A Short History of Newton's Method

Peter Deuflhard

2010 Mathematics Subject Classification: 01A45, 65-03, 65H05, 65H10, 65J15, 65K99
Keywords and Phrases: History of Newton's method, Simpson, Raphson, Kantorovich, Mysoskikh, geometric approach, algebraic approach

If an algorithm converges unreasonably fast,
it must be Newton's method.

John Dennis (private communication)

It is an old dream in the design of optimization algorithms, to mimic Newton's method due to its enticing quadratic convergence. But: Is Newton's method really Newton's method?

Linear perturbation approach

Assume that we have to solve a *scalar equation* in one variable, say

$$f(x) = 0$$

with an appropriate guess x^0 of the unknown solution x^* at hand. Upon introducing the *perturbation*

$$\Delta x = x^* - x^0,$$

Taylor's expansion dropping terms of order higher than linear in the perturbation, yields the approximate equation

$$f'(x^0)\Delta x \doteq -f(x^0) \ ,$$

which may lead to an iterative equation of the kind

$$x^{k+1} = x^k - \frac{f(x^k)}{f'(x^k)}, \quad k = 0, 1, \dots$$

assuming the denominator to be non-zero. This is usually named *Newton's method*. The perturbation theory carries over to rather general nonlinear *operator equations*, say

$$F(x) = 0, \quad x \in D \subset X, \quad F : D \to Y,$$

where X, Y are Banach spaces. The corresponding Newton iteration is then typically written in the form

$$F'(x^k)\Delta x^k = -F(x^k), \quad x^{k+1} = x^k + \Delta x^k, \quad k = 0, 1, \ldots$$

For more details and extensions see, e.g., the textbook [1] and references therein.

Convergence

From the linear perturbation approach, local quadratic convergence will be clearly expected for the scalar case. For the general case of operator equations $F(x) = 0$, the convergence of the generalized Newton scheme has first been proven by two Russian mathematicians: In 1939, L. Kantorovich [5] was merely able to show local *linear* convergence, which he improved in 1948/49 to local *quadratic* convergence, see [6, 7]. Also in 1949, I. Mysovskikh [9] gave a much simpler independent proof of local *quadratic* convergence under slightly different theoretical assumptions, which are exploited in modern Newton algorithms, see again [1]. Among later convergence theorems the ones due to J. Ortega and W.C. Rheinboldt [11] and the affine invariant theorems given in [2, 3] may be worth mentioning.

Geometric approach

The standard approach to Newton's method in elementary textbooks is given in Figure 1. It starts from the fact that any root of f may be interpreted as the intersection of the *graph* of $f(x)$ with the real axis. In Newton's method, this graph is replaced by its *tangent* in x_0; the first iterate x_1 is then defined as the intersection of the tangent with the real axis. Upon repeating this geometric process, a close-by solution point x^* can be constructed to any desired accuracy. On the basis of this geometric approach, this iteration will converge *globally* for *convex* (or concave) f.

At first glance, this geometric derivation seems to be restricted to the scalar case, since the graph of $f(x)$ is a typically one-dimensional concept. A careful examination of the subject in more than one dimension, however, naturally leads to a topological path called *Newton path*, which can be used for the construction of modern adaptive Newton algorithms, see again [1].

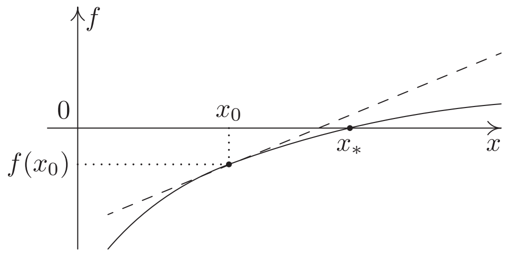

Figure 1: Newton's method for a scalar equation

Historical road

The long way of Newton's method to become Newton's method has been well studied, see, e.g., N. Kollerstrom [8] or T.J. Ypma [13]. According to these articles, the following facts seem to be agreed upon among the experts:

- In 1600, Francois Vieta (1540–1603) had designed a *perturbation* technique for the solution of the scalar polynomial equations, which supplied one decimal place of the unknown solution per step via the explicit calculation of successive polynomials of the successive perturbations. In modern terms, the method converged *linearly.* It seems that this method had also been published in 1427 by the Persian astronomer and mathematician al-Kāshī (1380–1429) in his *The Key to Arithmetic* based on much earlier work by al-Biruni (973–1048); it is not clear to which extent this work was known in Europe. Around 1647, Vieta's method was simplified by the English mathematician Oughtred (1574–1660).

- In 1664, Isaac Newton (1643–1727) got to know Vieta's method. Up to 1669 he had improved it by *linearizing* the successively arising polynomials. As an example, he discussed the numerical solution of the cubic polynomial
$$f(x) := x^3 - 2x - 5 = 0\ .$$
Newton first noted that the integer part of the root is 2 setting $x_0 = 2$. Next, by means of $x = 2 + p$, he obtained the polynomial equation
$$p^3 + 6p^2 + 10p - 1 = 0\ .$$
He neglected terms higher than first order setting $p \approx 0.1$. Next, he inserted $p = 0.1 + q$ and constructed the polynomial equation
$$q^3 + 6.3q^2 + 11.23q + 0.061 = 0\ .$$
Again neglecting higher order terms he found $q \approx -0.0054$. Continuation of the process one further step led him to $r \approx 0.00004853$ and therefore to the third iterate
$$x_3 = x_0 + p + q + r = 2.09455147\ .$$

Note that the relations $10p - 1 = 0$ and $11.23q + 0.061 = 0$ given above correspond precisely to

$$p = x_1 - x_0 = -f(x_0)/f'(x_0)$$

and to

$$q = x_2 - x_1 = -f(x_1)/f'(x_1) \; .$$

As the example shows, he had also observed that by keeping all decimal places of the corrections, the number of accurate places would *double* per each step – i.e., *quadratic convergence*. In 1687 (Philosophiae Naturalis Principia Mathematica), the first nonpolynomial equation showed up: it is the well-known equation from astronomy

$$x - e\sin(x) = M$$

between the *mean anomaly* M and the *eccentric anomaly* x. Here Newton used his already developed polynomial techniques via the series expansion of *sin* and *cos*. However, no hint on the derivative concept is incorporated!

- In 1690, Joseph Raphson (1648–1715) managed to avoid the tedious computation of the successive polynomials, playing the computational scheme back to the original polynomial; in this now fully *iterative* scheme, he also kept all decimal places of the corrections. He had the feeling that his method differed from Newton's method at least by its derivation.
- In 1740, Thomas Simpson (1710–1761) actually introduced derivatives ('fluxiones') in his book 'Essays on Several Curious and Useful Subjects in Speculative and Mix'd Mathematicks [No typo!], Illustrated by a Variety of Examples'. He wrote down the true *iteration* for one (nonpolynomial) equation and for a system of two equations in two unknowns thus making the correct extension to *systems* for the first time. His notation is already quite close to our present one (which seems to date back to J. Fourier).

The interested reader may find more historical details in the book by H. H. Goldstine [4] or even try to read the original work by Newton in Latin [10]; however, even with good knowledge of Latin, this treatise is not readable to modern mathematicians due to the ancient notation. That is why D.T. Whiteside [12] edited a modernized English translation.

What is Newton's method?

Under the aspect of historical truth, the following would come out:

- For scalar equations, one might speak of the Newton–Raphson method.
- For more general equations, the name Newton–Simpson method would be more appropriate.

Under the convergence aspect, one might be tempted to define Newton's method via its quadratic convergence. However, this only covers the pure Newton method. There are plenty of variants like the simplified Newton method, Newton-like methods, quasi-Newton methods, inexact Newton methods, global Newton methods etc. Only very few of them exhibit quadratic convergence. In fact, even the Newton–Raphson algorithm for scalar equations as realized in hardware within modern calculators converges only linearly due to finite precision, which means they asymptotically implement some Vieta algorithm. Hence, one will resort to the fact that Newton methods simply exploit derivative information in one way or the other.

Acknowledgement

The author wishes to thank E. Knobloch for having pointed him to several interesting historical sources.

References

[1] P. Deuflhard. *Newton Methods for Nonlinear Problems. Affine Invariance and Adaptive Algorithms*, volume 35 of *Computational Mathematics*. Springer International, 2004.

[2] P. Deuflhard and G. Heindl. Affine Invariant Convergence theorems for Newton's Method and Extensions to related Methods. *SIAM J. Numer. Anal.*, 16:1–10, 1979.

[3] P. Deuflhard and F.A. Potra. Asymptotic Mesh Independence of Newton-Galerkin Methods via a Refined Mysovskii Theorem. *SIAM J. Numer. Anal.*, 29:1395–1412, 1992.

[4] H. H. Goldstine. *A history of Numerical Analysis from the 16th through the 19th Century*. Spri9nger, 1977.

[5] L. Kantorovich. The method of successive approximations for functional equations. *Acta Math.*, 71:63–97, 1939.

[6] L. Kantorovich. On Newton's Method for Functional Equations. (Russian). *Dokl. Akad. Nauk SSSR*, 59:1237–1249, 1948.

[7] L. Kantorovich. On Newton's Method. (Russian). *Trudy Mat. Inst. Steklov*, 28:104–144, 1949.

[8] N. Kollerstrom. Thomas Simpson and 'Newton's Method of Approximation': an enduring myth. *British Journal for History of Science*, 25:347–354, 1992.

[9] I. Mysovskikh. On convergence of Newton's method. (Russian). *Trudy Mat. Inst. Steklov*, 28:145–147, 1949.

[10] I. Newton. *Philosophiae naturalis principia mathematica.* Colonia Allobrogum: sumptibus Cl. et Ant. Philibert, 1760.

[11] J.M. Ortega and W.C. Rheinboldt. *Iterative Solution of Nonlinear Equations in Several Variables.* Classics in Appl. Math. SIAM Publications, Philadelphia, 2nd edition, 2000.

[12] D.T. Whiteside. The Mathematical Papers of Isaac Newton (7 volumes), 1967–1976.

[13] T.J. Ypma. Historical Development of the Newton-Raphson Method. *SIAM Rev.*, 37:531–551, 1995.

Peter Deuflhard
Konrad-Zuse-Zentrum
für Informationstechnik
Berlin (ZIB)
Takustraße 7
14195 Berlin
Germany
deuflhard@zib.de