Nonlinear Equations

Antony Jameson

Department of Aeronautics and Astronautics, Stanford University, Stanford, California, 94305

Nonlinear equations

1. Simple iteration of $x = g(x)$, convergence conditions
2. Error estimate after n steps of computation
3. Propagation of round off error
4. Asymptotic convergence rate; quadratic convergence
5. Solution of $f(x) = 0$
6. Newton's method
7. Global convergence of Newton's method
8. Examples of Newton's method
9. Secant method
10. Δ^2 acceleration

1 Functional iteration

To solve

$$x = g(x)$$

set

$$x_{n+1} = g(x_n)$$

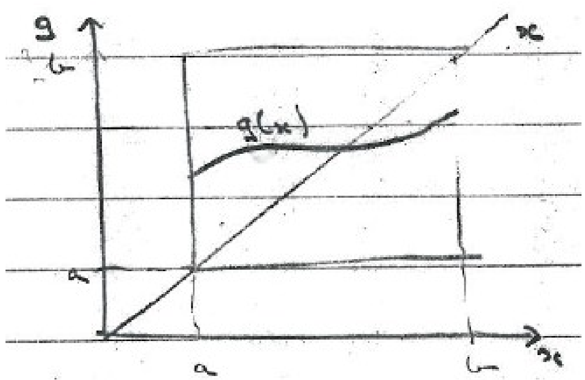

Figure 1:

Assume

1. On $[a,b]$ values of g lie in $[a,b]$
2. g is continuous with Lipschitz constant $L < 1$, $|g(x_1) - g(x_2)| \leq L|x_1 - x_2|$

Then

1. There exists a root s because $x - g(x)$ is continuous and changes sign
2. The root is unique since if s_1 and s_2 are roots

$$|s_1 - s_2| = |g(s_1) - g(s_2)| \leq L|s_1 - s_2|$$

   so

$$s_1 - s_2 = 0$$

3. Iteration starting from any x_0 in $[a,b]$ converges.

   Proof:

   (a) Since $g(x)$ lies in $[a,b]$ every iterate lies in $[a,b]$

   (b) Let s be the root, then

$$x_n - s = g(x_{n-1}) - s = g(x_{n-1}) - g(s)$$

$$
\begin{aligned}
|x_n - s| &\leq L|x_{n-1} - s| \\
&\leq L^n|x_0 - s| \\
&\rightarrow 0
\end{aligned}
$$

Corollary: Suppose that

$$|g'(x) \leq L < 1, \quad a \leq x \leq b$$

Then

$$g(x_2) = g(x_1) = g'(x^X)(x_2 - x_1), \quad x_1 \leq x^X \leq x_2$$

so the Lipschitz condition is satisfied with constant L and iteration converges.

The process is easily interpreted graphically.

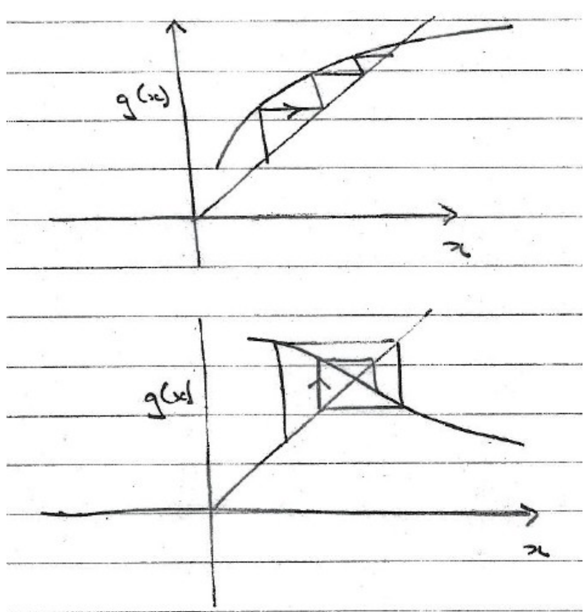

Figure 2:

This is an example of a 'contraction mapping'.

2 Error estimate in actual computation

The convergence proof gives an error estimate in terms of the root s which is not as yet known. To estimate the error after n steps of an actual computation we have

$$\begin{aligned} |x_{n+1} - x_n| &= |g(x_n) - g(x_{n-1})| \\ &\leq L|x_n - x_{n-1}| \\ &\leq L^n|x_1 - x_0| \end{aligned}$$

Also for fixed n

$$\begin{aligned} |x_m - x_n| &\leq |x_m - x_{m-1}| + |x_{m-1} - x_{m-2}| \cdots + |x_{n+1} - x_n| \\ &\leq (L^{m-1} \cdots + L^n)|x_1 - x_0| \\ &\leq \frac{L^n}{1-L}|x_1 - x_0| \end{aligned}$$

Thus since $x_m \to s$ as $m \to \infty$

$$|x_n - s| \leq \frac{L^n}{1-L}|x_1 - x_0|.$$

3 Propagation of round off errors

Due to round off the actual iterates are

$$x_{n+1} = g(x_n) + \delta_n$$

Let s be a root and let

$$|g(x) - g(s)| \leq L|x - s| \quad \text{when} \quad |x - s| < p$$

$$\delta_n \leq \delta$$

and let $|x_0 - s| \leq p_0$ where

$$0 < p_0 \leq p - \frac{\delta}{1 - L}$$

Then iterates lie in interval $|x_n - s| \leq p$ and

$$|x_n - s| \leq \frac{\delta}{1 - L} + L^n \left(p_0 - \frac{\delta}{1 - L} \right)$$

Proof: x_0 is in the interval. Assume $x_1, \ldots x_{n-1}$ are in the interval. Then

$$x_n - s = g(x_{n-1}) - g(s) + \delta_n$$

$$
\begin{aligned}
|x_n - s| &\leq L|x_{n-1} - s| + \delta \\
&\leq L(L|x_{n-2} - s| + \delta) + \delta \\
&\quad \dots \\
&\leq L^n|x_0 - s| + \frac{1 - L^n}{1 - L}\delta \\
&\leq L^n p_0 + \frac{\delta}{1 - L} - L^n \frac{\delta}{1 - L} \\
&\leq p_0 + \frac{\delta}{1 - L} \\
&\leq p
\end{aligned}
$$

so all iterates are in the required interval and we also obtain the convergence estimate from the 3rd line from the bottom.

4 Asymptotic convergence and quadratic convergence

Consider errors $e_n = x_n - s$ where s is a root. We have

$$
\begin{aligned}
e_{n+1} &= x_{n+1} - s \\
&= g(s + e_n) - g(s) \\
&= g'(s + \theta_n e_n)e_n, \quad 0 < \theta_n < 1.
\end{aligned}
$$

Also $e_n \to 0$. Thus

$$
\operatorname*{Lim}_{n \to \infty} \frac{e_{n+1}}{e_n} = g'(s)
$$

We can see that convergence rate improves as $g'(s)$ becomes smaller.

Consider case where $g'(s) = 0$ and g' is continuous. First note that from continuity of g'

$$|g'(x)| = |g'(x) - g'(s)| \leq L < 1$$

for a small enough interval, assuming convergence. Also now

$$\begin{aligned} e_{n+1} &= g(s + e_n) - g(s) \\ &= g'(s)e_n + g''(s + \theta_n e_n)\frac{e_n^2}{2} \end{aligned}$$

and since $g'(s) = 0$

$$\operatorname*{Lim}_{n\to\infty} \frac{e_{n+1}}{e_n^2} = \frac{1}{2}g''(s).$$

This is 'quadratic' convergence.

5 Iteration for $f(x) = 0$

Set

$$g(x) = x - Mf(x)$$

Then

$$g' = 1 - Mf'$$

Thus the method converges if x_0 lies in an interval $|x_0 - s| < p$ on which $0 < Mf' < 2$. It can be seen that M must have the same sign as $f'(s)$. The method amounts to setting

$$x_{n+1} = x_n - Mf(x_n)$$

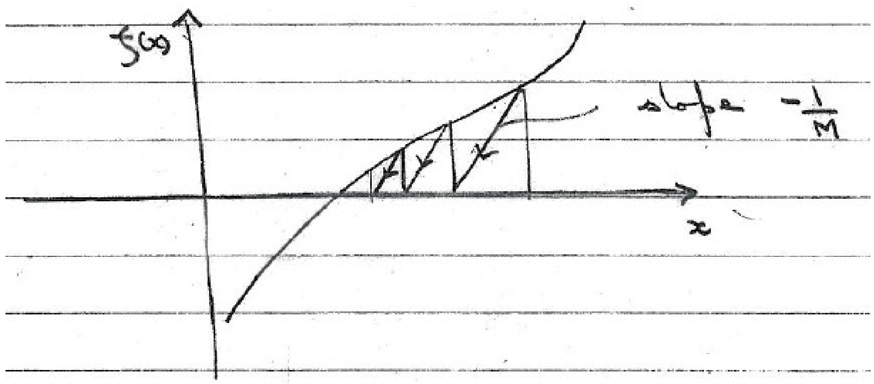

Figure 3:

<u>Round off error and functional iteration:</u>

Suppose $|L| < 1$ and

$$x_n - s = g(x_{n-1}) - g(s) + \delta_n$$

Then

$$\begin{aligned} |x_n - s| &\leq L|x_{n-1} - s| + \delta \\ &\leq L\left(L|x_{n-2} - s| + \delta\right) + \delta \\ &\leq L^n|x_0 - s| + \delta(1 + L + L^2 \cdots + L^{n-1}) \\ &= L^n|x_0 - s| + \frac{1 - L^n}{1 - L}\delta \\ &\leq L^n|x_0 - s| + \frac{\delta}{1 - L} \end{aligned}$$

6 Newton's method

We have seen that if $g'(s) = 0$ convergence is quadratic, so we try to define g with this property. Let

$$g(x) = x - m(x)f(x)$$

$$g'(x) = 1 - m'(x)f(x) - m(x)f'(x)$$

$$g'(s) = 1 - m(s)f'(s)$$

Thus the required property is obtained by choosing

$$m(x) = \frac{1}{f'(x)}$$

Then

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$

This is Newton's method. Geometrically it means simply setting x_{n+1} as the intercept of the tangent with the x axis.

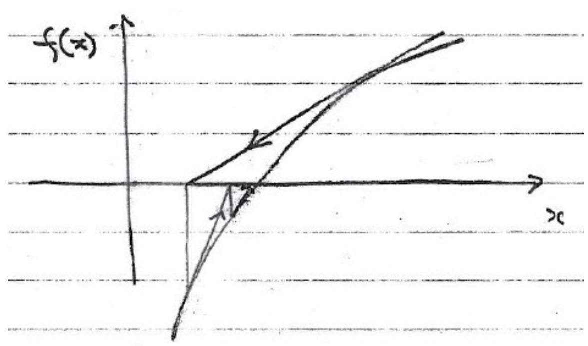

Figure 4:

In the case of a multiple root we have, say

$$f(x) = (x - s)^p h(x)$$

where $h(x)$ has a second derivative and $h(s) \neq 0$. Then

$$\begin{aligned} f'(x) &= p(x-s)^{p-1}h(x) + (x-s)^p h'(x) \\ &= (x-s)^{p-1}\left(ph(x) + (x-s)h'(x)\right) \end{aligned}$$

$$\begin{aligned} f''(x) &= p(p-1)(x-s)^{p-2}h(x) + 2p(x-s)^{p-1}h'(x) + (x-s)^p h''(x) \\ &= (x-s)^{p-2}\big(p(p-1)h(x) + 2p(x-s)h'(x) + (x-s)^2 h''(x)\big) \end{aligned}$$

Also

$$g(x) = x - \frac{f}{f'}$$

$$g'(x) = \frac{ff''}{f'^2}$$

Thus

$$\begin{aligned} g'(s) &= \frac{p(p-1)h^2(s)}{p^2h^2(s)} \\ &= 1 - \frac{1}{p} \end{aligned}$$

Thus $|g'(x)| < 1$ for x sufficiently near s and the method converges but only linearly.

<u>Newton's method:</u>

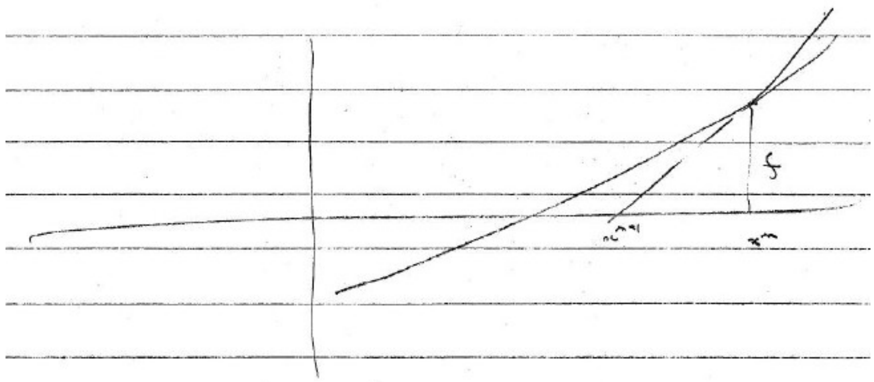

Figure 5: Slope is $\dfrac{f}{x^n - x^{n+1}} = f'$, i.e. $x^{n+1} - x^n = -\dfrac{f}{f'}$.

7 Global convergence of Newton's method

General conditions for convergence of Newton's method are as follows:

Theorem: Let f be defined and twice continuously differentiable on $[a, b]$, and let

1. $f(a)f(b) < 0$
2. $f'(x) \neq 0$ on $[a, b]$
3. $f''(x) \geq 0$ or $f''(x) \leq 0$ on $[a, b]$
4. if c is endpoint at which $|f'(x)|$ is smaller, $\left|\dfrac{f(c)}{f'(c)}\right| \leq b - a$

Consider $f(a) < 0$, $f(b) > 0$, $f''(x) \leq 0$

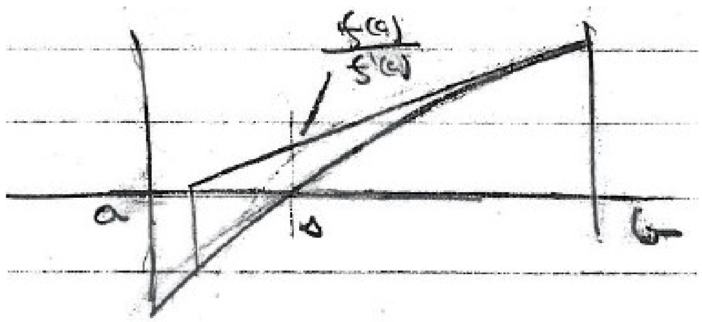

Figure 6:

Then $f'(x) > 0$. Let $a \leq x_0 \leq s$. Then

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \leq x_0$$

since $f(x_0) < 0$. We show by induction that $x_n \leq s$, $x_{n+1} \geq x_n$. By the mean value theorem

$$-f(x_n) = f(s) - f(x_n) = (s - x_n)f'(x_n^*), \quad x_n \leq x_n^* \leq s$$

Since $f'' \leq 0$ f' is decreasing,

$$f'(x_n^*) \leq f'(x_n)$$

$$-f(x_n) \leq (s - x_n)f'(x_n)$$

$$x_{n+1} = x_n - \frac{f(x_n}{f'(x_n)} \leq x_n + s - x_n = s$$

****

Thus x_n is a bounded monotone sequence and must reach a limit q say. Then

$$q = q - \frac{f(q)}{f'(q)}$$

so since $f'(q) \neq 0$ $f(q) = 0$ and q must be the unique root s.

If $s < x_0 \leq b$

$$f(x_0) = f'(x_0^*)(x_0 - s), \quad s < x_0^* < x_0$$

and since f' is decreasing

$$f(x_0) \geq (x_0 - s) f'(x_0)$$

$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)} \leq x_0 - (x_0 - s) = s$$

On the other hand

$$f(x_0) = f(b) - (b - x_0) f'(x_0^+), \quad x_0 \leq x_0^+ \leq b$$

whence

$$f(x_0) \leq f(b) - (b - x_0) f'(b)$$

and using (4)

$$\begin{aligned} x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)} \\ &\geq x_0 - \frac{f(x_0)}{f'(b)} \\ &\geq x_0 - \frac{f(b)}{f'(b)} + b - x_0 \\ &\geq a \end{aligned}$$

Now we are in the same situation as before.

8 Special cases of Newton's method

(a) Square roots. To find $\sqrt{c}$ set

$$\begin{aligned} f(x) &= x^2 - c \\ x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} \\ &= x_n - \frac{x_n^2 - c}{2x_n} \\ &= \frac{1}{2}\left(x_n + \frac{c}{x_n}\right). \end{aligned}$$

Then $f'(x) > 0$, $f''(x) > 0$ so function is convex.

(b) To find k^{th} root set

$$\begin{aligned} f(x) &= x^k - c \\ x_{n+1} &= x_n - \frac{x_n^k - c}{kx_n^{k-1}} \\ &= \left(1 + \frac{1}{k}\right) x_n + \frac{c}{kx_n^{k-1}}. \end{aligned}$$

9 Secant method

Sometimes it is difficult to evaluate f'. For example f may be determined numerically. Then a simple procedure is to approximate f' by a difference

$$f'(x_n) \approx \frac{f(x_n) - f(x_n - h)}{h}$$

If we take for $x_n - h$ the previous value x_n we obtain the secant method, which can be regarded as a 'quasi Newton' method. Denote $f(x_n)$ by f_n. After evaluating f_0, f_1, set

$$x_{n+1} = x_n - \frac{(x_n - x_{n-1})f_n}{f_n - f_{n-1}} \tag{9.1}$$

This amounts to travelling along chords. The formulation

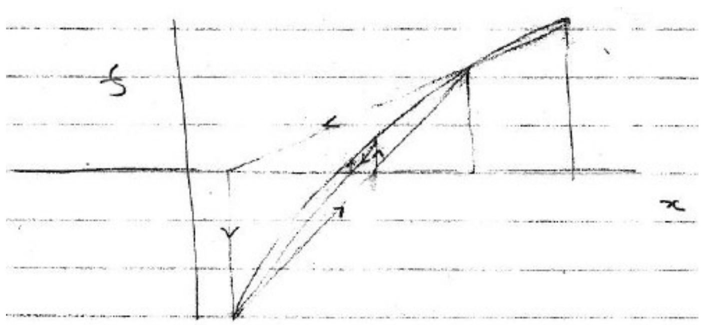

Figure 7:

$$x_{n+1} = \frac{x_{n-1}f_n - x_n f_{n-1}}{f_n - f_{n-1}}$$

is more sensitive to cancellation and should be avoided.

To estimate the order of convergence of the secant method we have for any x, from equation 9.1

$$(x_{n+1} - x)\frac{f_n - f_{n-1}}{x_n - x_{n-1}} = (x_n - x)\frac{f_n - f_{n-1}}{x_n - x_{n-1}} - f_n$$

But according to the error estimate for linear interpolation

$$f(x) = f_n - \frac{x_n - x}{x_n - x_{n-1}}(f_n - f_{n=1}) + \frac{1}{2}(x_n - x)(x_{n-1} - x)f''(\xi)$$

where ξ lies in the interval $(x_{n-1}, x_n$. Let s be a root, $f(s) = 0$. Setting $x = s, x_n - s = e_n$, we obtain

$$e_{n+1}\frac{f_n - f_{n-1}}{x_n - x_{n-1}} = \frac{1}{2}f''(\xi)e_n e_{n-1}$$

Also by the mean value theorem

$$\frac{f_n - f_{n-1}}{x_n - x_{n-1}} = f'(\eta)$$

where η lies in the interval (x_{n-1}, x_n) Thus

$$e_{n+1} = \frac{f''(\xi)}{2f'(\eta)}e_n e_{n-1}$$

Assume that $\left|\frac{f''(\xi)}{2f'(\eta)}\right| \leq M$ in some interval about the root. Then setting $d_n = |Me_n|$,

$$d_{n+1} \leq d_n d_{n-1}$$

Thus if $\delta = \max(d_0, d_1)$

$$d_2 \leq \delta^2$$

$$d_3 \leq \delta^3$$

$$d_4 \leq \delta^5$$

$$d_k \leq \delta^{m_k}$$

where

$$m_{k+1} = m_k + m_{k-1}$$

The indices m_k form a Fibonacci sequence. Solving the difference equation we obtain

$$m_k = \frac{1}{\sqrt{5}}\left(r_+^{k+1} - r_-^{k+1}\right)$$

where

$$r_+ = \frac{1+\sqrt{5}}{2}, \;\; r_- = \frac{1-\sqrt{5}}{2}$$

In the limit

$$m_k \rightarrow \frac{1}{\sqrt{5}} r_+^{k+1}$$

and

$$d_k \leq \delta^{\frac{1}{\sqrt{5}} r_+^{k+1}} = d_{k-1}^{r_+} = d_{k-1}^{1.618}$$

Thus the method has a fractional order of 1.618.

A variant of the secant method is the regula falsi method. In this method the chord is drawn between f_n and f_m where f_m is the most recent value of opposite sign to f_n.

With this method if f is convex on (x_0, x_1) f_0 is always retained so that

$$e_{n+1} = \frac{f''(\xi)}{2f'(\eta)} e_n e_0$$

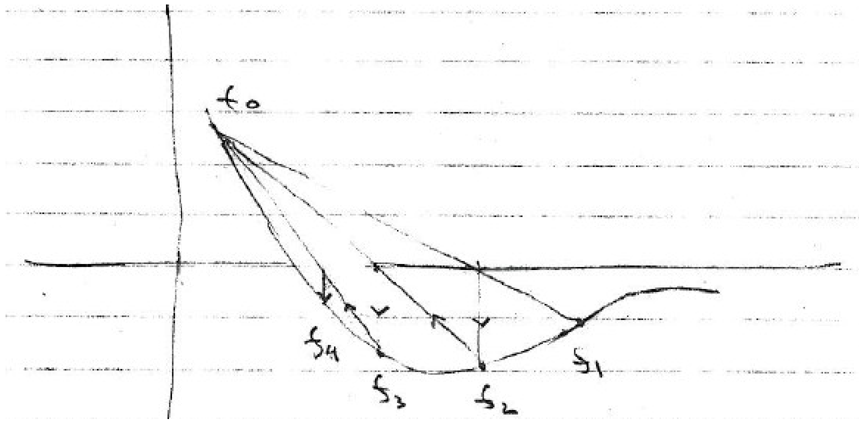

Figure 8:

or

$$|e_{n+1}| \leq |Me_0||e_n|$$

and convergence is linear. The scheme has the advantage, on the other hand, of being globally convergent, and it might be used in the early stages before switching to the secant method.

We note that with the standard secant method the function changes sign every third step near the root, where f' and f'' have a constant sign, so that $\frac{e_{n+1}}{e_n e_{n-1}}$ has a constant sign. Then if f_0 and f_1 have opposite signs the only possible sequences of signs are

$$+ - + + - + + - + + - + \ldots$$

or

$$+ - - + - - + - - + - - \ldots$$

Thus regula falsi can never coincide with the secant method.

Secant method:

Estimate

$$f'(x_n) = \frac{f_n - f_{n-1}}{x_n - x_{n-1}}$$

Then set

$$x_{n+1} = x_n - \frac{(x_n - x_{n-1})f_n}{f_n - f_{n-1}}$$

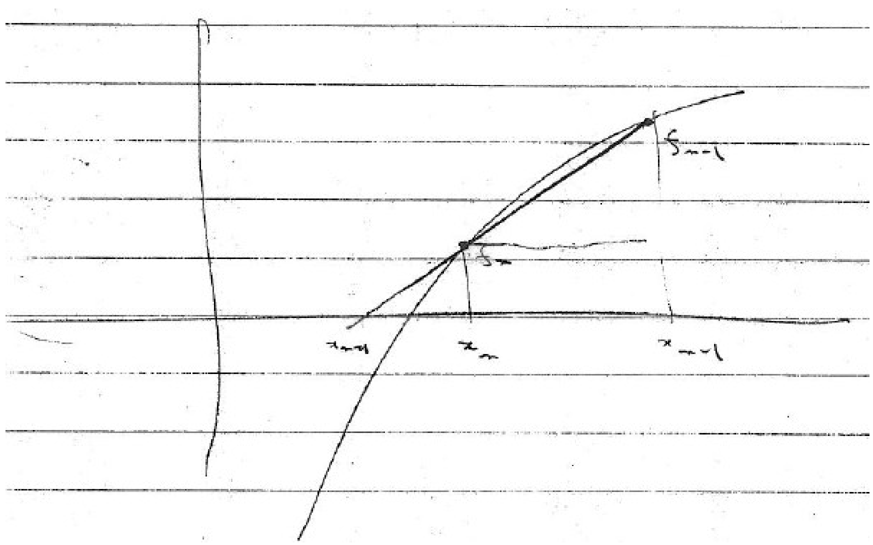

Figure 9:

10 Aitken Δ^2 method

We found that simple iteration yields, say

$$\text{Lim} \frac{x_{n+1} - s}{x_n - s} = g'(s) = A$$

If this were exact we would have

$$x_{n+1} - s = A(x_n - s)$$

$$x_{n+2} - s = A(x_{n+1} - s)$$

so that we could solve for A and s. Subtracting the first equation from the second.

$$A = \frac{x_{n+2} - x_{n+1}}{x_{n+1} - x_n},\ A - 1 = \frac{x_{n+2} - 2x_{n+1} + x_n}{x_{n+1} - x_n}$$

Then the 1^{st} equation gives

$$\begin{aligned} s &= \frac{x_{n+1} - Ax_n}{1 - A} \\ &= x_n + \frac{x_{n+1} - x_n}{1 - A} \\ &= x_n - \frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n} \end{aligned}$$

This suggests that the quantities

$$x_n' = x_n - \frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n}$$

will give a better approximation to s than x_n.

Write

$$\Delta x_n = x_{n+1} - x_n$$

$$\Delta^2 x_n = \Delta(\Delta x_n) = x_{n+2} - 2x_{n+1} + x_n$$

Then

$$x_n' = x_n + \frac{\Delta x_n^2}{\Delta^2 x_n}$$

This is Aitken's Δ^2 method for improving the convergence of a sequence.

Theorem: Let any sequence satisfy

$$e_{n+1} = (A + \delta_n) e_n$$

where $e_n = x_n - s$, $|A| < 1$, and $\delta_n \to 0$ as $n \to \infty$. Then the Δ^2 sequence converges faster in the sense that

$$\frac{x_n' - s}{x_n - s} \to 0 \quad \text{as } n \to \infty$$

Proof: We have

$$e_{n+1} = (A + \delta_n) e_n$$

Thus

$$\Delta x_n = \Delta e_n = (A - 1 + \delta_n) e_n$$

and

$$(\Delta x_n)^2 = (A - 1)^2 e_n^2 + \eta e_n^2$$

where

$$\eta = \delta_n^2 + 2(A-1)\delta_n$$

Also

$$e_{n+2} = (A + \delta_{n+1})(A + \delta_n)e_n$$

Thus

$$\begin{aligned} \Delta_2 x_n = \Delta^2 e_n &= e_{n+2} - 2e_{n+1} + e_n \\ &= \big((A-1)^2 + p\big)\, e_n \end{aligned}$$

where

$$p = A(\delta_n + \delta_{n+1}) - 2\delta_n + \delta_n\delta_{n+1}$$

Then

$$\begin{aligned} x_n' - s &= e_n - \frac{(\Delta x_n)^2}{\Delta^2 x_n} \\ &= e_n\left(1 - \frac{(A-1)^2 + \eta}{(A-1)^2 + p}\right) \end{aligned}$$

and

$$\frac{x_n' - s}{e_n} = \frac{p - \eta}{(A-1)^2 + p} \to 0 \quad \text{as } n \to \infty$$

11 Comments on Δ^2 method

1. The evaluation of expression such as

$$\frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n}$$

   involve taking small differences of numbers that need not be small, leading to possibility of serious round off errors.

2. The application of Δ^2 acceleration requires the x_n sequence to be evaluated 2 steps ahead of the x'_n sequence.

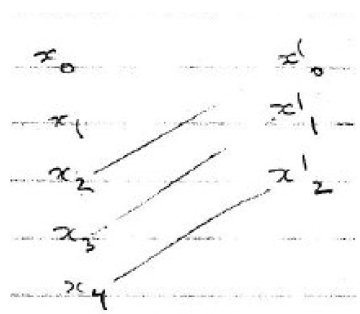

Figure 10:

12 Steffensen's method

One could apply a Δ^2 acceleration after 2 steps or 3 function evaluations at x_0, x_1, x_2. Then

$$x_0^{(1)} = x_0 - \frac{-(x_1 - x_0)^2}{x_2 - 2x_1 + x_0}$$

Now one could take 2 steps using simple iteration from $x_0^{(1)}$ and apply Δ^2 acceleration again to obtain

$$x_0^{(2)} = x_0^{(1)} - \frac{-\left(x_1^{(1)} - x_0^{(1)}\right)^2}{x_2^{(1)} - 2x_1^{(1)} + x_0^{(1)}}$$

This is equivalent to solving $x = G(x)$ by simple iteration where

$$G(x) = x - \frac{(g(x) - x)^2}{g\left(g(x)\right) - 2g(x) + x}$$

At $x = s$ $G(s)$ is defined by L'Hopital's rule as

$$\begin{aligned} G(s) &= s - \frac{2\left(g(s) - s\right) g'(s)}{g'\left(g(s)\right) g'(s) - 2g'(s) + 1} \\ &= s \end{aligned}$$

To solve $f(x) = 0$ by Steffensen's method we can set

$$x = g(x), \quad g(x) = x + f(x)$$

Then we have

$$\begin{aligned} G(x) &= x - \frac{f^2(x)}{x + f(x) + f(x + f(x)) - 2(x + f(x)) + x} \\ &= x - \frac{f^2(x)}{f(x + f(x)) - f(x)} \end{aligned}$$

so the iteration becomes

$$x_{n+1} = x_n - \frac{f_n}{d_n}$$

where

$$f_n = f(x_n), \quad d_n = \frac{f(x_n + h) - f(x_n)}{h}$$

with

$$h = f_n.$$

Thus we can regard this as a quasi Newton method with a finite difference approximation to f' with a special choice of step length h.

To estimate the order of convergence of Steffensen's method in this case expand d_n as a Taylor series

$$\begin{aligned} d_n &= \frac{f(x_n + h) - f(x_n)}{h} \\ &= \frac{1}{h}\left(hf'(x_n) + \frac{h^2}{2}f''(\eta)\right) \end{aligned}$$

where η lies between x_n and x_nth, and $h = f_n$. Thus

$$d_n = f'(x_n) + \frac{1}{2}f_n f''(\eta) = f'_n\left(1 + \frac{1}{2}\frac{f_n}{f'_n}f''(\eta)\right)$$

and setting $e_n = x_n - s$, and remembering that f_n is small near a root we have

$$e_{n+1} = e_n - \frac{f_n}{f'_n}\left(1 - \frac{1}{2}\frac{f_n}{f'_n}f''(\eta) + O(f_n^2)\right)$$

But expanding $f(x)$ as a Taylor series we have

$$O = f(s) = f_n + (s - x_n)f'_n + \frac{(s - x_n)^2}{2}f''(\xi)$$

where ξ lies between x_n and s so

$$\frac{f_n}{f'_n} = e_n - \frac{e_n^2}{2}\frac{f''(\xi)}{f'_n}$$

Thus

$$\begin{aligned} e_{n+1} &= e_n - e_n\left(1 - \frac{e_n}{2}\frac{f''(\xi)}{f'_n}\right)\left(1 - \frac{1}{2}e_n f''(\eta) + O(e_n^2)\right) \\ &= \frac{e_n^2}{2}\left(\frac{f''(\xi)}{f'_n} + f''(\eta) + O(e_n^2)\right) \end{aligned}$$

and

$$\frac{e_{n+1}}{\frac{1}{2}e_n^2} \to \frac{f''(s)}{f'(s)}(1 + f'(s))$$

To determine the order of Steffensen's method in general let

$$g'(s) = g''(s) \ldots = g^{(p-1)}(s) = 0$$

$$g^{(p)}(s) = p!A \neq 0$$

$g^{(p+1)}(x)$ exists in $|x - s| \leq p$. Then

$$\begin{aligned} g(s+\epsilon) &= g(s) + A\epsilon^p + \frac{g^{(p+1)}(s+\theta\epsilon)}{(p+1)!}\epsilon^{p+1}, \quad 0 < \theta < 1 \\ &= s + \delta \end{aligned}$$

where

$$\delta = A\epsilon^p + B\epsilon^{p+1}$$

Also similarly

$$g(s+\delta) = s + A\delta^p + C\delta^{p+1}, \quad C = \frac{g^{(p+1)}(s+\phi\delta)}{(p+1)!}$$

Thus

$$\begin{aligned} G(s+\epsilon) &= \frac{(s+\epsilon)g(s+\delta) - (s+\delta)^2}{g(s+\delta) - 2(s+\delta) + A + \epsilon} \\ &= s - \frac{\delta^2 - A\epsilon\delta^p - C\epsilon\delta^{p+1}}{\epsilon - 2\delta + A\delta^p + C\delta^{p+1}} \end{aligned}$$

Substituting for δ we obtain $G(s+\epsilon)$ in terms of ϵ. If $p \geq 2$ we get

$$G(s+\epsilon) = s + A^2\epsilon^{2p-1} + O\left(\epsilon^{2p}\right)$$

For $p = 1$, provided $A = g'(s) \neq 1$ we get

$$G(s+\epsilon) = s - \frac{A}{1-A} g''(s) \frac{\epsilon^2}{2} + O(\epsilon^3)$$

The case $g(s) = 1$ corresponds to a multiple root of f, since $g = x - f(x)$. Then we get for a root of multiplicity m

$$G(s+\epsilon) = s + \left(1 - \frac{1}{m}\right)\epsilon + O(\epsilon^2)$$

Now if

$$G(s+\epsilon) = s + C\epsilon^q + O(\epsilon^{q+1})$$

then

$$G'(s) = G''(s) \ldots = G^{(q-1)}(s) = 0.$$

Proof: By Taylor theorem

$$G(s+\epsilon) = s + \epsilon G'(s) \ldots + \frac{\epsilon^q}{q!} G^q(s) + \frac{\epsilon^{q+1}}{(q+1)!} G^{(q+1)}(s+\theta\epsilon)$$

Then equate powers of q.

We deduce:

If functional iteration is order p, the method is order $2p-1$. If functional iteration is order 1, the method is order 2 except for multiple root where it is of order 1.

13 Systems of nonlinear equations

We now consider

$$x = g(x)$$

where x and g are vectors, i.e.

$$x_i = g_i(x_1, x_2 \ldots, x_n), \quad i = 1, 2 \ldots, n$$

Convergence Theorem:

Suppose there exists a region R, $a_i \le x_i \le b_i$, in which

(1) g_i is continuous

(2) $g(x)$ lies in R if x lies in R

(3) $\|g(x_1) - g(x_2)\| \leq L\|x_1 - x_2\|, \quad L < 1$ (generalized Lipschitz condition)

Then

(a) there is a solution s in R and it is unique

(b) for any x_0 in R the sequence $x_{n+1} = g(x_n)$ converges to s

(c) $\|x_n - s\| \leq \dfrac{L^n}{1-L}\|x_1 - x_0\|$

Note: Assumption (3) states that the mapping $x = g(x)$ reduces the 'distance' between any 2 points. It is therefore called a 'contraction mapping'.

Proof of theorem:

As in scalar case (3) implies

$$\|x_{n+1} - x_n\| \leq L^n\|x_1 - x_0\|$$

Now consider n fixed. Then

$$\begin{aligned}
\|x_m - x_n\| &\leq \|x_m - x_{m-1}\| + \|x_{m-1} - x_{m-2}\| \ldots + \|x_{n+1} - x_n\| \\
&\leq (L^{m-1} + L^{m-2} \ldots + L^n)\|x_1 - x_0\| \\
&\leq \frac{L^n}{1-L}\|x_1 - x_0\|
\end{aligned}$$

The expression on the right $\to 0$ with n independent of m, so x_n is a Cauchy sequence, and has a limit s. Since R is compact s is in R. Also from the continuity of g

$$\operatorname{Lim} g(x_n) = g(s)$$

and

$$s = \operatorname*{Lim}_{n\to\infty} x_n = \operatorname*{Lim}_{n\to\infty} x_{n+1} = \operatorname{Lim} g(x_n) = g(s)$$

so s is a root. It is unique because if s_1,s_2 are roots

$$\|s_1 - s_2\| = \|g(s_1) - g(s_2)\| \leq L\|s_1 - s_2\|$$

where $L < 1$.

14 Estimate of Lipschitz constant

To estimate the Lipschitz constant L we have

$$g(x) = g(y) + J(\xi)(x - y)$$

where $J = \left[\frac{\partial g_i}{\partial x_j}\right]$ is Jacobian.
Thus

$$\|g(x) - g(y)\| \leq \|J(\xi)\|\|x - y\|$$

15 Newton's method for a system of equations

For a system of equations Newton's method is easily generalized. To solve

$$f(x) = 0$$

let J be Jacobian matrix $\left[\frac{\partial f_i}{\partial x_j}\right]$. Then for a small change we have

$$f(x+\delta x) = f(x) + J\delta x + \text{higher order terms}$$

We choose δx so that to this order

$$f(x+\delta x) = 0.$$

That is we set

$$\delta x = -J^{-1} f$$

$$x_{n+1} = x_n - J^{-1}(x_n) f(x_n).$$

At each step we have to solve the linear system of equations

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

To show that the method is 2^{nd} order we have

$$g(x) = x - J^{-1} f(x)$$

$$\frac{\partial g(x)}{\partial x_j} = \frac{\partial x}{\partial x_j} - J^{-1}\frac{\partial f(x)}{\partial x_j} - \frac{\partial}{\partial x_j}(J^{-1}) f(x)$$

Setting $x = s,\ f(s) = 0$, we get

$$\frac{\partial g(s)}{\partial x} = I - J^{-1}J = 0.$$

Rate of convergence:

$$\begin{aligned}\frac{e_{n+1}}{e_n} &= \frac{x_{n+1}-s}{x_n-s} = \frac{g(x_n)-g(s)}{x_n-s} \\ &= g'(\xi)\end{aligned}$$

for ξ in interval between x_n and s.

16 Generalized quasi Newton or secant method

If it is hard to obtain the Jacobian $J = \left[\frac{\partial f_i}{\partial x_j}\right]$ we can estimate by finite differences using the steps made so far, as in the regula falsi method. In fact to save the inversion of J we can directly estimate J^{-1}. If J were constant we would have

$$\delta f = J\delta x$$

$$\delta x = J^{-1}\delta f$$

which would be exact as system was then linear. From n linearly independent δf we could then solve for J^{-1}. Start with an initial guess H_0 for J^{-1}, typically $H_0 = I$. Then choose H_1 so that

$$H_1\delta f_0 = \delta x_0$$

over 1^{st} step.

Let

$$H_1 = H_0 + D_0$$

Then

$$D_0 \delta f_0 = \delta x_0 - H_0 \delta f_0$$

This is satisfied by

$$D_0 = \frac{(\delta x_0 - H_0 \delta f_0) z^T}{z^T \delta f_0}$$

for any z for which the denominator exists.

Thus we have the update rule

$$H_{i+1} = H_i + \frac{(\delta x_i - H_i \delta f_i) z_i^T}{z_i^T \delta f_i}$$

$$H_{i+1} \delta f_i = \delta x_i.$$

Now if we choose z_i as a direction orthogonal to all previous δf_i, then

$$H_{i+1} \delta f_j = H_{j+1} \delta f_j = \delta x_j$$

so after n steps we have

$$H_n \delta f_j = \delta x_j, \quad j = 0, \ldots, n-1$$

so if the δf are independent and J is constant, $H = J^{-1}$. At each iteration we then make the estimated Newton step

$$\delta x_i = x_{i+1} - x_i = -H_i f_i.$$

To choose the z_i we can use successively

$$z_0 = \delta f_0$$

$$z_1 = \delta f_1 - \beta_{10} z_0$$

$$z_2 = \delta f_2 - \beta_{20} z_0 - \beta_{21} z_1$$

$$\ldots$$

Then

$$z_1^T \delta f_0 = \delta f_1^T \delta f_0 - \beta_{10} \delta f_0^T \delta f_0 = 0$$

giving

$$\beta_{10} = \frac{\delta f_1^T \delta f_0}{\delta f_0^T \delta f_0}$$

Next

$$z_2^T \delta f_0 = \delta f_2^T \delta f_0 - \beta_{20} \delta f_0^T \delta f_0 = 0$$

giving

$$\beta_{20} = \frac{\delta f_2^T \delta f_0}{\delta f_0^T \delta f_0}$$

and

$$z_2^T \delta f_1 = \delta f_2^T \delta f_1 - \beta_{20} \delta f_0^T \delta f_1 - \beta_{21} \delta f_1^T \delta f_1 = 0$$

giving

$$\beta_{21} = \frac{\delta f_2^T \delta f_1 - \beta_{20} \delta f_0^T \delta f_1}{\delta f_1^T \delta f_1}$$

etc. This is Gram Schmidt orthogonalization.

Conclusions

Acknowledgements

References