ANNALES UNIVERSITATIS MARIAE CURIE-SKLODOWSKA

LUBLIN-POLONIA

VOL. XLV, 6

SECTIO A

1991

Instytut Matematyki, UMCS

S. M. GRZEGÓRSKI

Remarks on the Convergence of Newton-like Methods at Regular Singularities

Uwagi o zbieżności metod newtono-podobnych w punktach regularnie osobliwych

Abstract. The paper deals with the convergence behaviour of a class of least-change secant and inverse-secant methods for nonlinear equations at regularly singular roots. It turns out that these methods are locally and Q-linearly convergent with the asymptotic error constant $(\sqrt{5}-1)/2$.

1. Introduction. Consider the system of nonlinear equations

$$(1.1) F(x) = 0,$$

where $F: \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear mapping with the following property:

A1. There exists an x° such that $F(x^{\circ}) = 0$ and F is twice continuously differentiable in a neighbourhood of x° .

Provided $F'(x^*)$ is nonsingular, it is well known (for example [8], [13]) that the Newton method

(1.2)
$$x_{k+1}^N = x_k^N - [F'(x_k^N)]^{-1} F(x_k^N), \quad \text{for } k = 0, 1, \dots$$

converges Q-quadratically to x° . The behaviour of the sequence (1.2) for the problems with singular Jacobian $F'(x^{\circ})$ has been studied by a number of authors [3], [4], [10], [11], [14]. The convergence and the rate of convergence depend on the nature of the singularity of $F'(x^{\circ})$.

Without loss of generality we assume that $F'(x^*)$ is symmetric. If this is not the case, then there exists a nonsingular map L such that $LF'(x^*)$ is symmetric, and it is sufficient to consider LF instead of F.

Let N denote a null space of $F'(x^{\circ})$. P_N be an orthogonal projection onto N and X be a subspace of \mathbb{R}^n orthogonal to N such that $\mathbb{R}^n = X \oplus N$. Let $P_X = I - P_N$. The simplest singular structure occurs when

This research was sponsored in part by the Polish Academy of Science under Grant No CPBP 1.1/4.5

A2. there exists $F''(x^*)$ and $P_N F''(x^*)(v,v) \neq 0$ for $v \in N$, $v \neq 0$.

If the assumption A2 is satisfied then the singularity is called regular. In this case, the rate of convergence of the Newton method is linear and

(1.3)
$$\lim_{k \to \infty} \frac{\|x_{k+1}^N - x^*\|}{\|x_k^N - x^*\|} = \frac{m}{m+1},$$

where $m = \dim(N)$ is the order of singularity. A. Griewank and M. R. Osborne [11] have analyzed the behaviour of the sequence $\{x_k^N\}$ in the neighbourhood of irregular singularities, and then the Newton method converges with a limiting ratio of about $\frac{2}{3}$, or diverges from arbitrarily close starting points, or behaves in a certain sense chaotically.

Although the Newton method has a very nice local convergence property, it has one drawback. That is, we need to evaluate the $n \times n$ Jacobian matrix $F'(x_k)$. To avoid this disadvantage, the Newton-like methods of the form

(1.4)
$$x_{k+1} = x_k - B_k^{-1} F(x_k) , \quad B_k \cong F'(x_k) ,$$

have been proposed and studied intensively [1], [2], [5-9], [12]. When $F'(x^*)$ is nonsingular, the matrices B_k satisfy the secant equation

$$(1.5) B_{k+1}(x_{k+1}-x_k) = F(x_{k+1}) - F(x_k)$$

and

$$||B_{k+1} - B_k|| \to 0 \quad \text{as } k \to \infty ,$$

then the sequence (1.4) is locally and Q-superlinearly convergent to x° . For the Broyden method the matrix updating is given by

(1.7)
$$B_{k+1} = B_k + \frac{F(x_{k+1})s_k^T}{s_k^T s_k} \quad \text{for } k = 0, 1, \dots,$$

where $s_k = x_{k+1} - x_k$. Assuming A1, A2 and

> $m = \dim(N) = 1.$ A3.

D. W. Decker and C. T. Kelley [5] have proved that if the starting point zo is chosen in a special region, B_0 is nonsingular and sufficiently close to $F'(x^*)$, then the sequence (1.4) with the update (1.7) converges Q-linearly to x^{\bullet} and

2 762 webtienen en hundföhrer i

0.

(1.8)
$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{\sqrt{5} - 1}{2},$$

(1.9)
$$\lim_{k \to \infty} \frac{\|P_X(x_{k+1} - x^*)\|}{\|P_N(x_k - x^*)\|^2} =$$

37

Let a sequence of symmetric and positive-definite matrices $\{W_k\}$ be given. Then we define the sequence of norms

(1.10)
$$||X||_k = \{ \operatorname{tr}(X^T W_k X) \}^{1/2} \text{ for } X \in \mathbb{R}^{n \times n}, \ k = 0, 1, \dots$$

In this paper we consider a class of Broyden's methods which are defined as follows: from current approximations x_k and G_k to x° and $F'(x^{\circ})$, respectively, a next iterate is computed by

$$(1.11) x_{k+1} = x_k - G_k^{-1} F(x_k) .$$

and the new approximate Jacobian G_{k+1} is a solution to the problem

(1.12)
$$\min_{G \in R^{n \times n}} \|G - G_k\|_k^2$$

under constraint

(1.13)
$$Gs_k = F(x_{k+1}) - F(x_k), \qquad (s_k = x_{k+1} - x_k).$$

The solution of this subproblem has the form

$$G_{k+1} = G_k + \frac{F(x_{k+1})s_k^T W_k^{-1}}{s_k^T W_k^{-1} s_k} .$$

If we denote $p_k = W_k^{-1} s_k$, then

(1.14)
$$G_{k+1} = G_k + \frac{F(x_{k+1})p_k^T}{p_k^T s_k}, \quad p_k \in \mathbb{R}^n$$

If $W_k = I$, i.e. $p_k = s_k$, then we get the Broyden method.

We want to know if the rate of convergence of the sequence $\{x_k\}$ with the update (1.14) is dependent on the choice of the sequence $\{p_k\}$. We will show that the update (1.14) with every sequence $\{p_k\}$ such that for $\alpha > 0$

(1.15)
$$|p_k^T s_k| \ge \alpha ||p_k|| ||s_k||, \quad p_k \ne 0 \quad k = 0, 1, \dots,$$

guarantees local Q-linear convergence of the sequence (1.11). The relations (1.8) and (1.9) are also satisfied. It means that the rate of convergence of the method (1.11), (1.14) is independent of the choice of the sequence $\{p_k\}$, provided (1.15) holds. This fact is proved in Section 2. In Section 3 we consider the inverse secant method, where the sequence $\{x_k\}$ is defined as follows: given x_k and H_k , the next iterate is computed by

.

$$(1.16) x_{k+1} = x_k - H_k F(x_k)$$

and

(1.17)
$$H_{k+1} = H_k - \frac{H_k F(x_{k+1}) q_k^T}{q_k^T y_k}$$

S. M. Grzegorski

where $y_k = F(x_{k+1}) - F(x_k)$ and $q_k \in \mathbb{R}^n$ is such that some analogous condition to (1.15) holds.

In this case the sequence $\{x_k\}$ also satisfies the properties (1.8) and (1.9). Numerical effectiveness of the algorithm (1.16) with the update (1.17) has been verified on four singular problems.

2. Analysis of convergence. We state the rate of convergence of the sequence (1.11) with the update (1.14). For this reason, in the same way as D. W. Decker and C. T. Kelley [5], we define the set of allowable starting points in the following way

(2.1)
$$U(\delta, \theta, \nu) = \left\{ x \in \mathbb{R}^n : 0 < \|x - x^*\| < \delta, \|P_X(x - x^*)\| \le \theta \|P_N(x - x^*)\|^{\nu} \right\}$$
for $\delta > 0, \ \theta > 0, \ \nu > 0$.

Theorem 1. Assume A1, A2, A3. Let $x_0 \in U(\delta, \theta, 1)$, $\gamma > 0$, $\nu > 0$ and the matrix $G_0 \in \mathbb{R}^{n \times n}$ be such that

$$\|(G_0 - F'(x_0))P_X\| \leq \gamma \delta$$

$$\|(G_0 - F'(x_0))P_N\| \le \mu \delta^2$$

If the sequence $\{p_k\}$ satisfies (1.15), then for sufficiently small δ and θ the sequence (1.11) with the update (1.14) converges to x° and

(2.4)
$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{\sqrt{5} - 1}{2},$$

(2.5)
$$\lim_{k\to\infty} \frac{\|P_X(x_k-x^*)\|}{\|P_N(x_k-x^*)\|^2} = 0.$$

Proof. Let us denote $N = \operatorname{span}(\Phi)$, $\|\Phi\| = 1$, $P_N(x_k - x^{\circ}) = \delta_k \Phi$, $\delta_0 > 0$,

$$\lambda_{k} = \delta_{k+1}/\delta_{k} \quad (i.e. \ P_{N}(x_{k+1} - x^{*}) = \lambda_{k}P_{N}(x_{k} - x^{*})), \\ A_{k} = P_{X}G_{k}P_{X} , \quad B_{k} = P_{X}G_{k}P_{N} , \quad C_{k} = P_{N}G_{k}P_{X} , \\ D_{k} = P_{N}G_{k}P_{N} , \quad u = F''(x^{*})(\Phi, \Phi) .$$

and let $\beta_m(x)$ denote an element of \mathbb{R}^n or an operator defined on \mathbb{R}^n such that

(2.6)
$$\|\beta_m(x)\| = O(\|x - x^*\|^m).$$

In the same way as D. W. Decker and C. T. Kelley [5], we can see that

$$\begin{aligned} x_1 &\in U(\delta, \theta_1, 2) , \quad \theta_1 < \theta , \\ P_X(x_1 - x^*) &= \theta \beta_2(x_0) + \beta_3(x_0) , \\ P_N(x_1 - x^*) &= \lambda_0 \delta_0 \Phi , \quad \text{where } \frac{3}{4} < \lambda_0 < \frac{4}{4} , \end{aligned}$$

and $x_0 \in U(\delta, \theta_0, 2)$ with $\theta_0 = \theta/\delta$. The proof is by induction. Assume that for a given k, the matrices G_i are nonsingular for all $i \leq k$ and that

(2.8)
$$P_N(x_{i+1} - x^*) = \delta_{i+1} \Phi = \lambda_1 \delta_i \Phi = \lambda_i P_N(x_i - x^*), \quad \frac{3}{8} < \lambda_i < \frac{4}{5},$$

(2.9)
$$||G_i - F'(x^*)|| \le \epsilon_0 + c_1 \delta_0 \sum_{j=1}^{6} \left(\frac{4}{\delta}\right)^j = \epsilon_i , \quad c_1 > 0 ,$$

(2.10)
$$\theta_{i+1} \leq c_2 \left((\varepsilon_i + \delta_i)(\theta_i + \theta_{i-1}) + \delta_i \right) \leq 1, \quad c_2 > 0$$

$$||A_i^{-1}|| \le 2||V^{-1}||,$$

where V^{-1} denotes the inverse of $F'(x^*)$ when $F'(x^*)$ is restricted to X. In order to show that if (2.7)-(2.11) hold for $i \leq k$, then they also hold for i = k + 1, we notice that

$$\|(G_{k+1} - G_k)v\| \le \frac{\|F(x_{k+1})\| \|v\|}{\alpha \|s_k\|} \quad \text{for } x \in \mathbb{R}^n .$$

(2.12)
$$||G_{k+1} - G_k|| \le \frac{||F(x_{k+1})||}{\alpha ||s_k||}$$

The expansion of F in Taylor series about x^* yields

(2.13)
$$F(x_k) = F'(x^*)(x_k - x^*) + \frac{1}{2} \delta_k^2 u + \beta_3(x_k) ,$$

and since $x_{k+1} \in U(\delta, \theta_{k+1}, 2)$ we get

(2.14) $||F(x_{k+1})|| \le c ||P_N(x_{k+1} - x^*)||^2 + c_2 ||P_X(x_{k+1} - x^*)|| \le c_3 \delta_{k+1}^2$, $c_3 > 0$. The fact that

$$\|P_X s_k\| \le \|P_X(x_{k+1} - x^{\bullet})\| + \|P_X(x_k - x^{\bullet})\| \le \\\le \theta_{k+1} \delta_{k+1}^2 + \theta_k \delta_k^2 \le 2\delta_k^2$$

implies

$$||s_k|| \ge ||P_N s_k|| - ||P_X s_k|| \ge (1 - \lambda_k)\delta_k - 2\delta_k^2 \ge (\frac{1}{5} - 2\delta_k)\delta_k .$$

Since $\delta_{k+1} = \lambda_k \delta_k < \delta_k < \ldots < \delta_0 \leq \delta$, we get

(2.15)
$$||s_k|| \ge \frac{1}{10} \, \delta_k \,, \quad \text{if } \delta \le \frac{1}{20}$$

From the inequalities (2.12), (2.14) and (2.15) we conclude

$$||G_{k+1} - F'(x^*)|| \le ||G_k - F'(x^*)|| + \frac{||F(x_{k+1})||}{\alpha ||s_k||} \le \le \varepsilon_k + c_1 \lambda_k \lambda_{k-1} \dots \lambda_0 \delta_0 \le \varepsilon_k + c_1 \left(\frac{4}{k}\right)^{k+1} \delta_0,$$

where $c_1 \geq 8c_3/\alpha$.

Since $||P_N|| = ||P_X|| = 1$, we also have

$$|P_N G_{k+1} P_X|| = ||P_N (G_{k+1} - F'(x^*)) P_X|| \le \varepsilon_{k+1} .$$

Using the formula $G_{k+1}s_k = G_ks_k + F(x_{k+1})$ we can write

$$G_{k+1}P_Ns_k = G_kP_Ns_k - (G_{k+1} - G_k)P_Xs_k + F(x_{k+1}) .$$

Then $P_N s_k = -(1 - \lambda_k) \delta_k \Phi$ implies that

(2.16)
$$G_{k+1}\Phi = G_k\Phi - \frac{F(x_{k+1})}{(1-\lambda_k)\delta_k} + \frac{(G_{k+1}-G_k)P_Xs_k}{(1-\lambda_k)\delta_k}$$

Furthermore,

$$\frac{\|(G_{k+1} - G_k)P_X s_k\|}{(1 - \lambda_k)\delta_k} \le \frac{\|F(x_{k+1})\| \|P_X s_k\|}{\alpha \|s_k\|(1 - \lambda_k)\delta_k} \le \frac{10c_3\delta_{k+1}^2 2\delta_k^2}{\alpha \delta_k^2}$$

Hence

(2.17)
$$\frac{(G_{k+1} - G_k)P_X s_k}{(1 - \lambda_k)\delta_k} = \beta_2(x_{k+1})$$

Note that $s_k = P_N s_k + P_X s_k = (\lambda_k - 1)\delta_k \Phi + P_X s_k$, and $G_k s_k = -F(x_k)$. From this we obtain

and an Th

(2.18)
$$G_k \Phi = \frac{G_k P_X s_k + F(x_k)}{(1 - \lambda_k) \delta_k}$$

The relations (2.16)-(2.18) and (2.13) yield

(2.19)
$$G_{k+1}\Phi = \frac{G_k P_X s_k + F(x_k) - F(x_{k+1})}{(1 - \lambda_k)\delta_k} + \beta_2(x_{k+1}) = \frac{1}{2}(1 + \lambda_k)\delta_k u + \frac{(G_k - F'(x^*))P_X s_k}{(1 - \lambda_k)\delta_k} + \beta_2(x_{k+1}).$$

The further part of the proof is the same as the proof of Theorem 2.10 ([5]) and we omit it.

3. Inverse secant update methods. In this section, we state analogue of Theorem 1, which is appropriate for rescaled least-change inverse-secant update methods. For solving (1.1) we consider iterative procedures

(3.1)
$$x_{k+1} = x_k - H_k F(x_k), \quad k = 0, 1, \dots,$$

which employ the approximates

$$(3.2) H_k \cong F'(x_k)^{-1}$$

If the matrices H_k satisfy the inverse-secant equation

where $y_k = F(x_{k+1}) - F(x_k)$, then we get inverse-secant algorithms. For instance, the inverse Broyden update has the form

$$H_{k+1} = H_k + \frac{(s_k - H_k y_k) y_k^T}{y_k^T y_k}$$
, if $y_k \neq 0$.

Assuming that a matrix $G_k \cong F'(x_k)$ is nonsingular and $y_k^T G_k^{-1} p_k \neq 0$ and using the Sherman-Morrison-Woodbury formula [13] to the matrix G_{k+1} , given by (1.14), one can see that

$$G_{k+1}^{-1} = G_k^{-1} - \frac{G_k^{-1} F(x_{k+1}) p_k^{t} G_k^{-1}}{p_k^T (s_k + G_k^{-1} F(x_{k+1}))}$$

Now let $H_k = G_k^{-1}$ and $q_k = (G_k^T)^{-1} p_k$. Since $s_k = -G_k^{-1} F(x_k)$ this implies

(3.4)
$$H_{k+1} = H_k - \frac{H_k F(x_{k+1}) q_k^T}{q_k^T y_k}$$

If $p_k = G_k^T y_k$, then we get the inverse Broyden method. Thus, to establish the local linear convergence of the sequence (3.1) with the update (3.4), it suffices to apply Theorem 1, which in this case has the following formulation:

Theorem 1a. Assume A1, A2, A3. Let $x_0 \in U(\delta, \theta, 1)$, $\gamma > 0$, $\mu > 0$ and the matrix $H_0 \in \mathbb{R}^{n \times n}$ be such that

$$\|(H_0^{-1} - F'(x_0))P_X\| \le \gamma \delta , \\ \|(H_0^{-1} - F'(x_0))P_N\| \le \mu \delta^2$$

If the sequence $\{q_k\}$ is such that

$$|q_k^T F(x_k)| \ge \alpha ||s_k|| || (H^{-1})^T q_k ||, \quad q_k \neq 0, \ \alpha > 0,$$

then for every sufficiently small δ and θ the sequence (3.1) with the update (3.4) converges to x^* and

$$\lim_{k \to \infty} \frac{\|x_{k+1} - x^*\|}{\|x_k - x^*\|} = \frac{\sqrt{5} - 1}{2}$$
$$\lim_{k \to \infty} \frac{\|P_X(x_k - x^*)\|}{\|P_N(x_k - x^*)\|^2} = 0.$$

Note that for the inverse-secant method we have

Ċ

$$\operatorname{let}(H_{k+1}) = -\operatorname{det}(H_k)q_k^T F(x_k)/q_k^T y_k .$$

If we choose q_k in such a way that $q_k^T y_k = -q_k^T F(x_k)$, i.e. $q_k^T F(x_{k+1}) = 0$ and $q_k^T F(x_k) \neq 0$, then

$$\det(H_{k+1}) = \det(H_k) \ .$$

This means that there is a possibility to construct an algorithm with the property $det(H_k) = det(H_0)$, for k = 1, 2, ... Moreover, the cost of one step of the Broyden method is $O(n^3)$, while the cost of one step of the inverse-secant algorithm is $O(n^2)$.

Numerical experiments, executed together with A. Wisnioch, show some effectiveness of the inverse secant updates for singular problems. We have taken under consideration the following algorithms:

MN. the Newton method,

- M1. the inverse Broyden method with $H_0 = [F'(x_0)]^{-1}$,
- M2. the inverse Broyden method with $H_0 = I$,
- **M3.** the update (3.4) with $q_k = H_k^T s_k$ and $H_0 = [F'(x_0)]^{-1}$,
- M4. the update (3.4) with $q_k = H_k^T s_k$ and $H_0 = I$.

Here we give the results of the following problems:

Problem 1.	$F(x) = \begin{bmatrix} \exp(x_1^2) - x_1 x_2 - 1 \\ \\ x_1^2 + x_1 x_2^2 + x_2 \end{bmatrix}$
Problem 2.	$F(x) = \begin{bmatrix} x_1 + x_2^2 \\ 1.5x_1x_2 - x_2^2 + x_3^3 \\ x_1^3 + x_3 \end{bmatrix}$
	$F(x) = \begin{bmatrix} x_1 + x_2^2 \\ 1.5x_1x_2 + x_2^2 + x_2^3 \end{bmatrix} ([5])$
Problem 4.	$F(x) = \begin{bmatrix} x_1 + x_2^3 \\ x_1 x_2^2 + x_2^3 + x_2^4 \end{bmatrix} .$

For all methods it was used the same stopping criterion $||x_k||_{\infty} \leq 10^{-6}$. The results contained in Table 1 indicate that, in practice, for some singular problems the matrix $H_0 = I$ assures the better results than the matrix $H_0 = [F'(x_0)]^{-1}$. It happens since the matrix $F'(x_0)$ is almost singular.

Problem	Starting point	Method				
		MN	M1	M2	M3	M4
1	(0.5, 0.05)	16	19	23	23	20
	(1,0.1)	22	24	26	F	25*
2	(0.0001, 0.01, 0.0001)	10	F	15	20	14
	(0.001, 0.05 , 0.005)	14	F	19	F	19
3	(0.01, 0.1)	13	23	21	21	20
	(0.1, 1.0)	18	F	24	27°	29*
4	(0.05, 0.5)	32	46	27	40 *	25
	(0.1, 1.0)	29	45	40	34	46*

Tab 1. Number of iterations required by Methods MN, M1, M2, M3 and M4

An * in the iteration column designates that the result is obtained after one restart, i.e. if the denominator of the formula (3.4) $s_k^T H_k y_k = 0$, then the matrix H_{k+1} is calculated in the same way as the matrix H_0 . An F denotes that at least two restarts were necessary to attain the stopping criterion or the sequence $\{x_k\}$ diverges.

Acknowledgment. The author would like to thank Professor Andrzej Kiełbasiński for his valuable comments, which have greatly improved the paper.

REFERENCES

- Broyden, C. G., A class of methods for solving nonlinear simultaneous equations, Math. Comp., 19, (1965), 577-593.
- [2] Broyden, C. G., Dennis, J. E., Moré, J. J., On the local and superlinear convergence of quasi-Newton methods, J. Inst. Math. Appl., 12, (1973), 223-246.
- [3] Decker, D. W., Keller, H. B., Kelley, C. T., Convergence rates for Newton's method at singular points, SIAM J. Numer. Anal., 20, (1983), 296-314.
- [4] Decker, D.W., Kelley, C. T., Newton's method at singular points I, SIAM J. Numer. Anal., 17, (1980), 66-70.
- [5] 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, (1985), 556-574.

- [6] Dennis, J. E., Moré, J. J., A characterisation of superlinear convergence and its application to quasi-Newton methods, Math. Comp., 28, (1974), 549-560.
 - [7] Dennis, J. E., Moré, J. J., Quasi-Newton methods, motivation and theory, SIAM Rev., 19, (1977), 46-89.
 - [8] Dennis, J. E., Schnabel, R. B., Numerical Methods for Nonlinear Equations and Unconstrained Optimization, Prentice-Hall, Englewood Cliffs, New York 1983.
- [9] Dennis, J. E., Walker, H. F., Convergence theorems for least-change secant update methods, SIAM J. Numer. Anal., 18, (1981), 949-987.
- [10] Griewank, A., Starlike domains of convergence for Newton's method at singularities, Numer. Math., 35, (1980), 95-111.
- [11] Griewank, A., Osborne, M. R., Analysis of Newton's method at irregular singularities, SIAM J. Numer. Anal., 20, (1983), 747-773.
- [12] Grsegórski, S. M., Orthogonal projections on convex sets for Newton-like methods, SIAM J. Numer. Anal., 22, (1985), 1208-1219.
- [13] Ortega, J. M., Rheinboldt, W. C., Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York 1970.
- [14] Reddien, G. W., On Newton's method for singular prolems, SIAM J. Numer. Anal., 15, (1978), 993-996.

STRESZCZENIE

W pracy zajmujemy się zachowaniem zbieżności rzutowych metod siecznych i metod odwrotnych dla układu równań nieliniowych w punktach regularnie osobliwych. Okazuje się, że metody te są lokalnie Q-liniowo zbieżne ze stałą asymptotyczną ($\sqrt{5} - 1$)/2.

(received June 20, 1990)