

Newton's Method

Newton's method is the most effective method for finding roots by iteration.

$$f(x) = 0$$

The method consists of the following steps:

1. Pick a point x_0 close to a root. Find the corresponding point $(x_0, f(x_0))$ on the curve.
2. Draw the tangent line to the curve at that point, and see where it crosses the x -axis.
3. The crossing point, x_1 , is your next guess. Repeat the process starting from that point.

Mathematical details

We pick a point

$$(x_0, f(x_0)) \tag{1}$$

on the curve and find the equation of the tangent line at that point. The slope of the tangent line at that point is

$$m = f'(x_0) \tag{2}$$

The tangent line has general form

$$f'(x_0) = \frac{y - f(x_0)}{x - x_0}$$

This line intersects the x -axis when $y = 0$.

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

$$f'(x_0) (x - x_0) = -f(x_0)$$

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

The expression on the right of the last equation becomes the basis for our iteration scheme. Specifically, we introduce the function

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

and note that any root of $f(x)$ is a fixed point of $g(x)$ and vice-versa. Thus we have converted the root finding problem into a fixed point finding problem that can be solved by iteration.

Newton's Method is a very good method

Like all fixed point iteration methods, Newton's method may or may not converge in the vicinity of a root. As we saw in the last lecture, the convergence of fixed point iteration methods is guaranteed only if $|g'(x)| < 1$ in some neighborhood of the root. Even Newton's method can not always guarantee that condition. When the condition is satisfied, Newton's method converges, and it also converges faster than almost any other alternative iteration scheme based on other methods of converting the original $f(x)$ to a function with a fixed point.

In order to start to get a handle on why Newton's method is unusually effective for a fixed point iteration, we start with a couple of definitions.

Definition A sequence of fixed-point iterates

$$p_n = g(p_{n-1})$$

converges linearly to a limiting value p if there exists a constant $0 < \lambda < 1$ and a positive integer N such that

$$|p_{n+1} - p| < \lambda |p_n - p|$$

for all $n > N$.

Definition A sequence of fixed-point iterates

$$p_n = g(p_{n-1})$$

converges quadratically to a limiting value p if there exists a constant $0 < \lambda$ and a positive integer N such that

$$|p_{n+1} - p| < \lambda |p_n - p|^2$$

for all $n > N$.

Both of these definitions state that the distance from p_n to p shrinks as we progress through the sequence. The shrinkage is much more dramatic in the second case due to the presence of the square term.

The fixed point theorem we saw in the last lecture is sufficient to guarantee linear convergence provided that certain simple conditions on $g(x)$ are satisfied. Unfortunately, that theorem does not guarantee quadratic convergence. For that we need something special.

Quadratic Convergence Theorem

Let p be a fixed point of a function $g(x)$. If $g'(p) = 0$ and $g''(x)$ is continuous with $|g''(x)| < M$ on an open interval $(p - \delta, p + \delta)$ any iterated sequence starting from a $p_0 \in (p - \delta, p + \delta)$ will converge quadratically to p . Moreover, for large n we will have

$$|p_{n+1} - p| < \frac{M}{2} |p_n - p|^2$$

Proof Expanding $g(x)$ in a first order Taylor polynomial about $x = p$ gives

$$g(x) = g(p) + g'(p)(x - p) + \frac{g''(\xi)}{2} (x - p)^2$$

where ξ is some point between x and p . Noting that p is a fixed point and that $g'(p) = 0$ gives

$$g(x) = p + \frac{g''(\xi)}{2} (x - p)^2$$

Substituting p_n for x and rearranging gives

$$p_{n+1} - p = \frac{g''(\xi_n)}{2} (p_n - p)^2$$

where ξ_n is a point between p and p_n . Since $g'(p) = 0$ and $g'(x)$ is continuous near p we can conclude that $|g'(x)| < 1$ for all x in some neighborhood of p . If we choose δ to make the interval $(p-\delta, p+\delta)$ fit inside that interval we can use the original fixed point theorem to conclude that the sequence of p_n points converges to p . Since the ξ_n points are trapped between p and p_n they also converge to p . Thus,

$$\frac{g''(\xi_n)}{2} < \frac{M}{2}$$

for n large enough. It follows that

$$|p_{n+1} - p| < \frac{M}{2} |p_n - p|^2$$

for n large enough and the sequence of p_n points converges quadratically to p .

Newton's Method converges quadratically

I leave it as an exercise for the reader to verify that the Newton iteration function

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

satisfies the condition that $g'(p) = 0$ at the fixed point. In cases when it also satisfies the restriction that $|g''(x)| < M$ on an open interval $(p-\delta, p+\delta)$ we have enough to guarantee quadratic convergence of the Newton's method sequence.

Newton's Method is not always applicable

Newton's method is a lovely method that you should try to apply any time you are faced with a root finding problem. Newton's method has one small flaw, though. To apply the method you have to be able to compute the derivative $f'(x)$. At first, you might think that this not such a big deal. Almost any reasonable function that one can write down can be differentiated, so the derivative step doesn't look like a problem.

The problem in practice is that functions come in many forms, and not all of these forms lend themselves to computing derivatives. Here are several different ways that functions can be defined.

1. The function is defined via a closed form formula involving elementary functions.

$$f(x) = \frac{e^x + x^2}{\sqrt{\sin x}}$$

2. The function is defined via an integral

$$f(x) = \int_1^x \frac{\sin t}{t} dt$$

3. The function is defined via a convergent power series.

$$f(x) = \sum_{n=0}^{\infty} \frac{(-1)^n}{n^2 + 1} x^n$$

4. The function is the solution to a differential equation.

$$y'' + y y' = e^x ; y(0) = 1, y'(0) = 0$$

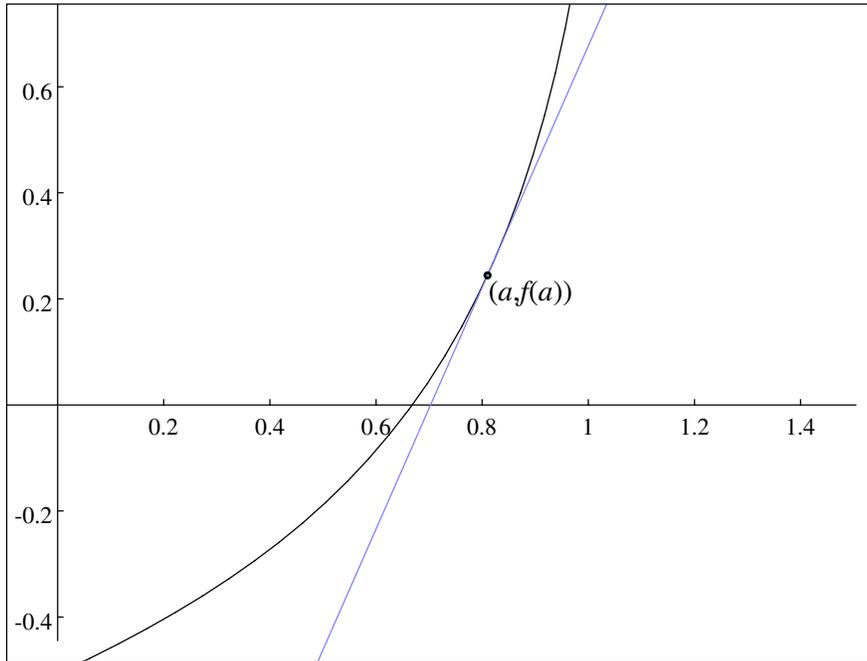
5. The function is defined recursively.

$$f(x) = \begin{cases} \sqrt{f(-2x-2) + f(x+10)} & -10 < x < 10 \\ 2 & x \leq -10 \\ -2 & x \geq 10 \end{cases}$$

Only about the first two and a half of these methods produce functions whose derivatives can be readily computed. In the absence of derivative information we can deploy some alternative algorithms.

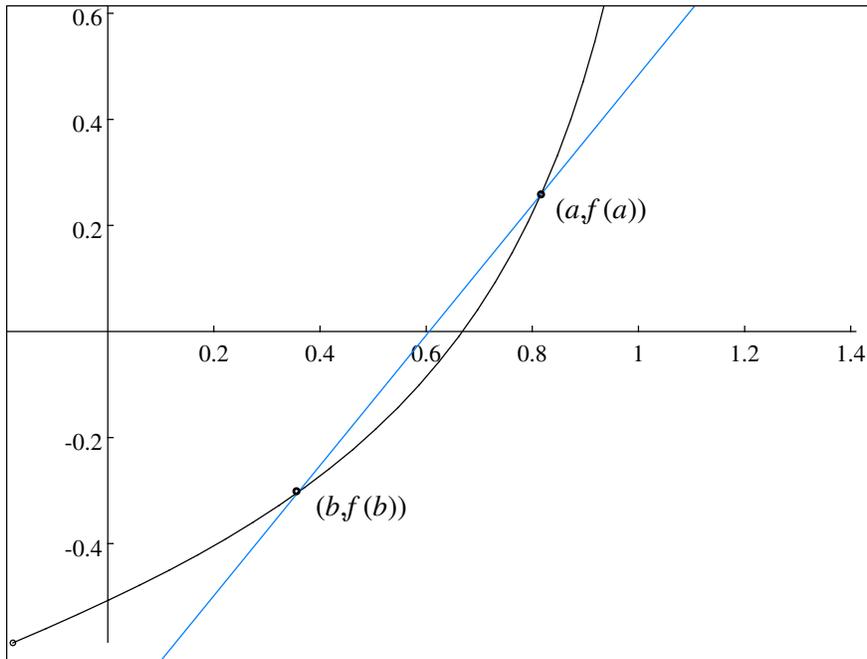
The Secant Method

Here is an obvious geometrical way to modify Newton's method. Newton's method is based on using a tangent line to find the next approximation of the root.



To compute the equation of the tangent line we need to know the derivative of the function.

An alternative to using a tangent line is to use a secant line, which touches the curve at two points.



The equation of the secant line is easy to compute:

$$\frac{y - f(a)}{x - a} = \frac{f(b) - f(a)}{b - a}$$

and it is easy to determine where this secant line will cross the x -axis.

$$\frac{0 - f(a)}{x - a} = \frac{f(b) - f(a)}{b - a}$$

$$-f(a) \frac{b - a}{f(b) - f(a)} = x - a$$

$$x = a - f(a) \frac{b - a}{f(b) - f(a)}$$

Once the secant line crosses the axis, we can replace either a or b with the crossing point and repeat the process. To make the correct replacement, we use the following logic.

1. If the root is between b and x , the product will $f(b)*f(x)$ will be negative. In that case, replace a with x .
2. If the root is between a and x , the product will $f(a)*f(x)$ will be negative. In that case, replace b with x .

Accelerating Convergence

Several alternative methods for finding fixed points are based on taking a fixed point sequence you already have and modifying the sequence to make it converge faster.

Aitken's Δ^2 method takes a linearly convergent sequence of points $\{p_n\}$ and modifies the sequence to make it converge faster. The method is based on the observation that in a linearly convergent sequence the ratio of successive distances from the limit approaches a limit as n gets larger:

$$\frac{p_{n+1} - p}{p_n - p} \approx \frac{p_{n+2} - p}{p_{n+1} - p}$$

$$p \approx \frac{p_{n+2} p_n - p_{n+1} p_{n+1}}{p_{n+2} - 2 p_{n+1} + p_n} = p_n - \frac{(p_{n+1} - p_n)^2}{(p_{n+2} - p_{n+1}) - (p_{n+1} - p_n)}$$

The latter formula can be used directly, but is more commonly rewritten by introducing the *forward difference operator*

$$\Delta p_n = p_{n+1} - p_n$$

and its *iterated form*

$$\Delta^2 p_n = \Delta(\Delta p_n) = \Delta(p_{n+1} - p_n) = (p_{n+2} - p_{n+1}) - (p_{n+1} - p_n)$$

Using this notation leads to

$$\hat{p}_n = p_n - \frac{(\Delta p_n)^2}{\Delta^2 p_n}$$

To apply the Aitken Δ^2 method we start by generating some number of terms $\{p_n\}$ in a linearly convergent

sequence. We then feed those terms through the formula above to generate a second sequence $\{\hat{p}_n\}$ which converges more quickly than the first.

Steffenson's method is a variant of the Aitken method that uses the Aitken formula to generate a better sequence directly:

1. Make a starting guess p_0 .
2. Compute $p_1 = g(p_0)$ and $p_2 = g(p_1)$.
3. Use the Aitken formula to compute \hat{p}_0 :

$$\hat{p}_0 = p_0 - \frac{(p_1 - p_0)^2}{(p_2 - p_1) - (p_1 - p_0)}$$

4. Go back to step 2 with with $p_0 = \hat{p}_0$.