Error Bounds Associated with Newton's Iteration

W. Kahan
Math. Dept.
Univ. of Calif.
Berkeley CA 94720-3840

Suppose we seek a root z of the equation $f(z) = o$ where $f(x)$ is a continuously differentiable vector-valued function of the vector x in a normed space. Newton's iteration replaces a guess x by $x - f'(x)^{-1}f(x)$ repeatedly in the hope of convergence to z. Is there some neighborhood around z within which iteration must converge to z? And if convergence is assured, yet we cannot iterate forever; at some point we shall accept an approximate z_o despite $f(z_o) \ne o$. How close is z_o to z?

Claim 1: If f is twice differentiable about z, and if for some positive δ we find that $\| f'(x)^{-1}f''(y) \| < 2/\delta$ as x and y range independently throughout a ball $\| x - z \| < \delta$, then Newton's iteration will converge to the root z from every starting point x in that ball. Moreover, convergence is at least quadratic.

This claim merely reassures us that, under normal circumstances, Newton's iteration will converge rapidly to a root z from any starting point close enough to z; under normal circumstances $f'(z)^{-1}$ does exist and $f''(x)$ does stay bounded at all x near z, so a positive radius δ has to exist.

Proof of Claim 1: According to Newton's divided difference formula, $f(z) = f(x) + f'(x)(z-x) + \Delta^2 f((z,x,x))(z-x)(z-x)$ where $\Delta^2 f$ is the second divided difference of f and, according to Hermite's integral representation for divided differences,
$\Delta^2 f((a,b,c))$ = The uniformly weighted average of $f''(y)/2$ as y runs over the triangle whose vertices are a, b, c.
If x lies in the ball $\|x-z\| < \delta$, so does the degenerate triangle whose vertices are z, x, x, and therefore so does y, and therefore $\|f'(x)^{-1}f''(y)/2\| < 1/\delta$, and therefore

$$\begin{aligned}\|f'(x)^{-1}\Delta^2 f((z,x,x))\| &= \| \text{Average of } f'(x)^{-1}f''(y)/2 \ldots\| \\ &\le \text{Average of } \|f'(x)^{-1}f''(y)/2\| \ldots \\ &< 1/\delta .\end{aligned}$$

Newton's iteration replaces x by $New(x) = x - f'(x)^{-1}f(x)$; by using Newton's divided difference formula with $f(z) = o$ we find $New(x) - z = f'(x)^{-1}\Delta^2 f((z,x,x))(x-z)(x-z)$, whence follows that $\|New(x) - z\|/\delta \le \|f'(x)^{-1}\Delta^2 f((z,x,x))\| \, \|x-z\|^2/\delta < (\|x-z\|/\delta)^2$. Therefore, starting the iteration $x_{n+1} := New(x_n)$ at any x_o in the ball leads to a sequence $\{x_n\}$ that stays in the ball and

$$\log(\|x_{n+1}-z\|/\delta) < 2 \log(\|x_n-z\|/\delta) < \ldots < 2^{n+1} \log(\|x_o-z\|/\delta) \to -\infty \text{ as } n \to \infty ,$$

as claimed.

Who can wait until $n \to \infty$? Instead we shall accept some x_n as good enough; call it z_o. How close is it to a desired root z?

Claim 2: If we can find some positive $\delta > \| f'(x)^{-1}f(z_0) \|$
~~~~~~~~ throughout the ball $\| x - z_0 \| < \delta$, then at least one root z lies in that ball.

Proof: Define a trajectory $x = y(\tau)$ by solving the initial value problem

$$y(0) = z_0 ,$$
$$d\,y(\tau)/d\tau = -f'(y(\tau))^{-1}f(y(\tau)) \quad \text{for all } \tau \ge 0 .$$

Our first task is to discover for how large an interval $0 \le \tau < T$ the trajectory stays inside the ball $\| y(\tau) - z_0 \| < \delta$. We know that the trajectory exists inside the ball for some sufficiently small $T > 0$; let T be the largest value for which the trajectory stays inside the ball throughout $0 \le \tau < T$. Along the trajectory we find that $d\,f(y(\tau))/d\tau = f'(y(\tau))\,dy(\tau)/d\tau =$ $= f'(y(\tau))\,(-f'(y(\tau))^{-1}f(y(\tau))\,) = -f(y(\tau))$, whence follows that $f(y(\tau)) = e^{-\tau}f(z_0)$. Therefore the length of the trajectory from $\tau = 0$ up to $\tau = T$ is

$$\begin{aligned}\int_0^T \| dy(\tau)/d\tau \|\, d\tau &= \int_0^T \| f'(y(\tau))^{-1}f(y(\tau)) \|\, d\tau \\ &= \int_0^T \| f'(y(\tau))^{-1}f(z_0) \|\, e^{-\tau}\, d\tau \\ &< \int_0^T \delta\, e^{-\tau}\, d\tau = (1 - e^{-T})\,\delta \\ &< \delta .\end{aligned}$$

If T were finite we would observe that $\| y(T) - z_0 \|$ could not exceed the length of the trajectory from $\tau = 0$ to $\tau = T$, so $y(T)$ would have to lie strictly inside the ball, so T could not be the largest such value. Therefore the trajectory $y(\tau)$ lies in the ball for all $\tau \ge 0$ and has finite length less than the radius δ of the ball. Therefore a limit z exists inside the ball such that $y(\tau) \to z$ and $f(y(\tau)) = e^{-\tau}f(z_0) \to f(z) = o$ as $\tau \to \infty$. Thus is claim 2 proved.

Comment 1 : The same conclusion could be drawn from a different hypothesis $\delta > \|f'(x)^{-1}\|\,\|f(z_0)\|$. Besides being stronger (harder to satisfy), this hypothesis is affected by scaling; replacing $f(x)$ by $Lf(x)$ for any invertible linear operator L alters the stronger hypothesis but does not affect the root z , nor Newton's iteration, nor the Claims proved above. And yet the stronger hypothesis is the one more often applied: Given a constant $\sigma > \|f'(x)^{-1}\|$ throughout a region known to be farther than δ from all roots other than z , we infer that $\|z_0-z\| < \delta$ wherever $\|f(z_0)\| < \delta/\sigma$.

Comment 2 : Claim 2 concerns "at least one root" instead of "just one root" because the ball may contain more than one root. For example, let x be a complex variable, so that $\|x\| = |x|$; and let $f(x) = 1 + \exp(x)$, so that $z = \pm\iota\pi$. Choose $\delta = 2\pi$ and $z_0 = 3.13\iota$. Then $f'(x)^{-1}f(z_0) = \exp(-x)(1 + \exp(3.13\iota))$, so $|f'(x)^{-1}f(z_0)| = 2\exp(-\mathit{Re}(x))\cos(1.565)$. Inside the ball $|x - z_0| < \delta$, we find $-\mathit{Re}(x) < \delta = 2\pi$, whence follows that $|f'(x)^{-1}f(z_0)| < 2\exp(2\pi)\cos(1.565) = 6.2077\ldots < \delta$ too. Two roots $z = \pm\iota\pi$ lie in the ball.

Claim 3: If f' varies so slowly that $\|I - f'(z_o)^{-1}f'(x)\| < 1$
~~~~~~~~ for all x in some convex region R that includes z_o , then $f(x)$ can vanish at most once in R .

Proof: We shall find that $F(x) := x - f'(z_o)^{-1} f(x)$ "contracts" R in the sense that $\|F(x) - F(y)\| < \|x - y\|$ for all distinct x and y in R, and consequently that f can take no value, o or not, more than once in R. This last inequality follows from

$$\begin{aligned}\|F(x) - F(y)\| &= \| x - y - \int_0^1 f'(z_o)^{-1} f'(y + \tau(x-y))d\tau\,(x-y) \| \\ &\le \int_0^1 \| I - f'(z_o)^{-1} f'(y + \tau(x-y)) \| d\tau \; \|x - y\| \\ &< \|x - y\|, \quad \text{as claimed.}\end{aligned}$$

Here's an example more typical than before. Let's apply Newton's iteration to solve $(\lambda-\upsilon)(\lambda+\upsilon) + 2\lambda = 0$ and $2\upsilon(\lambda-3) = 0$ for roots $(\lambda, \upsilon) = (0, 0)$, $(-2, 0)$ or $(3, \pm\sqrt{15})$. Knowing the roots makes error bounds easier to compare with errors. Now $x = \begin{pmatrix}\lambda\\ \upsilon\end{pmatrix}$, $f(x) = \begin{pmatrix}(\lambda-\upsilon)(\lambda+\upsilon)+2\lambda\\ 2\upsilon(\lambda-3)\end{pmatrix}$, $f'(x) = 2\begin{pmatrix}\lambda+1 & -\upsilon\\ \upsilon & \lambda-3\end{pmatrix}$, and $\det(f'(x)) = 4((\lambda-1)^2+\upsilon^2-2^2)$, so $f'(x)^{-1}$ exists everywhere except on the boundary of a circle of radius 2 around $x = \begin{pmatrix}1\\ 0\end{pmatrix}$. Indeed, $f'(x)^{-1} = 2\begin{pmatrix}\lambda-3 & \upsilon\\ -\upsilon & \lambda+1\end{pmatrix}/\det(f'(x))$.

Starting the iteration with $\upsilon = 0$ keeps $\upsilon = 0$ in all iterates, so the iteration behaves just like Newton's iteration applied to solve $\lambda^2 + 2\lambda = 0$, except possibly at $\lambda = 3$ where $0/0$ may occur. Convergence to $\lambda = 0$ or $\lambda = -2$ is rapid from any initial λ not too big nor too near -1. Suppose iteration is stopped at $\lambda = 0.001$, so $z_0 = \begin{pmatrix}0.001\\ 0\end{pmatrix}$ and $f(z_0) = \begin{pmatrix}0.002001\\ 0\end{pmatrix}$, and then $f'(x)^{-1}f(z_0) = 0.0010005\begin{pmatrix}3-\lambda\\ \upsilon\end{pmatrix}/(4 - \upsilon^2 - (1-\lambda)^2)$. Using the norm $\|\begin{pmatrix}\lambda\\ \upsilon\end{pmatrix}\| = \max\{|\lambda|, |\upsilon|\}$, we find throughout a ball $\|x-z_0\| < 0.005$ that $\|f'(x)^{-1}f(z_0)\| < \delta = 0.00101 < 0.005$. Hence $\|z-z_0\| < \delta$. With a litle work we may infer from Claim 3 that no other root x can satisfy $\|x - z_o\| < 0.9$ roughly.

Starting the iteration with $\lambda = 3$ keeps $\lambda = 3$ in all iterates, so the iteration behaves just like Newton's iteration applied to solve $15 - \upsilon^2 = 0$, except possibly at $\upsilon = 0$ where $0/0$ may occur. Other starting points lead to more complicated behaviour.

Suppose iteration is stopped at $\lambda = 3.001$, $\upsilon = 3.87$, so $z_0 = \begin{pmatrix}3.001\\ 3.87\end{pmatrix}$ and $f(z_0) = \begin{pmatrix}0.031101\\ 0.00774\end{pmatrix}$; then

$$f'(x)^{-1}f(z_0) = \begin{pmatrix}0.0155505(\lambda-3) + 0.00387\upsilon\\ -0.0155505\upsilon + 0.00387(\lambda-3)\end{pmatrix} / ((\lambda-3)(\lambda+1)+\upsilon^2)$$

Throughout $\|x-z_0\| < 0.005$, we find $\|f'(x)^{-1}f(z_0)\| < \delta = 0.0033$. After some work, we may infer from Claim 3 that no other root x can satisfy $\|x - z_o\| < 1.8$ roughly.

After four roots z have been found, how can we be sure that there are no more? That is an algebraic problem for which, in general, no simple numerical solution exists.

The foregoing examples were comparatively easy to handle because the derivative f' was not too complicated. In general, the derivative can be far too complicated to manipulate symbolically by hand; then different approaches are needed:

See Grcar's books and papers on Automated Differentiation; and see books on Interval Arithmetic.