Newton approximation

Newton-Raphson, or the Babylonian method

Suppose we have a first approximation to $\sqrt{N}$, call it x. We wish to find a better guess as to the value of $\sqrt{N}$. Let

$$xy = N, \qquad y = \frac{N}{x}$$

x and y ”straddle” the true value. If $x < \sqrt{N}$ then $y > \sqrt{N}$ and vice-versa.

Proof.

Factoring the equation above and rearranging, we have that

$$\frac{g}{\sqrt{N}} = \frac{\sqrt{N}}{x}$$

Suppose that $x < \sqrt{N}$ ($x^2 < N$), so then $\sqrt{N}/x > 1$ and

$$\frac{y}{\sqrt{N}} = \frac{\sqrt{N}}{x} > 1 \quad \Rightarrow \quad y > \sqrt{N}$$

while if $x > \sqrt{N}$

$$\frac{y}{\sqrt{N}} = \frac{\sqrt{N}}{x} < 1 \quad \Rightarrow \quad y < \sqrt{N}$$

The geometric mean of x and y is

$$\sqrt{x \cdot y} = \sqrt{x \cdot \frac{N}{x}} = \sqrt{N}$$

which would give us the precise value. But of course that assumes exactly what we're trying to find.

The arithmetic mean of x and y might do.

$$g = \frac{1}{2}\ (x + N/x)$$

It must be closer to $\sqrt{N}$ than at least one of x or y but it is not guaranteed to be closer than both. (Counterexample: if we have $N = 2$ and suppose $x = 0.2$ then $y = N/x = 10$. The average is 4.9 which is not as close to $1.414\ldots$ as x is.)

Newton-Raphson

My source calls this the "Babylonian method", but I've always known it as the Newton-Raphson method. It's a linear approximation.

Technically, the Newton method applies to (almost) any $f(x)$ and is written in terms of $f'(x)$, while the Babylonian method is strictly for the square root function.

A simplified derivation is as follows. We can formulate the square root problem as $y = x^2 - N$ where we want to find the positive root x such that $y = 0 = x^2 - N$ since then $x^2 = N$.

For this parabola, the slope of the tangent line at any point such as $(x, x^2 - N)$ is $2x$ (from basic calculus or inspired geometry).

Let the zero of the tangent line be at the point $(g, 0)$, then we can write the point-slope equation of the tangent line as

$$\frac{\Delta y}{\Delta x} = 2x = \frac{(x^2 - N) - 0}{x - g}$$
$$2x(x-g) = x^2 - N$$
$$2x - 2g = x - \frac{N}{x}$$
$$g = \frac{1}{2}(x + \frac{N}{x})$$

Using the tangent line as an approximation to the parabola, then the point where the tangent line crosses the y-axis is close to the zero of the parabola, that is, to $g \approx \sqrt{N}$.

As an example, 7/4 is a reasonable first approximation of $\sqrt{3}$. Then

$$x = \frac{1}{2}(\frac{7}{4} + \frac{4}{7} \cdot 3) = \frac{1}{2}(\frac{49+48}{28}) = \frac{97}{56}$$

A more sophisticated derivation is from here:

`http://www.math.ubc.ca/~anstee/math104/104newtonmethod.pdf`

It goes like this. Let r be the actual value of the zero of $f(x)$. Let x_0 be a good estimate of r, and the difference $h = r - x_0$. Linear approximation gives

$$f(r) = f(x_0 + h) \approx f(x_0) + f'(x_0) \cdot h$$

Starting from the value of the function at x_0, a small change h gives a change in the value of the function of the derivative times the small change h.

And then (provided $f'(x_0)$ is not near zero):

$$f(r) = 0 \approx f(x_0) + f'(x_0) \cdot h$$

$$h \approx -\frac{f(x_0)}{f'(x_0)}$$

so

$$r = x_0 + h \approx x_0 - \frac{f(x_0)}{f'(x_0)}$$

In the case of the square root problem, the numerator is $x_0^2 - N$ and $f'(x_0) = 2x_0$ so

$$r \approx x_0 - \frac{x_0^2 - N}{2x_0}$$

Let us call the new value x_1

$$\begin{aligned} x_1 &= x_0 - \frac{x_0^2 - N}{2x_0} \\ &= \frac{1}{2}\ (x_0 + \frac{N}{x_0}) \end{aligned}$$

secant method

There is another method called the *secant* method. In Newton's method, we need the derivative. The secant method approximates the derivative by using the slope connecting two points on the curve.

Recall that $f(x) = x^2 - N$. Suppose we have two approximations to the square root, call them x_1 and x_2. For these two points on the curve we have that the slope of the secant line connecting them is

$$\begin{aligned} m = \frac{y_2 - y_1}{x_2 - x_1} &= \frac{(x_2^2 - N) - (x_1^2 - N)}{x_2 - x_1} \\ &= \frac{x_2^2 - x_1^2}{x_2 - x_1} \end{aligned}$$

Since the numerator is a difference of squares, we see that m is just equal to $x_1 + x_2$.

Now let g also be on the same line, at the point where the line goes through the x-axis and $y = 0$. The point-slope equation (again) is

$$m = x_1 + x_2 = \frac{f(x_1) - 0}{x_1 - g} = \frac{x_1^2 - N}{x_1 - g}$$

$$x_1 - g = \frac{x_1^2 - N}{x_1 + x_2}$$

$$g = \frac{N - x_1^2}{x_1 + x_2} + x_1$$

$$= \frac{N - x_1^2 + x_1^2 + x_1 x_2}{x_1 + x_2}$$

$$= \frac{N + x_1 x_2}{x_1 + x_2}$$

As an example, suppose we use 5/3 and 7/4 as approximations to $\sqrt{3}$. Then

$$g = \frac{3 + (5/3 \cdot 7/4)}{5/3 + 7/4}$$

It's a fraction of fractions, but the denominators cancel. So

$$g = \frac{36 + 5 \cdot 7}{4 \cdot 5 + 3 \cdot 7} = \frac{71}{41}$$

The last reference above also discusses the secant method, its connection to the Newton method, and also something about what Newton actually did.