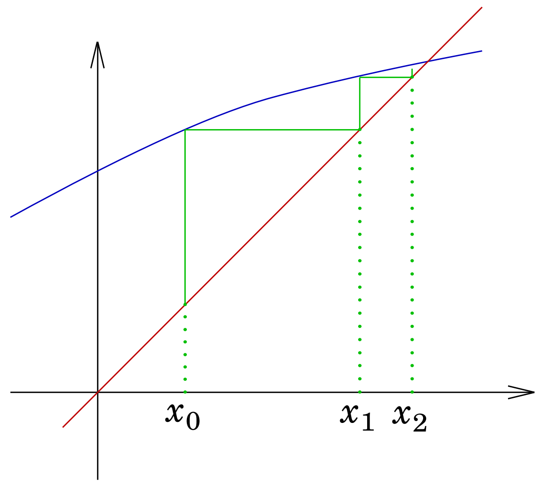
Non-linear equations

Iterative methods [continued]

Diagrammatically, we can see what is happening in the case of first-order convergence:



For each value of x, we go up to the curve to find $g(x)$; that value replaces x, corresponding to going across the diagram until we hit the line $y = x$, which gives us the next value of x.

If you try the same thing with $g(x)$ steeper than 1, you see that its graph crosses $y = x$ from below, and the same construction takes you further and further away. If you try with the slope of $g(x)$ negative, you will find that the construction 'spirals' around the solution, making a so-called 'cobweb' diagram; for small negative slope it spirals in, for large negative slope it spirals out. Both of these constructions are worth drawing your own diagrams for!

Note that although the *theory* is about the value of $g'(p)$, where p is the root, in real life we don't know p so can't test this value directly. We usually 'suck it and see'; but in some cases it may be worth evaluating $g'(x_n)$ once in a while, to see whether the method seems to be working.

Special case:

In the case where $|g'(x)| < 1$ for *all* x, it is easy to see that, in the previous diagram, the curve crosses the line once and once only. That is, the equation $g(x) = x$ has exactly one solution. [This is a topological theorem—look up 'contraction mapping'!] You won't often spot such cases, but it does mean that equations like $x = 2 + \sin(x)$ [and a host of similar equations] have exactly one solution which we can guarantee to find [though perhaps not quickly] by the obvious iteration.

Convergence of Newton-Raphson

For Newton-Raphson,

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

we have

$$g'(x) = 1 - \frac{f'(x)^2 - f(x)f''(x)}{f'(x)^2} = \frac{f(x)f''(x)}{f'(x)^2}$$

[using the quotient rule], so that $g'(p) = 0$ at the root, $f(p) = 0$, *provided that* $f'(p) \neq 0$. So N-R *usually* has quadratic convergence, once it gets near the root. But beware the special case, which will happen if f has a multiple root at p, and will very nearly happen in the cases we talked about earlier where f' happens to be very small.

Three cases of historical interest

(a) If we try to solve the quadratic $x^2 = a$, then Newton-Raphson gives us:

$$x \to x - \frac{x^2 - a}{2x} = \tfrac{1}{2}(x + a/x).$$

This is the actual way most calculators/computers find square roots, as it converges very rapidly from any reasonable starting value. *E.g.* to find $\sqrt{2}$, so $x \to \frac{1}{2}(x+2/x)$, we find

$$1 \to 1.5 \to 1.4166666667 \to 1.4142156863 \to 1.4142135624$$

[correct to my calculator accuracy]. *Exercise:* Try $x^n = a$ for values of n other than 2.

(b) If $1/x = a$, then Newton-Raphson gives

$$x \to x - \frac{1/x - a}{-x^{-2}} = 2x - ax^2 = x \times (2 - ax).$$

This gives a way of finding reciprocals without dividing! Early computers took ages to divide, so a few times around N-R from a decent starting value was significantly faster. Then computer chips got better at dividing. But more recently, RISC chips often can't divide again, and this method has come back into use on some computers.

(c) If you solve the equation $x^3 = 1$ using N-R, and yes I know we all know what the answer is, then it converges for almost all starting values x_0. If you do this with *complex* values of x_0, then it sometimes converges to $x = 1$, but also sometimes to the complex roots, $x = -(1 \pm \sqrt{3}\,\mathrm{i})/2$. If you colour the starting point by which root it converges to, you get a pretty fractal picture in the Argand diagram.

Aitken's Device

This is a 'trick' to make slow convergence faster. If we have first-order convergence, then we have

$$x_0 = p+h; \; x_1 \approx p+rh; \; x_2 \approx p+r^2h,$$

where p is the [unknown] root, h is the [unknown] error, and r is the [unknown] rate of convergence, $r = g'(p)$. But we have three approximate equations for these three unknowns! We can solve [approximately] for p:

$$h = x_0-p; \; rh \approx x_1-p; \; r^2h \approx x_2-p;$$

so

$$(p-x_0)\times(p-x_2) \approx (p-x_1)^2 \; [\; \approx r^2h^2\,].$$

That is,

$$p^2 - (x_0+x_2)p + x_0x_2 \approx p^2 - 2x_1p + x_1^2,$$

or

$$p \approx \frac{x_0x_2 - x_1^2}{x_0 + x_2 - 2x_1}.$$

If we take x_3 to be given by this formula instead of by $x_3 = g(x_2)$, then resume the iteration to find x_4, x_5 normally, then x_6 by this 'Aitken' formula, and so on, then, lo and behold, we get quadratic convergence.

Note that we get quadratic convergence *even if* $r > 1$, that is, even if the first-order process is diverging! [This sounds too good to be true, and sadly it often is, but it's worth trying.] Oh, something to watch is that we get serious cancellation errors if all the x_i are close together, so it's better at getting you close to the root than at finding it with great accuracy.

Final word on iterations:

Don't get hung up on doing iterations exactly. The whole point is that they [should] converge from more-or-less anywhere sensible. It doesn't matter if you make a mistake, as long as it's not too gross, and as long as you don't do it too often; blunders are self-correcting. You can work to 1dp, then 2dp, then 3, 4, 5, … as the results converge, in case you are having to type in numbers to your calculator or computer. You don't have to 'wait' for a process to converge—if you can 'guess' where it's heading [for example, if it seems to be oscillating about some value and only slowly converging], just go there and try again. It's a severely practical exercise!