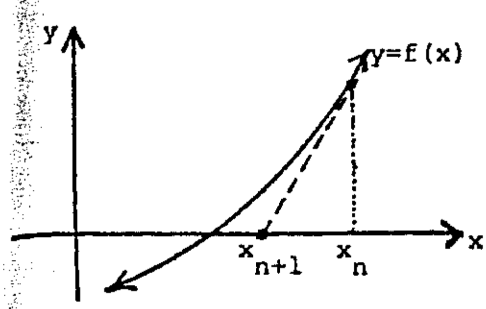
Vol 6 No 2 1981

CAN NEWTON'S METHOD GO WRONG?

by Brailey Sims,
University of New England

Let us recall that Newton's Method is the iterative procedure for approximating a zero (or root) of the function f(x) according to the scheme

$$x_{n+1} = x_n - f(x_n)/f'(x_n).$$



In one of the more popular N.S.W. school text books one finds the following statement:

> The key factor in applying this method is to obtain a good first approximation. If this first approximation is a good one, then it will be certainly true that the next approximation will be a better one, and so on.
>
> **In actual fact, the number of decimal places of accuracy doubles with each successive application of Newton's Method. Thus, if one approximation is good to 1 decimal place, ten the next approximation is good to 2 decimal places, and the following one will be good to 4 decimal places, and so on.**

As it stands this pale shadow of a partial truth is completely wrong. A correct statement, and probably the one intended in the text, is as follows:

THEOREM: Provided f' is continuous at the zero x_0 [that is, $\lim_{x\to x_0} f'(x) = f'(x_0)$] and $f'(x_0) \neq 0$ and provided the initial estimate x_1 is sufficiently near to x_0 (and this may have to be very near to x_0 indeed), the successive iterates produced by Newton's Method will converge to x_0. Indeed, given any positive number $\epsilon < 1$, for x_1 sufficiently near to x_0 (the "nearness" required depending upon ϵ) we have

$$|x_n - x_0| \leq \epsilon^{n-1}|x_1 - x_0|.$$

PROOF: From the continuity of f' at x_0, for any x sufficiently near x_0 we have that $f'(x) \neq 0$ and so we may form $x - f(x)/f'(x)$. Now at such an x

$$\frac{[x-f(x)/f'(x)] - x_0}{x - x_0} = 1 - \frac{f(x)}{f'(x)(x-x_0)}$$

$$= \frac{1}{f'(x)}\left[f'(x) - \frac{f(x)}{x-x_0}\right]$$

$$= \frac{1}{f'(x)}\left[f'(x) - \frac{f(x)-f(x_0)}{x-x_0}\right] \quad (\text{as } f(x_0) = 0)$$

In the limit as $x \to x_0$ this last expression becomes

$$\frac{1}{f'(x_0)}\left[f'(x_0) - \lim_{x\to x_0}\frac{f(x)-f(x_0)}{x-x_0}\right]$$

$$= \frac{1}{f'(x_0)}[f'(x_0) - f'(x_0)] \quad (\text{definition of derivative})$$

$$= 0$$

and so we conclude that

$$\lim_{x\to x_0} \frac{[x-f(x)/f'(x)] - x_0}{x-x_0} = 0$$

In particular then, for all x sufficiently near x_0 we have

$$\left|\frac{[x-f(x)/f'(x)] - x_0}{x-x_0}\right| < \epsilon$$

or $|[x-f(x)/f'(x)] - x_0| < \epsilon|x-x_0|$.

Thus provided x_1 is chosen sufficiently near to x_0 we have

$$|x_2 - x_0| = |[x_1 - f(x_1)/f'(x_1)] - x_0| < \epsilon|x_1 - x_0|$$

while

$$|x_3 - x_0| = |[x_2 - f(x_2)/f'(x_2)] - x_0| < \epsilon|x_2 - x_0| < \epsilon^2|x_1 - x_0|.$$

Similarly,

$$|x_4 - x_0| < \epsilon|x_3 - x_0| < \epsilon^3|x_1 - x_0|.$$

Repeating this argument $(n-1)$ times establishes the result.

Provided f satisfies the conditions of the theorem, choosing $\epsilon = \frac{1}{10}$ and starting with x_1 appropriately near to x_0 we obtain an improvement in accuracy of one decimal place at each iteration.*

* By further restricting f, for example by also requiring f'' to exist and be bounded in a neighbourhood of x_0, it is possible to guarantee a faster rate of convergence leading to the doubling in precision claimed in the text-book cited above. The proof is, however, more complicated and so will not be considered here.

Without assumptions on f such as those of the theorem any of several possibilities can occur:

a) The scheme may converge, but possibly at a slower rate than suggested by the theorem.

b) The successive iterates may diverge away from the zero, no matter how near our initial approximation may be. They may, of course, eventually approach some other zero of f.

c) The successive iterates may "oscillate" about the zero without either approaching or receding.

To illustrate these possibilities with relatively simple functions it suffices to consider power functions of the form $y = x^{a}$ $(a > 0)$.

Each function has as its only zero $x = 0$ and, as a simple calculation will show, for this class of functions Newton's Method becomes

$$x_{n+1} = \left[\frac{a-1}{a}\right] x_n .$$

Thus starting with an initial approximation x_1 we have as the nth iterate

$$x_n = \left[\frac{a-1}{a}\right]^{n-1} x_1 .$$

a) For a a positive integer $\frac{a-1}{a}$ is between 0 and 1 and so the successive iterates will converge to 0. The error after each iteration is $\left|\frac{a-1}{a}\right|$ that for the previous estimate.

By choosing a large we may make $\frac{a-1}{a}$ as near to 1 as we please, thus decreasing the rate at which x_n converges to 0. (Note $\left|\frac{a-1}{a}\right|$ is fixed by the function and cannot be varied as the ϵ in the above theorem.) For example; starting with $x_1=1$ it takes 4 successive iterations to improve the accuracy of the estimate by one decimal place when $a=2$, while for $a=10$ it requires 23 iterations, and this would remain true no matter what the starting value.

b) Choosing $a=\frac{1}{3}$ we have $\frac{a-1}{a} = -2$ and so the sequence of iterates,

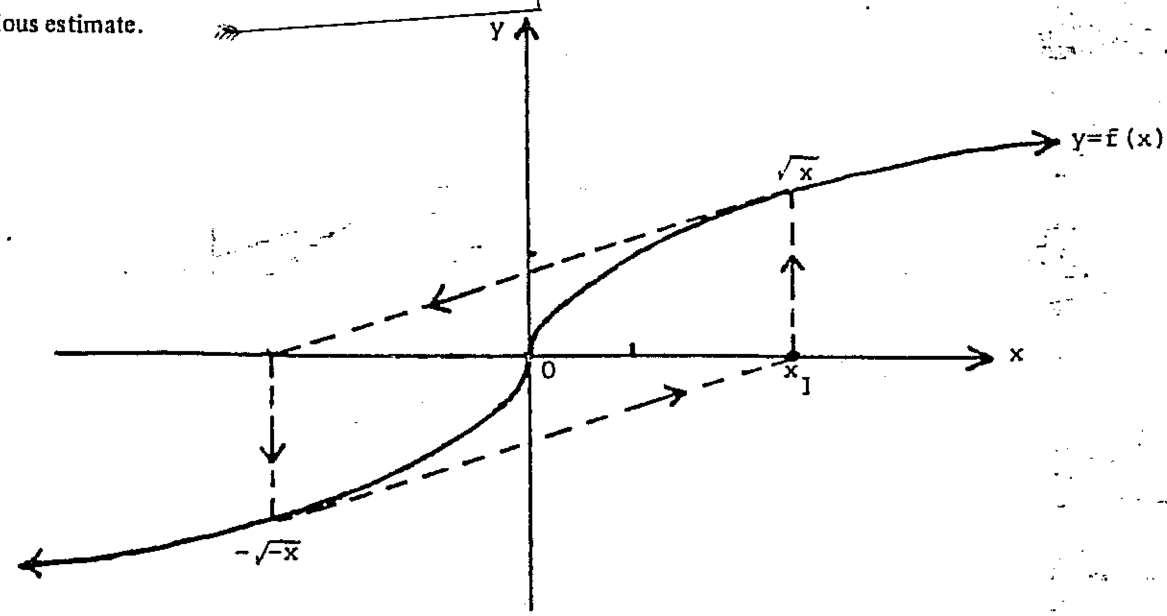
$x_n = (-2)^{n-1} x_1$, fails to converge to 0. For example with $x_1 = 1$ we have

$x_2 = -2, \quad x_3 = 4, \quad x_4 = -8, \quad x_5 = 16, \quad x_6 = -32$

etc.

c) To obtain a function which exhibits oscillatory behaviour we choose $a = \frac{1}{2}$ and extend the function $y = \sqrt{x}$ as an odd function to the whole of the real line according to the formula

$$f(x) = \begin{cases} \sqrt{x} & \text{for } x \geq 0 \\ -\sqrt{-x} & \text{for } x < 0 \end{cases}$$



In this case $\frac{a-1}{a} = -1$ and so the iterates are alternatively $+x_1$ and $-x_1$. For example with $x_1 = 1$ we have successive "approximations" to the zero 0;

1, −1, 1, −1, 1, −1, 1, −1, etc.

Newton's Method works well for a great many functions, however, as the above examples illustrate it is fallible and care must be exercised when applying it. For more complicated functions the method can "misbehave" in more diverse ways than those suggested above; for example, the sequence of iterates may first appear to approach a zero before eventually diverging.

A major problem with Newton's Method is the difficulty in obtaining useful error estimates. The "half–interval method" is an alternative procedure which overcomes this deficiency.

Approximation procedures are frequently called for in the application of Mathematics to real situations (even more so in the modern age of computers). Developing skill in their selection and use is an important part of a mathematical education.