$$

Example 7. [14]

$$F(x) = \begin{bmatrix} x + xy + y^2 \\ x^2 - 2x + y^2 \\ x + z^2 \end{bmatrix}, \quad x^* = (0,0,0)\ N = \mathrm{span}(0,1,1),\ x^0 = (1,0.5,1)$$

For all methods we use $B_0 = F'(x^0)$ with $\alpha = 0.7$ in Examples 3-6 and $\alpha = 0.5$ in Example 7 when applying Martínez's method and $P_0 = 0.0005I$ in Thomas' method. The stopping criterion is the number of iterations and for all methods we calculate 30 iterations. In Tables 7 and 8 the results for Examples 3-6 are presented. The last (30th) iteration is given for the three considered Martínez methods (MCUM1,MCUM2, MCUM12) and Thomas' method. Letter D denotes numerical singularity of an update so the iterative process stopped before reaching 30th iteration. Clearly the choice of update column for Martínez's method has strong influence again. A few comments are due concerning the empirical convergence rate $\|x^* - x^{k+1}\|/\|x^* - x^k\|$. Examples 3 and 5 consists of one linear and one nonlinear equation and hence for all convergent methods the empirical rate approaches $(\sqrt{5}-1)/2$, the same as in Broyden's method. In Examples 2 and 4 there is no linear component and hence the empirical rate mainly oscillates around 1 indicating sublinear convergence. The exception is Thomas' method at Example 6 where the empirical rate is again close to $(\sqrt{5}-1)/2$.

Table 9 contains results for Martínez's and Thomas' methods applied to Example 7. As already mentioned, this example does not satisfy the regular singularity condition since $rank(F'(x^*)) = 1$, but the behaviour of all considered methods was again in line with the previous examples. A surprising fact is that Thomas' method generated a sequence with empirical rate oscillating around 0.5 which is the rate of convergence for the Newton method. All Martínez's methods demonstrate sublinear convergence in this example while Table 9 clearly indicates that the best results were obtained using MCUM123 i.e. the method without fixed update column.

	MCUM1	MCUM2
Example	(x_1^{30}, x_2^{30})	(x_1^{30}, x_2^{30})
3	D	$(0,\ -3.88075(-6))$
4	$(-3.01222(-4), -8.08037(-9))$	D
5	$(1,\ 0.999997)$	(1,1)
6	$(-4.62268(-4), 2.75342(-2))$	$(-3.24604(-9), 9.58753(-4))$

Table 7: Martínez's methods for fixed update column

	MCUM12	Thomas'
Example	(x_1^{30}, x_2^{30})	(x_1^{30}, x_2^{30})
3	$(0, -3.1446(-6))$	$(0,\ -3.39033(-6))$
4	$(-3.01222(-4), -8.08037(-9))$	$(-3.27623(-3), 7.79816(-7))$
1.3	(1,1)	$(1, 0.999999)$
1.4	$(8.7883(-8), 3.44936(-4))$	$(9.9435(-13), 2.79228(-6))$

Table 8: Martínez's and Thomas' methods

Method	x^{30}
MCUM1	$(8.65285(-5), 2.06108(-2), 1.13313(-2))$
MCUM2	$(-3.54096(-7), -6.15288(-4), 6.58251(-5))$
MCUM3	$(-1.03782(-5), -2.96838(-3), 3.10629(-3))$
MCUM123	$(-3.45326(-11), 1.35062(-5), 6.5033(-3))$
Thomas	$(-7.64324(-5), -1.5785(-2), 2.05642(-2))$

Table 9: Martínez's and Thomas' methods

3 Algorithms for singular roots

In the previous section we have mentioned that QN methods for singular problem are slower than the same methods for nonsingular case. In fact the convergence in the null space is very slow and it deteriorates the overall convergence rate.

Let us start with the modified Newton method suggested in [13].

Algorithm N: Modification of the Newton method

Let $x^0 \in \mathbb{R}^n$ be given.

For $k = 0, 1, 2, \ldots$

Step 1. Solve $F'(x^k)w^k = -F(x^k)$.

Step 2. Compute $v^k = x^k + w^k$.

Step 3. Solve $F'(v^k)s^k = -F(v^k)$

Step 4. Compute

$$x^{k+1} = v^k + (2 - C \cdot \|s^k\|^\alpha)s^k \tag{10}$$

The iterative sequence $\{x^k\}$ defined by the above algorithm with $C \in \mathbb{R}$ and $\alpha \in (0,1)$ is locally superlinearly convergent, i.e.

$$||x^* - x^{k+1}|| \leq K||x^* - x^k||^{1+\alpha}$$

under assumptions A1-A3. This method generalizes the modified Newton method for the case $f : \mathbb{R} \to \mathbb{R}$. The value 2 which appears in Step 4 given by (10) provides fast convergence of the components in the null space, while $C \cdot \|s^k\|^\alpha$ keeps the modified iterations in the region of convergence W. Obviously two sequences are obtained, where $\{x^k\}$ is the sequence of iterates and $\{v^k\}$ is the sequence of mid-iterates.

The Shamanski method could be easily adapted to handle singular problems with superlinear convergence, [14].

Algorithm S: Shamanski method

Let $x^0 \in \mathbb{R}^n$ be given.
For $k = 0, 1, 2, \ldots$

Step 1. Solve $F'(x^k)w^k = -F(x^k)$.

Step 2. Compute $v^k = x^k + w^k$.

Step 3. Solve $F'(x^k)s^k = -F(v^k)$

Step 4. Compute

$$x^{k+1} = v^k + (4 - C \cdot \|s^k\|^\alpha)s^k \tag{11}$$

Both methods, given by Algorithm N and Algorithm S, can deal with general singular problems with $rank(F'(x^*)) = n - p$ using slightly different rule for the modified iterations. Superlinear convergence of Algorithm S for $p = 1$ is given by the following theorem.

Theorem 3.1 *[14] Assume that singular assumptions A1-A3 hold. Let x^{-1} be given and $x^0 = x^{-1} - F'(x^{-1})^{-1}F(x^{-1})$. Consider the sequence $\{x^k\}$ defined by Algorithm S. Let $\alpha \in (0, (\sqrt{5}-1)/2)$ and $C \neq 0$ Then for ρ, θ and μ sufficiently small and $x^{-1} \in W(\rho, \theta, \mu)$, we have $x^k \in W(\rho, \theta, \mu)$ for all $k \geq 0$, and the sequence $\{x^k\}$ converges $q-$superlinearly to x^* with $q-$order $1 + \alpha$.*

In this paper we suggest two modifications of Quasi-Newton methods based on the ideas of Algorithm N and Algorithm S. The general QN algorithm for singular roots is the following.

Algorithm SQN: Modification of QN method

Let M, $C > 0$, $\alpha \in (0, 1)$, $x^0 \in \mathbb{R}^n$ and nonsingular $B_0 \in \mathbb{R}^{n \times n}$ be given.
For $k = 0, 1, 2, \ldots$

Step 1. Solve

$$B_k w^k = -F(x^k) \tag{12}$$

Step 2. Compute

$$v^k = x^k + w^k \tag{13}$$

Step 3. Update B'_k

Step 4. Solve

$$B'_k s^k = -F(v^k)$$

Step 5. Compute
$$x^{k+1} = v^k + (M - C \cdot \|s^k\|^\alpha)s^k \tag{14}$$
Step 6. Update B_{k+1}

In the Algorithm SQN two choices arise. The first one is the choice of real parameters M, C, and α, which define the new iteration. In comparison with Algorithm N the role of M is to increase the rate of convergence in the null space while C and α should keep the new iteration within the convergence region. On the other hand, we need two QN matrices - Step 3 and Step 6. We will analyze both question here starting with the choice of update rule for B_k.

If we keep in mind Newton's and Shamanski's method two possibilities arise for B'_k and B_{k+1}. The Shamanski approach leads to the update rule I given below, while the Newton method gives rise to QN update rule II.

I In this algorithm we only use sequence $\{x^k\}$ for the update of B_k. Both the mid-iteration v^k and the new iteration x^{k+1} are obtained from linear systems with the same matrix and different left-hand side vectors. Therefore the algorithm is quite cheap since it requires only one matrix update per iteration. Steps 3 and 6 in Algorithm SQN become:

Step 3. Compute
$$B'_k = B_k \tag{15}$$

Step 6. Compute
$$B_{k+1} = G(B_k, x^k, x^{k+1}). \tag{16}$$

II For each update of B_k we use the last iteration and the last mid-iteration. Steps 3 and 6 are defined by

Step 3. Compute
$$B'_k = G(B_k, x^k, v^k). \tag{17}$$

Step 6. Compute
$$B_{k+1} = G(B_k, v^k, x^{k+1}). \tag{18}$$

Here G is the function for update of matrix B_k and we consider three possibilities (3), (4) and (5). This way we obtain a modification of the Broyden, Martínez and Thomas method, respectively.

Both algorithms are tested in Section 4. Results show that rule I is more efficient. Since it is also cheaper the empirical conclusion would be that it is better than rule II. Clearly, this conclusion is influenced by the collection of test examples used in this paper. On the other hand, proofs in [14] and [15] suggest the same conclusion. Theoretical analyzes will be a subject of future work and here we focus on numerical results presented in the next Section.

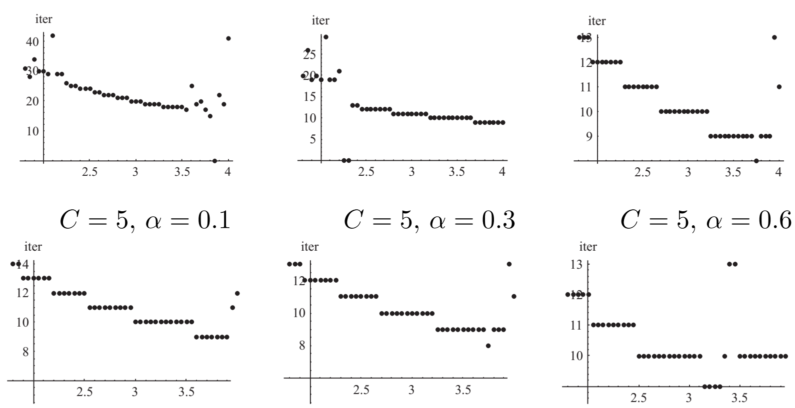
4 Numerical results

In this section we present the results of numerical experiments for both iterative rules I and II in Algorithm SQN. First we will deal with parameters C, α and M in the example we already used. Afterwards we present more test examples analyzing only the influence of M on the number of iterations.

As already explained the role of M is to speed up the convergence in the null space while C and α should keep the sequence $\{x^k\}$ in the convergence region. Optimal values of these parameters are not known. For Newton's method we have $M = 2$ as determined in [13] and for Shamanski's method $M = 4$, see [14]. A very rough guess would be that for QN method with rules I and II, M should be somewhere between 2 and 4. Therefore we tested the case $M \in [2, 4]$.

Let us first present some results of numerical testing for different values of C and α in the following example.

Example 8. Let F be defined in Example 1 and $x^0 = (0.5, 0.8)$. Figure 2 presents the number of iterations needed to satisfy the exit criterion $\|F(x^k)\| \leq 10^{-8}$ versus M for different values of C and α.



$C = 5,\ \alpha = 0.1$ $C = 5,\ \alpha = 0.3$ $C = 5,\ \alpha = 0.6$

$C = 1,\ \alpha = 0.1$ $C = 1,\ \alpha = 0.3$ $C = 1,\ \alpha = 0.6$

Figure 2: MUCM I

This example indicates that the influence of these parameters is not very strong. The same conclusion holds for all other examples and it is in concordance with the tests presented in [15, 14].

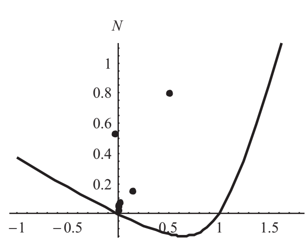


Figure 3:MCUM I, $M = 2.7$

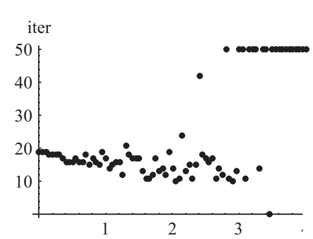


Figure 4: MCUM II, $C = 1; \alpha = 0.6$

it	x_1^k	x_2^k	$\frac{\|\|x^*-x^{k+1}\|\|}{\|\|x^*-x^k\|\|}$
0	0.500000	0.800000	–
1	-0.019404	0.209186	0.222689
2	-0.000910	0.076603	0.364658
3	0.000012	0.032083	0.418789
4	-0.000000	0.011722	0.365365
5	0.000000	0.004394	0.374868
6	-0.000000	0.001620	0.368691
7	0.000000	0.000603	0.372021
8	-0.000000	0.000223	0.369297
9	0.000000	0.000083	0.370982

Table 10: MCUM I, $M = 2.5$,$C = 1$, $\alpha = 0.6$

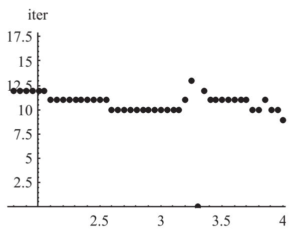


Figure 5: Thomas I

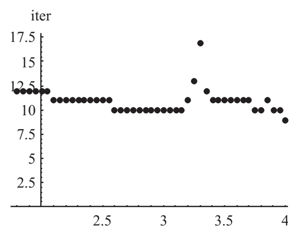


Figure 6: Broyden I

The first 7 iterations of MCUM I method for $M = 2.7, C = 1, \alpha = 0.6$ are given in Figure 3 together with the hyperbole of singular points. The second update rule (II) for B_k in MCUM is illustrated in Figure 4 with the number of iterations required for the exit criterion $\|F(x^k)\| \leq 10^{-8}$ versus M. Table 10 gives the results for MCUM and rule I. Clearly, the results are much better that those given in Tables 1-4. Finally, Figures 5 and 6 show the number of iterations needed for the same exit criterion as before, versus M for the Thomas I and Broyden I method. It is clear from all these cases that for the methods of type I we can clearly distinguish intervals with almost the same number of iterations. The same kind of results is obtained for other test functions that follow. In other words, it was quite possible to fix the value of M for all considered methods and all examples and to get good results. On the other hand, rule II generally produces results like those shown in Figure 4. It appears that the methods of type II are much more sensitive to the choice of M and that "optimal" M depends on both the method and example. Therefore we present only the results for methods of type I for all other examples.

In the examples which follow we claim that the method converges if the condition $||F(x^k)|| \leq 10^{-8}$ is satisfied.

Let us now describe the other nonlinear singular system used for testing in this paper. For the construction of singular system we use the Schnabel-Frank transformation of nonsingular problem, [27]. If $F(x) = 0$ is a standard nonsingular test function with solution x^*, $A \in \mathbb{R}^{n\times k}$ has full column rank with $1 \leq k \leq n$ and

$$\widehat{F}(x) = F(x) - F'(x^*)A(A^TA)^{-1}A^T(x - x^*)$$

then $\widehat{F}(x) = 0$ is a singular problem and $\operatorname{rank}(\widehat{F}(x^*)') = n - \operatorname{rank}(A)$. In the following experiments matrix $A = [1, 1, \ldots, 1]^T$ is used.

The set of test problems is obtained using three functions F_1, F_2, F_3 from [29] while F_4 is from [16]. All $F_i : \mathbb{R}^n \to \mathbb{R}$ and the results are given for $n = 10$ and $n = 100$. The components f_k of F^i as well as the initial approximations are given below.

F_1

$$\begin{aligned} f_k(x) &= x_k - 0.1x_{k+1}^2, \qquad && 1 \leq k < n, \\ f_n(x) &= x_n - 0.1x_1^2, && \\ x^0 &= (2, 2, \ldots, 2)^T && \end{aligned}$$

F_2

$$\begin{aligned} f_1(x) &= x_1, && \\ f_k(x) &= \cos x_{k-1} + x_k - 1, \qquad && 1 < k \leq n, \\ x^0 &= (0.5, 0.5, \ldots, 0.5)^T. && \end{aligned}$$

F_3

$$f_k(x) = \begin{cases} 1 - x_k, & k \equiv_2 1 \\ 10(x_k - x_{k-1}^2), & k \equiv_2 0 \end{cases}$$

$$x^0 = \begin{cases} -1.2, & k \equiv_2 1 \\ 1 & k \equiv_2 0. \end{cases}$$

F_4

$$f_k(x) = \begin{cases} x_k + 10x_{k+1}, & k \equiv_4 1 \\ \sqrt{5}(x_{k+1} - x_{k+2}), & k \equiv_4 2 \\ (x_{k-1} - 2x_k)^2, & k \equiv_4 3 \\ \sqrt{10}(x_{k-3} - x_k)^2, & k \equiv_4 0, \end{cases}$$

$$x^0 = \begin{cases} 3, & k \equiv_4 1 \\ -1, & k \equiv_4 2 \\ 0, & k \equiv_4 3 \\ 1, & k \equiv_4 0. \end{cases}$$

F_i	n	Broyden	MCUM	Thomas'
1	10	20	20	20
1	100	21	21	21
2	10	20	20	20
2	100	21	21	21
3	10	24	24	24
3	100	26	25	26
4	10	23	22	23
4	100	24	24	24

Table 12: QN methods

F_i	n	Broyden	MCUM	Thomas
1	10	8	8	8
1	100	9	9	9
2	10	6	6	6
2	100	8	8	8
3	10	10	10	10
3	100	10	11	10
4	10	9	8	9
4	100	10	10	10

Table 13: Modified QN methods

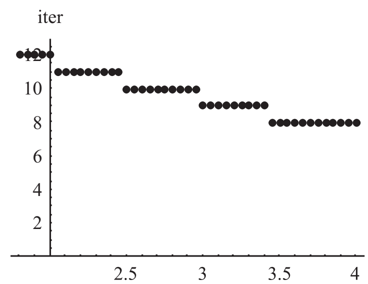


Figure 7: MCUM I, $\widehat{F}_1(x) = 0, n = 10$

it.	$\frac{\|\|x^{k+1}-x^*\|\|}{\|\|x^k-x^*\|\|}$
1	0.197701
2	0.421388
3	0.166771
4	0.462255
5	0.126589
6	0.551790
7	0.084208

Table 14: MCUM I, $\widehat{F}_1(x) = 0, n = 10$

Table 12 gives the results of all three considered QN methods without modification while Table 13 consists of results for the modified QN methods with update rule I and $M = 3.7$. Figure 7 presents the number of iterations needed to satisfy the exit criterion versus M. Finally, Table 14 gives the empirical convergence rate for MCUM I method.

5 Conclusion

The purpose of this paper was to analyze in detail the numerical behavior of Quasi-Newton methods in the case of singular roots and to suggest an improved two-step procedure. We were mainly interested in nonlinear systems with regular singularity ($rank(F'(x^*)) = n-1$). It is well known that the convergence rate of Newton's and Broyden's method deteriorates significantly if one approaches a singular root and hence some modified method should be used. Given that

Martínez' and Thomas' methods are quite competitive with Broyden's method in non-singular and semismooth case we first tested their behavior using a set of small dimensional examples in Section 2. The numerical experiments clearly demonstrate large influence of the initial approximation B_0 and decrease in the convergence rate. The choice $B_0 = F'(x^0)$ seemed to be significantly better that the other commonly used alternative $B_0 = I$, while the convergence rate was linear or sublinear. In Section 3 we suggested two possibilities for two-step modification of the considered QN methods. The idea stems from the well-known modification of Newton's method. The two-step procedure (denoted by I) which uses only whole iterations for QN updates seems to be more stable and is definitely cheaper than the procedure based on mid-iteration and full-iteration. Therefore we adopted the procedure I and tested it on the set of problems of larger dimensions, $n = 10$ and $n = 100$.

The numerical results presented in Section 4 suggested several conclusions. The newly proposed algorithm depends on three real parameters. Two of them are meant to keep iterations in the region of convergence while the third parameter speeds up the convergence in the null space. That parameter M has the strongest influence of all three. In our tests values close to $M = 3.7$ seem to be feasible although such conclusion is influenced by the test collection. Nevertheless one should notice that $M = 2$ is the optimal value for Newton's method while $M = 4$ is valid for Shamanski's method. The convergence of Martínez's method depends heavily on the choice of column to be updated. Not surprisingly the best results are obtained if the update is made on the column that corresponds to the null space. Thomas' and Broyden's two-step method were competitive as in the case of nonsingular and semismooth problems. Although all conclusions are based only on numerical evidence one should notice the consistency of numerical results on problems with different properties. In particular the proposed two-stage algorithms clearly outperformed QN methods without modification in terms of computation effort. Further more the algorithm was able to generate a convergent sequence in all tested cases. Several theoretical questions remain open and certainly deserve further research.

Acknowledgements. The authors are grateful to the anonymous referee for valuable suggestions and comments.

References

[1] Bouaricha, A., Schnabel, R. B.: Tensor methods for nonlinear least squares problems. SIAM J. Sci. Comput. **21**, 1199-1221 (1999)

[2] Broyden, G.: A Class of Methods for Solving Nonlinear Simultaneous Equations. Math.Comp. **19**, 577-593 (1965)

[3] Buhmiler, S., Krejić, N.: A new smoothing quasi-Newton method for nonlinear complementarity problems. J. Comput. Appl. Math. Vol. **211**, 141-155 (2008)

[4] Decker, D.W., Kelley, C.T.: Newton's method at singular point I. SIAM J. Numer.Anal., **17**, 66-70 (1980)

[5] Decker, D.W., Kelley, C.T.: Sublinear convergence of the chord method at singular point. Numer. Math. **42**, 147-154 (1983)

[6] Decker, D. W., Kelley, C. T.: Broyden's method for a class of problems having singular Jacobian at the root. SIAM J. Numer. Anal. **22**, 3, 566-573 (1985)

[7] Feng, D., Frank, P.D., Schnabel, R.B.: Local convergence analysis of tensor methods for nonlinear equations. Math. Prog. **62**, 427-459 (1993)

[8] Griewank, A.: On solving nonlinear equations with simple singularities or nearly singular solutions. SIAM Review **27**, 537-563 (1985)

[9] Griewank, A.: The "global" convergence of Broyden-like methods with a suitable line search. J. Austral. Math. Soc. Ser. B **28**, 75-92 (1986)

[10] Hauser, R., Nedić, J: The continuous Newton-Raphson method can look ahead. SIAM J. Optim. **15**, 915-925 (2005)

[11] Hauser, R., Nedić, J.: On the relationship between the convergence rates of iterative and continuous processes. SIAM J. Optim. **18**, 52-64 (2007)

[12] Janovsky, V., Seigre, V.: Qualitative analysis of Newton's flow. SIAM J. Numer. Anal. **33**, 2068-2097 (1996)

[13] Kelley, C.T., Suresh, R.: A new acceleration method for Newton's method at singular point. SIAM J. Numer. Anal. **20**, 1001-1009 (1983)

[14] Kelley, C.T.: A Shamanskii-like acceleration scheme for nonlinear equations at singular point. Math. Comput. **47**, 609-623 (1986)

[15] Kelley, C.T., Xue, Z.Q.: Inexact Newton methods for singular problems. Optim. Meth. Soft. **2**, 249-267 (1993)

[16] Lukšan,L.: Inexact Trust Region Method for Large Sparse Systems of Nonlinear Equations. J Opt. Theory Appl., **81**, 569-590 (1994)

[17] Martínez, J.M.: A Quasi-Newton Method with Modification of one Column per Iteration. Computing **3**, 353-362. (1984)

[18] Martínez, J.M.: Practical quasi-Newton methods for solving nonlinear systems. J. Comput. Appl. Math. **124**, 97-121 (2000)

[19] de Mendonça, L.F., Pérez, R., Lopes, V.L.R.: Inverse q-Columns updates methods for solving nonlinear system of equations. J. Comput. Appl. Math. **158**, 317-337 (2003)

[20] Neuberger, J.W.: The continuous Newton's method, inverse functions and Nash-Moser. American Math. Monthly **114**, 432-437 (2007)

[21] Nie, P.Y.: A null space method for solving sistem of equations. App. Math. Comp. **149**, 215-226 (2004)

[22] Reddien, G. W.: On Newton's method for singular problems. SIAM J. Numer. Anal. **15**, 5, 993-996 (1978)

[23] Riaza, R.: Attraction domains of degenerate singular equilibria in quasilinear ODEs. SIAM J. Math. Anal. **36**, 678-690 (2004)

[24] Riaza, R.: On the local classification of smooth maps induced by Newton's method. J. Differential Equations **217**, 377-392 (2005)

[25] Riaza, R.: Local dynamics of quasilinear ODEs with folded singular equilibria. Nonlinear analysis **68**, 2242-2249 (2008)

[26] Riaza, R., Zufiria, P.J.: Discretization of implicit ODEs for singular root-finding problems. J. Comput. Appl. Math. **140**, 695-712 (2002)

[27] Schnabel, R.B., Frank, P. D.: Tensor methods for nonlinear equations. SIAM J. Numer. Anal. **21**, 815-843 (1984)

[28] Shen Y.-Q., Ypma, T.J.: Newton's method for singular nonlinera equations using approximate left and right nullspaces of the Jacobian. Appl. Numer. Math. **54**, 2,256-265 (2005)

[29] Spedicato,E., Huang,Z.: Numerical Experience with Newton-like Methods for Nonlinear Algebraic Systems. Computing **58**, 69-89 (1997)

[30] Thomas, S.: Sequential Estimation Tehniques For Quasi-Newton Algorithms, Ph.D. Thesis, TR 75-227,Cornell University, 1975

[31] Yamamoto, T.: Historical developments in convergence analysis for Newton's and Newton-like methods. J. Comput. Appl. Math. **124**, 1-23 (2000)