

Mathematical Association of America

The Continuous Newton's Method, Inverse Functions, and Nash-Moser

Author(s): J. W. Neuberger

Source: *The American Mathematical Monthly*, Vol. 114, No. 5 (May, 2007), pp. 432-437

Published by: Mathematical Association of America

Stable URL: <http://www.jstor.org/stable/27642222>

Accessed: 25-10-2015 17:39 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at <http://www.jstor.org/page/info/about/policies/terms.jsp>

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.



Mathematical Association of America is collaborating with JSTOR to digitize, preserve and extend access to *The American Mathematical Monthly*.

<http://www.jstor.org>

NOTES

Edited by Ed Scheinerman

The Continuous Newton's Method, Inverse Functions, and Nash-Moser

J. W. Neuberger

1. INTRODUCTION. The conventional Newton's method for finding a zero of a function $F : R^n \rightarrow R^n$, assuming that $(F'(y))^{-1}$ exists for at least some y in R^n , is the familiar iteration: pick z_0 in R^n and define

$$z_{k+1} = z_k - (F'(z_k))^{-1} F(z_k) \quad (k = 0, 1, 2, \dots),$$

hoping that z_1, z_2, \dots converges to a zero of F . What can stop this process from finding a zero of F ? For one thing, there might not *be* a zero of F . For another, the process might terminate for some integer k in the event that $F'(z_k)$ does not have an inverse.

A domain of attraction corresponding to a given root of F consists of the set of all starting values z_0 that lead, through convergence of z_1, z_2, \dots , to this root. Newton's method can lead to chaotic domains of attraction, even for simple choices of F (see [8]). This can lead to striking pictures but constitutes a nightmare for the numerical analyst. This fact, if nothing else, leads one to the damped Newton's method, which consists of

$$z_{k+1} = z_k - \delta_k (F'(z_k))^{-1} F(z_k) \quad (k = 0, 1, 2, \dots),$$

where $\delta_1, \delta_2, \dots$ are chosen from $(0, 1)$ in an attempt to gain more reasonable domains of attraction. The continuous Newton's method is, in a sense, a limiting case of a sequence of damped Newton's methods: Pick $T > 0$ and for each positive integer m consider damped Newton's method running m steps with $\delta_k = T/m$, ($k = 1, \dots, m$). Denote the resulting z_{m+1} by x_m . If x_1, x_2, \dots converges, the limit of this sequence is said to be the result of the continuous Newton's method at the number T . The preceding is just a rough way of saying the following: consider finding a function $z : [0, \infty) \rightarrow R^n$ so that

$$z(0) = x \in R^n, \quad z'(t) = -(F'(z(t)))^{-1} F(z(t)) \quad (t \geq 0) \quad (1)$$

in the hope that $u = \lim_{t \rightarrow \infty} z(t)$ exists and $F(u) = 0$.

The continuous Newton's method seems much less familiar than the conventional Newton's method. An application to finding roots of complex polynomials is found in [8] and the references contained therein. In [8] it is shown that for the continuous Newton's method there are no fractal domains of attraction. A starting place from which to gain an historical overview of the continuous Newton's method is [1]. This reference contains a survey of path-following methods (to which the continuous Newton's method belongs) and also contains a very extensive set of references. The papers [10], [11] and the references contained therein have detailed discussions of various aspects

of the continuous Newton's method. The continuous Newton's method has a close connection with inverse and implicit function results; the present note is a contribution to this connection. See also [5] in this regard.

We pause to show an important feature of the continuous Newton's method, a feature that has no counterpart in the ordinary or damped Newton's method. This feature is at the heart of what led to the results in the rest of this note. Suppose that z satisfies (1). Then

$$F'(z(t))z'(t) = -F(z(t)) \quad (t \geq 0),$$

which can be rewritten as

$$(F \circ z)'(t) = -F(z(t)),$$

whence

$$F(z(t)) = \exp(-t)F(z(0)).$$

So, we see that the "residual" $F(z(t))$ (what we want to drive to zero) doesn't change direction, only magnitude, as t increases. Some thought indicates that this fact leads to a philosophy of mountain climbing when the function F represents the gradient of a function ϕ that specifies the height of a mountain. The philosophy says to proceed up a mountain in such a way that the *direction* of the gradient doesn't change as you climb (think of Stone Mountain in Georgia, Enchanted Rock in Texas, Ayers Rock in Australia).

In this note some results arising from a study of the continuous Newton's method are given. Included are some zero finding results, some inverse function results and a Nash-Moser type result that avoids some of the dreaded "derivative loss" associated with the conventional Newton's method applied to problems in partial differential equations.

2. TWO ZERO FINDING RESULTS. Suppose that each of m and n is a positive integer. If $s > 0$ and y is in R^n , then $b_s(y)$ and $B_s(y)$ denote the open and closed balls (in R^n), respectively, that have center y and radius s . The following theorem gives a condition on an initial estimate that is sufficient for convergence:

Theorem 1. *Suppose that $r > 0$, that $x \in R^n$, and that F is a continuous function from $B_r(x)$ to R^m with the property that for each y in $b_r(x)$ there is an h in $B_r(0)$ such that*

$$\lim_{t \rightarrow 0^+} \frac{1}{t} (F(y + th) - F(y)) = -F(x).$$

Then there exists u in $B_r(x)$ such that $F(u) = 0$.

Proof. Suppose that $\epsilon > 0$. Define

$$S = \{s \in [0, 1] : \exists y \in B_{rs}(x) \text{ such that } \|F(y) - (1 - s)F(x)\| \leq \epsilon s\}.$$

Note that S is closed, since $[0, 1]$ and $B_r(x)$ are compact and F is continuous. Denote $\sup S$ by λ and suppose that $\lambda < 1$. Pick y in $B_{\lambda r}(x)$ for which

$$\|F(y) - (1 - \lambda)F(x)\| \leq \epsilon \lambda,$$

and then choose h in $B_r(0)$ and δ in $(0, 1 - \lambda]$ so that

$$\left\| \frac{1}{\delta}(F(y + \delta h) - F(y)) + F(x) \right\| \leq \epsilon,$$

that is,

$$\|(F(y + \delta h) - F(y)) + \delta F(x)\| \leq \epsilon \delta.$$

Then $\|y + \delta h\| \leq (\lambda + \delta)r$ and

$$\begin{aligned} & \|F(y + \delta h) - (1 - \delta - \lambda)F(x)\| \\ & \leq \|F(y + \delta h) - F(y) + \delta F(x)\| + \|F(y) - (1 - \lambda)F(x)\| \leq \epsilon(\delta + \lambda), \end{aligned}$$

so $\delta + \lambda$ belongs to S , a contradiction. Therefore $\lambda = 1$.

Hence, for each $\epsilon > 0$ there is u_ϵ in $B_r(x)$ so that $\|F(u_\epsilon)\| \leq \epsilon$. By the continuity of F and the compactness of $B_r(x)$, there exists u in $B_r(x)$ such that $F(u) = 0$. ■

The following shows that things simplify if the inverse of the derivative exists for all y in $B_r(x)$:

Theorem 2. *Suppose that $r > 0$, that $x \in R^n$, and that F is a C^2 -function from $B_r(x)$ to R^n with the property that if $y \in B_r(x)$, then $(F'(y))^{-1}$ exists and*

$$\|(F'(y))^{-1}F(x)\| \leq r. \tag{2}$$

Then there exists u in $B_r(x)$ such that $F(u) = 0$. Moreover, such an element is given by $u = z(1)$, where z satisfies

$$z(0) = x, \quad z'(t) = -(F'(z(t)))^{-1}F(x) \quad (t \in [0, 1]). \tag{3}$$

Proof. Note that (3) has a unique solution z on all of $[0, 1]$ since, due to (2), a solution can't "escape" from $B_r(x)$ in less than "time one" (that F is of class C^2 makes local existence and uniqueness hold). Thus

$$(F'(z(t)))z'(t) = -F(x) \quad (t \in [0, 1]),$$

that is,

$$(F \circ z)'(t) = -F(x),$$

hence

$$F(z(t)) - F(z(0)) = -tF(x).$$

We conclude that

$$F(z(t)) = (1 - t)F(x) \quad (t \in [0, 1]).$$

In particular, $F(z(1)) = 0$. ■

If the function F in Theorem 2 were only of class C^1 the conclusion would still hold, even though there might not be uniqueness of the solution to (3) (which isn't needed anyway).

3. THE CONTINUOUS NEWTON'S METHOD. How do Theorems 1 and 2 relate to the continuous Newton's method? Suppose that H is a Banach space and that F is a C^1 -function from H to itself for which $(F'(x))^{-1}$ exists and has domain all of H for each x in H . If, as in the introduction, $z : [0, \infty) \rightarrow H$ and

$$z(0) = x \in H, \quad z'(t) = -(F'(z(t)))^{-1} F(z(t)) \quad (t \geq 0), \quad (4)$$

then

$$F(z(t)) = \exp(-t)F(z(0)) \quad (5)$$

and from (4) we obtain

$$z(0) = x, \quad z'(t) = -\exp(-t)(F'(z(t)))^{-1} F(x). \quad (6)$$

Now make a change of scale from $[0, \infty)$ to $[0, 1)$ by defining $w : [0, 1) \rightarrow H$ as follows:

$$w(t) = z\left(\ln\left(\frac{1}{1-t}\right)\right)$$

Then

$$w'(t) = -(F'(w(t)))^{-1} F(x). \quad (7)$$

Note that $\lim_{t \rightarrow 1^-} w(t)$ exists if and only if $\lim_{t \rightarrow \infty} z(t)$ exists. The former limit exists if and only if w in (7) can be solved on all of $[0, 1]$.

This is the key to Theorems 1 and 2 as well as to Theorem 3 in the next section. Theorems 1 and 3 can be regarded as applying an Euler-type method applied to (7).

4. ZERO FINDING IN INFINITE DIMENSIONAL SPACES. Theorem 1 can easily be generalized to cover a wide class of partial differential equations. The celebrated papers of Nash [7] and of Moser [6] are landmarks in the study of partial differential equations. These two works, together with [3], [4], and [5], use a version of the discrete Newton's method, and all involve long arguments running through quite a number of pages of difficult calculations. An early version of the continuous Newton's method for PDEs is found in [12], a later one in [2]. The present note is largely based on [9].

Details of applications to concrete partial differential equations won't be attempted here. Instead we state a version of Theorem 1 that applies to such equations. For the next theorem suppose that each of H , J , and K is a Banach space and that H is compactly embedded in J (meaning: the points of H form a dense linear subspace of J and if y_1, y_2, \dots is a sequence in H such that, for some $M > 0$, $\|y_n\|_H \leq M$ for all n , then y_1, y_2, \dots has a subsequence convergent in J to an element y of H satisfying $\|y\|_H \leq M$). Suppose also that $F : H \rightarrow K$ is a function that is continuous with respect to the topologies of J and K . When $s > 0$ and u is in H , $b_s(u)$ and $B_s(u)$ signify (only for Theorems 3 and 5) the open and closed balls in H , respectively, with center u and radius s .

Theorem 3. *Suppose that $x \in H$, that $r > 0$, and that for each y in $b_r(x)$ there is an h in $B_r(0)$ such that*

$$\lim_{t \rightarrow 0^+} \frac{1}{t} (F(y + th) - F(y)) = -F(x).$$

Then there is u in $B_r(x)$ such that $F(u) = 0$.

Proof. This parallels the proof of Theorem 1 except that in the two instances where compactness is mentioned we now appeal to the fact that any bounded sequence in H has a subsequence that converges in J to a member of H . ■

5. SOME INVERSE FUNCTION RESULTS.

Theorem 4. Suppose that $r > 0$, that G is a continuous function from R^n to R^m with $G(0) = 0$, and that g is a point of R^m . Suppose also that for each y in $b_r(0)$ there is an h in $B_r(0)$ such that

$$\lim_{t \rightarrow 0^+} \frac{1}{t}(G(y + th) - G(y)) = g. \quad (8)$$

Then there exists u in $B_r(0)$ such that $G(u) = g$.

Proof. Define $F : B_r(0) \rightarrow R^m$ by $F(y) = G(y) - g$. Then if y is in $b_r(0)$, there is an h in $B_r(0)$ for which (8) holds. Thus

$$\lim_{t \rightarrow 0^+} \frac{1}{t}(F(y + th) - F(y)) = g.$$

According to Theorem 1 there exists u in $B_r(0)$ such that $F(u) = 0$, whence $G(u) = g$. ■

If $m = n$, G is a C^1 function, and $(G'(y))^{-1}$ exists for each y in $B_r(0)$, then the hypothesis of Theorem 4 is satisfied by every g in R^n for which $\|(G'(u))^{-1}g\| \leq r$. From this we see that there is a ball in R^n centered at the origin that is filled with elements of the range of F .

A generalization to infinite dimensions of Theorem 2 follows. It is a Nash-Moser type theorem of use in partial differential equations (see [9]).

Theorem 5. Suppose H, J, K, r, b_r and B_r are as in Theorem 3 and that $G : B_r(0) \rightarrow K$ is continuous as a function on J . Suppose also that g belongs to K and that for each y in $b_r(0)$ there is an h in $B_r(0)$ such that

$$\lim_{t \rightarrow 0^+} \frac{1}{t}(G(y + th) - G(y)) = g.$$

Then there exists u in $B_r(0)$ such that $G(u) = g$.

Proof. This follows immediately from Theorem 3 by applying it to $F : B_r(0) \rightarrow K$ defined by $F(y) = G(y) - g$ for each y in $B_r(0)$. ■

6. TWO PROBLEMS. Here are two problems for the reader. The ideas of this note apply to these problems, but a reader might want to think of other ways to establish the results.

Problem 1. Suppose c and p are real numbers, that $r > 0$, and that f is a C^1 function with domain $[c - r, c + r]$. If

$$|p - f(c)| \leq r|f'(x)| \quad (x \in [c - r, c + r]),$$

show that there exists x in $[c - r, c + r]$ such that $f(x) = p$.

Preliminary to stating a second problem, we note that the maximal connected domain of existence of a solution u to

$$u' = 1 + u^2, \quad u(0) = 0, \tag{9}$$

is $(-\pi/2, \pi/2)$. This illustrates a common occurrence for nonlinear ordinary differential equations. Thus the assumptions of the following problem, as well as the conclusion, are strong.

Problem 2. Suppose that F is a C^1 function from R^n to R^n with $F(0) = 0$ and that $(F'(x))^{-1}$ exists for each x in R^n . Suppose also that for each v in R^n there is a function $z : [0, 1] \rightarrow R^n$ such that

$$z(0) = 0, \quad z'(t) = (F'(z(t)))^{-1}v.$$

Show that the range of F is all of R^n .

REFERENCES

1. E. L. Algower and K. Georg, Numerical path following, in *Handbook of Numerical Analysis*, vol. 5, P. G. Ciarlet and J. L. Lions, eds., North-Holland, Amsterdam, 1997, pp. 3–207.
2. A. Castro and J. W. Neuberger, An inverse function theorem via continuous Newton's method, *J. Nonlinear Analysis* **47** (2001) 3259–3270.
3. R. Hamilton, The inverse function theorem of Nash and Moser, *Bull. Amer. Math. Soc.* **7** (1982) 65–222.
4. L. Hörmander, On the Nash-Moser implicit function theorem, *Ann. Acad. Sci. Fenn., Ser. A1 Math.* **10** (1985) 255–259.
5. S. Krantz and H. Parks, *The Implicit Function Theorem*, Birkhäuser, Boston 2003.
6. J. Moser, A rapidly convergent iteration method and non-linear differential equations, *Ann. Scuola Normale Sup. Pisa*, **20** (1966) 265–315.
7. J. Nash, The imbedding problem for Riemannian manifolds, *Annals of Math.* **63** (1956) 20–63.
8. J. W. Neuberger, Continuous Newton's method for polynomials, *Math. Intelligencer* **21** (1999) 18–23.
9. ———, A near minimal hypothesis Nash-Moser theorem, *Int. J. Pure Appl. Math.* **4** (2003) 269–280.
10. H. Peitgen, M. Prüfer, and K. Schmitt, Global aspects of the continuous and discrete Newton's method: A case study, *Acta Appl. Math.* **13** (1988) 123–202.
11. D. Saupe, Discrete versus continuous Newton's method: A case study, *Acta Appl. Math.* **13** (1988) 59–80.
12. J. Schwartz, On Nash's implicit function theorem, *Comm. Pure Appl. Math.* **13** (1960) 509–530.

*Department of Mathematics, University of North Texas, Denton, TX 76203-1430
jwn@unt.edu*

A Note on Alternating Permutations

Anthony Mendes

A permutation $\sigma = \sigma_1 \cdots \sigma_n$ of a set of n distinct integers is *alternating* provided $\sigma_1 < \sigma_2 > \sigma_3 < \sigma_4 > \cdots$. Let $E_0 = 1$. For $n = 1, 2, \dots$, define E_n to be the number of alternating permutations in the set of permutations on n distinct integers. This sequence has a fantastic generating function. André showed that

$$\sum_{n=0}^{\infty} E_n \frac{t^n}{n!} = \sec t + \tan t \tag{1}$$

nearly 125 years ago [1], [2].