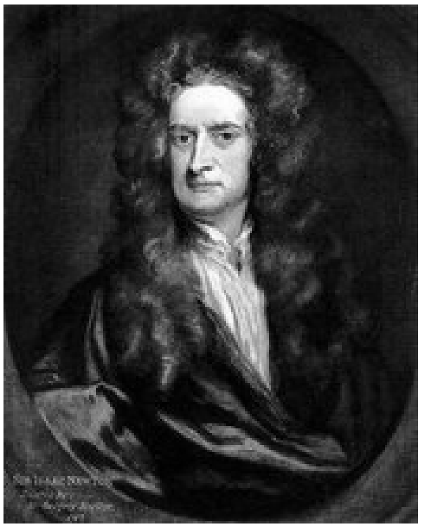
Convergence of
Newton method for solving nonlinear systems

Neculai Andrei

Research Institute for Informatics,
Center for Advanced Modeling and Optimization,
8-10, Averescu Avenue, Bucharest 1, Romania
E-mail: nandrei@ici.ro

One of the best known method for solving systems of nonlinear algebraic equations or for unconstrained optimization is the Newton method. The importance of this method does not consists of only in its excellent convergence properties when it is initialized in a point sufficiently close to the solution, but it is one of the most important ingredient in building up unconstrained and constrained optimization methods. For example, the interior points methods for solving large-scale linear programming problems, or more generally nonlinear programming are based on Newton method. In this work we shall consider some aspects of this method refering to its quadratic convergence for nonlinear algebraic systems of equations solving.



Isaac Newton (1642-1727)

Let us consider a system of nonlinear equations expressed in the form $F(x)=0,$ where $F: R^n \to R^n$ is a vector–valued, continuously differentiable, function of variable x, with components $f_i(x), \quad i=1,\ldots,n$, scalar functions. We are interested in finding x^* such that $F(x^*)=0.$ Let, as usually, $J(x)$ be the Jacobian of F at x. Then, given an estimation x_k of the solution x^*, the Newton's method compute the next estimation x_{k+1} by setting the local linear approximation to F at x_k, i.e.

$$F(x_k)+J(x_k)(x-x_k)=0.$$

Assuming the Jacobian $J(x_k)$ is nonsingular for every k, then x_{k+1} is defined as:

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

If J is singular, then The Newton method is undefined, and some transformations of the method must be considered in order to skip over this situation. In the following we assume that $J(x^*)$ is nonsingular, that means that the continuity of J will ensure that $J(x_k)$ is nonsingular for any x_k enough close to x^*.

We see that for solving $F(x)=0$, we need to consider an initial point x_0 from $dom(f_i)$, for all $i=1,\dots,n$, and to iterate (1) until a criteria for stopping the iterations is satisfied. Let us consider an example which illustrate the method.

Example 1. Let F be defined as:

$$F(x)=\begin{bmatrix} x_1^2+x_2^2-2 \\ \exp(x_1-1)-x_2 \end{bmatrix}.$$

Applying the Newton method with the initial point $x_0=\begin{bmatrix}1.4 & 0.8\end{bmatrix}$, the following results are obtained, as shown in tables 1 and 2.

Table 1. Solution of the system by Newton method

k	x_k^1	x_k^2
0	.1400000000000E+01	.8000000000000E+00
1	.1070918449595E+01	.1000892713209E+01
2	.1002402519583E+01	.9999422607128E+00
3	.1000002882268E+01	.9999999985242E+00
4	.1000000000004E+01	.1000000000000E+01
5	.1000000000000E+01	.1000000000000E+01

Table 2. Evolution of $\|x^*-x_k\|$ and $\|F(x_k)\|$

k	$\|x^*-x_k\|$	$\|F(x_k)\|$
0	.447213595499958E+00	.9157627488965E+00
1	.709240680574689E-01	.1654342180140E+00
2	.240321330115156E-02	.5302195270017E-02
3	.288226844402807E-05	.6442977199172E-05
4	.415378842433256E-11	.9288153280959E-11
5	.000000000000000E+00	.0000000000000E+00

We see that six iterations are sufficient to get the solution. In Andrei [2004a, p. 123] it shown that the complexity of Newton method is dependent on a number of constants and the initial point, being greater than or equal with 6. From table 2 we see that

$$\frac{\|x^*-x_2\|}{\|x^*-x_1\|^2}=0.477,\qquad \frac{\|x^*-x_3\|}{\|x^*-x_2\|^2}=0.499,\qquad \frac{\|x^*-x_4\|}{\|x^*-x_3\|^2}=0.5$$

suggesting that

$$\frac{\|x^*-x_{k+1}\|}{\|x^*-x_k\|^2}$$

is assymptotically constant as $k\to\infty$. Since the errors reduce very quickly to zero, this cannot be verified numerically. In turn the following theorem, which is the main result on the Newton method showing its quadratic convergence, can be proved.

Theorem 1. *Let* $F: R^n \to R^n$ *be a continuously differentiable vector–valued function and* $F(x^*) = 0$. *If:*

i) *The Jacobian* $J(x^*)$ *of* F *at* x^* *is nonsingular.*

ii) *The Jacobian* J *is Lipschitz continuous on a neighborhood of* x^*,

then, for all x_0 *sufficiently close to* x^*, *the Newton method defined by (1) generates a sequence* $x_0, x_1, \ldots$ *that converges quadratically to* x^*.

Before proving this theorem some comments are in order.

Firstly we characterize the Lipschitz condition of J. We say that $J: R^n \to R^{n \times n}$ is Lipschitz continuous on $S \subset R^n$ if there exists a positive constant L, known as *Lipschitz constant*, such that

$$\|J(x) - J(y)\| \leq L\|x - y\| \text{ for all } x, y \in S.$$

Roughly speaking, the difference $J(x) - J(y)$ is proportional in size with $x - y$. We notice that the Lipschitz condition on J is a stronger condition then that of continuity of J, but weaker then the condition F to be twice continuously differentiable.

The quadratic convergence is important in two respects. On one hand, it guarantees that if the Newton method can be initialized in a point close to the solution, then it is rapidly convergent to the solution. On the other hand, the Newton method has a very convenable *stopping criterion* as $\|x_k - x_{k-1}\| < \varepsilon$, where ε is a prespecified error tolerance.

To prove the quadratic convergence of the Newton method it is necessary to present some results from calculus, as follows.

Proposition 1. *Let* $F: R^n \to R^m$ *be continuous differentiable and* $a, b \in R^n$. *Then*

$$F(b) = F(a) + \int_0^1 J\big(a + t(b-a))(b-a)\big)\mathrm{d}t,$$

where J *is the Jacobian of* F.

Proposition 2. *Let* $F: R \to R^n$ *be integrable over the interval* $[a, b]$, *then*

$$\left\|\int_a^b F(t)dt\right\| \leq \int_a^b \|F(t)\|\mathrm{d}t.$$

Proposition 3. *Let* $J: R^m \to R^{n \times n}$ *be a continuous matrix-valued function. If* $J(x^*)$ *is nonsingular, then there exists* $\delta > 0$ *such that for all* $x \in R^m$ *with* $\|x - x^*\| < \delta$, $J(x)$ *is nonsingular and*

$$\|J(x)^{-1}\| < 2\|J(x^*)^{-1}\|.$$

The proposition 3 shows that the set of nonsingular matrices is an open set. The second part of the proposition follows from the fact that, if the aplication $x \to J(x)$ is continuous, then $x \to J(x)^{-1}$ is also continuous, as soon as it is defined. With these preparatives we can prove the theorem 1.

Prof of theorem 1. Let us assume that the estimation x_k is enough close to the solution x^* and $J(x_k)$ is nonsingular, and consider the iteration

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

Now, substract x^* from both sides, to obtain

$$x_{k+1} - x^* = x_k - x^* - J(x_k)^{-1} F(x_k).$$

But, $F(x^*) = 0,$ so we can write

$$x_{k+1} - x^* = x_k - x^* - J(x_k)^{-1}\left(F(x_k) - F(x^*)\right).$$

Now, using the proposition 1 we will estimate the difference $F(x_k) - F(x^*)$ as follows:

$$\begin{aligned} F(x_k) - F(x^*) &= \int_0^1 J(x^* + t(x_k - x^*))(x_k - x^*)\mathrm{d}t \\ &= \int_0^1 J(x^*)(x_k - x^*)\mathrm{d}t + \int_0^1 \left(J(x^* + t(x_k - x^*)) - J(x^*)\right)(x_k - x^*)\mathrm{d}t \\ &= J(x^*)(x_k - x^*) + \int_0^1 \left(J(x^* + t(x_k - x^*)) - J(x^*)\right)(x_k - x^*)\mathrm{d}t. \end{aligned}$$

Therefore,

$$\begin{aligned} &\left\| F(x_k) - F(x^*) - J(x^*)(x_k - x^*) \right\| \\ &= \left\| \int_0^1 \left(J(x^* + t(x_k - x^*)) - J(x^*)\right)(x_k - x^*)\mathrm{d}t \right\| \\ &\le \int_0^1 \left\| \left(J(x^* + t(x_k - x^*)) - J(x^*)\right)(x_k - x^*) \right\| \mathrm{d}t \\ &\le \int_0^1 \left\| J(x^* + t(x_k - x^*)) - J(x^*) \right\| \left\| x_k - x^* \right\| \mathrm{d}t \\ &\le \int_0^1 Lt \left\| x_k - x^* \right\|^2 \mathrm{d}t = \frac{L}{2} \left\| x_k - x^* \right\|^2. \end{aligned}$$

Now we have:

$$\begin{aligned} x_{k+1} - x^* &= x_k - x^* - J(x_k)^{-1}\left(F(x_k) - F(x^*)\right) \\ &= x_k - x^* - J(x_k)^{-1}\left(J(x^*)(x_k - x^*) + F(x_k) - F(x^*) - J(x^*)(x_k - x^*)\right) \\ &= \left(I - J(x_k)^{-1} J(x^*)\right)(x_k - x^*) - J(x_k)^{-1}\left(F(x_k) - F(x^*) - J(x^*)(x_k - x^*)\right). \end{aligned}$$

Therefore,

$$\begin{aligned} \left\| x_{k+1} - x^* \right\| &\le \left\| \left(I - J(x_k)^{-1} J(x^*)\right)(x_k - x^*) \right\| + \\ &\quad \left\| J(x_k)^{-1}\left(F(x_k) - F(x^*) - J(x^*)(x_k - x^*)\right) \right\| \\ &\le \left\| I - J(x_k)^{-1} J(x^*) \right\| \left\| x_k - x^* \right\| + \\ &\quad \left\| J(x_k)^{-1} \right\| \left\| F(x_k) - F(x^*) - J(x^*)(x_k - x^*) \right\| \end{aligned}$$

$$\le \left\|I - J(x_k)^{-1} J(x^*)\right\| \left\|x_k - x^*\right\| + \frac{L}{2} \left\|J(x_k)^{-1}\right\| \left\|x_k - x^*\right\|^2 .$$

Using again the Lipschitz continuity the following estimation can be obtained:

$$\left\|I - J(x_k)^{-1} J(x^*)\right\| = \left\|J(x_k)^{-1}\left(J(x_k) - J(x^*)\right)\right\|$$
$$\le \left\|J(x_k)^{-1}\right\| \left\|J(x_k) - J(x^*)\right\| \le L \left\|J(x_k)^{-1}\right\| \left\|x_k - x^*\right\|.$$

It follows that

$$\left\|x_{k+1} - x^*\right\| \le \frac{3L}{2} \left\|J(x_k)^{-1}\right\| \left\|x_k - x^*\right\|^2 .$$

Now using the proposition 3, for all x_k enough close to x^* we see that $\left\|J(x_k)^{-1}\right\| \le 2M$, where $M = \left\|J(x^*)^{-1}\right\|$. It follows that, for x_k sufficiently close to x^*,

$$\left\|x_{k+1} - x^*\right\| \le 3LM \left\|x_k - x^*\right\|^2 ,$$

proving the quadratic convergence of the Newton method. ■

We see that if $\left\|x_k - x^*\right\| \le \dfrac{1}{6LM}$, then $\left\|x_{k+1} - x^*\right\| \le \dfrac{1}{2}\left\|x_k - x^*\right\|$, showing the progress of the method near solution.

Example 1 (continued) Considering different initial points in the area $[1,3]\times[1,3]$ the number of iterations needed by the Newton method is shown in figure 1. We see that near solution it takes a relatively small number of iterations.

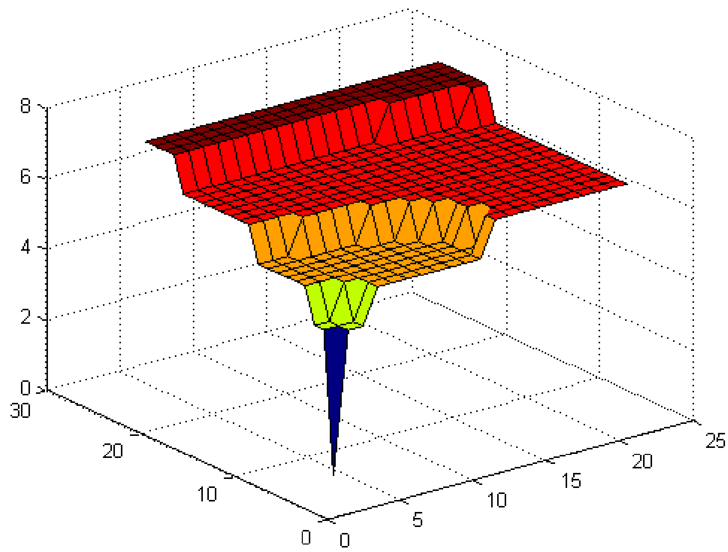


Fig. 1. The number of iterations for solving the system in example 1 when the Newton method is initialized in different points from area [1,3]x[1,3].

The Newton method has excellent *local* convergence properties. But, far away from solution there is no guarantee that its quadratic convergence property is conserved. The algorithm may not even be defined in the sense that there is an estimate x_k where the Jacobian $J(x_k)$ is singular. In these circumstances it is necessary to modify the Newton's method in order to obtain its *global* convergence, that is the method is convergent when it is initialized in a point far away from solution.

References

Andrei, N., (1999a) *Programarea Matematică Avansată - Teorie, Metode computaționale, Aplicații.* Editura Tehnică, Bucureşti, 1999.

Andrei, N., (1999b) *Programarea Matematică. Metode de Punct Interior.* Editura Tehnică, Bucureşti, 1999.

Andrei, N., (2004a) *Convergența Algoritmilor de Optimizare.* Editura Tehnică - Bucureşti, 2004.

Andrei, N., (2004b) *Teorie versus Empirism în Analiza Algoritmilor de Optimizare.* Editura Tehnică - Bucureşti, 2004,

Bazaraa, S.M., Sherali, H.D., Shetty, C.M., (1993) *Nonlinear Programming. Theory and Algorithms.* John Wiley & Sons, Inc., New York, 1993.

Boyd, S., Vandenberghe, L., (2003) *Convex optimization.* Texbook Stanford University, 2003. [http://www.stanford.edu/~boyd/cvxbook.html]

Dennis, J.E., Schnabel, R.B., (1983) *Numerical methods for unconstrained optimization and nonlinear equations.* Prentice Hall, Englewoods Cliffs, N.J., 1983.

Gill, Ph.E., Murray, W., Wright, M.H., (1981) *Practical Optimization.* Academic Press, London - New York, 1981.

Luenberger, D.G., (1984) *Introduction to Linear and Nonlinear Programming.* Addison-Wesley, Menlo-Park, 1984.

Nesterov, Y., Nemirovskii, A., (1994) *Interior-point polynomial algorithms in convex programming.* vol. 13 of Studies in Applied Mathematics, SIAM Publications, Philadelphia, PA, 1994.

Nocedal, J., Wright, S.J., (1999) *Numerical optimization.* Springer Series in Operations Research, Springer Verlag, Berlin, 1999.

December 6, 2004