Newton's Method

Tuesday, September 8, 2020 7:52 PM

Newton's method (a.k.a. Newton-Raphson method)

Fundamental, ubiquitous algorithm (partly because it extends beyond 1D root-finding to multi-dimensional root-finding and optimization)

Derivation

In equations, recall we want to solve $f(p)=0$

General technique in STEM: replace problem with a simpler approximation ... in particular, linearization

Do 1st order Taylor Series, $f(p) \approx f(p_0) + f'(p_0)\cdot(p-p_0)$, good approximation when $|p-p_0|$ small

so, solve $0 = f(p_0) + f'(p)\cdot(p-p_0)$ for p,

i.e., $p = p_0 - \dfrac{f(p_0)}{f'(p_0)}$

but this was only approximate, so do this repeatedly:

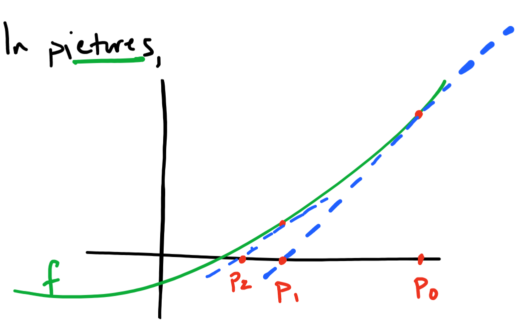
$$p_n = p_{n-1} - \frac{f(p_{n-1})}{f'(p_{n-1})}$$ Newton's Method

[Aside: next semester you'll see the multivariate version,

$$\vec{p}_n = \vec{p}_{n-1} - J(\vec{p}_{n-1})^{-1}\cdot F(\vec{p}_{n-1})$$

for solving $F(\vec{p}) = \vec{0}$, J is Jacobian of F]

In pictures,



Approximate the function $f(x)$ using its tangent line, and find a zero of the tangent line (which is easy since it's a line)

When the tangent line is a good approximation of the function, Newton's method will converge rapidly

Convergence, part 1

⚠ In general, Newton's method is not "globally convergent". That is, you can't start at any starting p_0. As we'll see later, this can be partially remedied, eg. with safeguarding and hybrid methods.

Thm 2.6 Local convergence of Newton's Method

Let $f \in C^2([a,b])$ and p is a root of f $(f(p)=0)$ inside (a,b), and p is simple root $(f'(p)\neq 0)$, then if Newton's method is initialized close enough to p, then the sequence (p_n) generated by Newton's method will converge to p

(i.e., $\exists\, \delta>0$ s.t. $(\forall\, p_0$ with $|p_0-p|<\delta)$, $p_n \to p$).

Proof

We can rewrite $\underbrace{p_{n+1} = p_n - \frac{f(p_n)}{f'(p_n)}}_{g(p_n)}$ as $p_{n+1} = g(p_n)$

i.e., we've converted root-finding $(f(p)=0)$ into a fixed-point problem, $p = g(p)$, $g(p) := p - \frac{f(p)}{f'(p)}$

$f(p)=0$ iff $-\frac{f(p)}{f'(p)} = 0$ (since we assume $f'(p)\neq 0$ at the root)

iff $p - \frac{f(p)}{f'(p)} = p$.

So, use contraction mapping theorem. We'll find an interval $x \in (p-\delta, p+\delta)$ such that
1. g maps this interval into this interval
2. g is a contraction, ie., $|g'(x)| \le k < 1$ on this interval.

First, is g well-defined? $\frac{1}{f'(x)}$ is a problem if $f'(x)=0$. We assume $f'(p)\neq 0$ at the true root p, and by continuity, there's also some region $(p-\delta_1, p+\delta_1)$ where $f'(x)\neq 0$. So g is well-defined and continuous on this region. In fact, since $f\in C^2$, $g \in C^1$ on this region.

Now, show g is a contraction, i.e., want $|g'(x)|$ small.

Well. $g'(p)=0$ in fact. To see this, use the quotient rule:

$$g'(x) = \frac{f(x)f''(x)}{(f'(x))^2} \quad \text{so} \quad g'(p) = \frac{\overbrace{\cancel{f(p)}}^{=0 \text{ since a root}} f''(p)}{(f'(p))^2} = 0$$

thus by continuity of g', there's an interval around p where $g'(x)$ is almost 0, i.e., $\forall K>0$ (in particular, choose some $K<1$) $\exists\, \delta_2$ s.t. $\forall\, x \in [p-\delta_2, p+\delta_2]$, $|g'(x)| \le K$. ✓

Let $\delta = \min(\delta_1, \delta_2)$,

$I = [p+\delta, p-\delta]$. It remains to show g maps I into I.

Let $x \in I$, then by the Mean Value Theorem, $\exists\, \xi$ between x and p such that

$$
\begin{aligned}
|g(x) - \underbrace{g(p)}_{=p}| &= |g'(\xi)| \cdot |x-p| \\
&\le K \cdot |x-p| \\
&< |x-p|
\end{aligned}
$$

so $|g(x) - p| < |x-p|$, so if $|x-p| < \delta$ then $|g(x)-p| < \delta$ so $g(x) \in I$. ✓

So, we can apply fixed-pt thm ("contraction mapping" / "Banach fixed pt.") to get convergence. $\square$

Convergence, part 2 (rate, ie., local quadratic convergence)

Helper Theorem for generic fixed-pt. iteration to solve $p = g(p)$

Thm 2.9 Let p be a solution to $x = g(x)$, and suppose $\underbrace{g'(p)=0}$ (as in Newton's method) and g'' is continuous and bounded $|g''(x)| < M$ on some open interval $(p-\delta_1, p+\delta_2)$, $\delta_1, \delta_2 > 0$. Then $\exists\, \delta > 0$ s.t. if $|p_0 - p| \le \delta$, $p_n := g(p_{n-1})$, then p_n converges quadratically to p and $\exists\, N$ st. $(\forall n \ge N)$ $|p_{n+1} - p| < \frac{M}{2} |p_n - p|^2$.

proof

As in previous theorem, close enough to p, we have $K<1$ with $|g'(x)| \le K$ near p, and g maps $[p-\delta, p+\delta]$ into $[p-\delta, p+\delta]$. (all due to continuity properties)

Now, Taylor expand around p:

$$
g(x) = \underbrace{g(p)}_{p \text{ fixed pt.}} + \underbrace{\cancel{g'(p)(x-p)}}_{g'(p)=0 \text{ by assumption}} + g''(\xi)/_2\, (x-p)^2, \quad \xi \text{ between } x \text{ and } p
$$

Choosing $x = p_n$, $p_{n+1} = g(p_n)$

so $p_{n+1} - p = g''(\xi)/_2\, (p_n - p)^2$ (or ξ_n)

In previous theorem, under these conditions, $p_n \to p$.

What about ξ? ξ_n between p and p_n, so by squeeze thm., $\xi_n \to p$ also.

Thus

$$\lim_{n\to\infty} \frac{|p_{n+1} - p|}{|p_n - p|^2} = \lim_{n\to\infty} \left| \frac{g''(\xi_n)}{2} \right| = \left| \frac{g''(p)}{2} \right| < M/2$$

which means $p_n \to p$ quadratically. $\square$

Recall our earlier discussion of simple roots

(Thm 2.11) A root p of $f \in C^1([a,b])$, $p \in (a,b)$, is called simple if $f'(p) \neq 0$.

(Careful with notation:

root finding $f(p) = 0$, then $f'(p) \neq 0$ is good since it means a simple root.

fixed-pt. $g(p) = p$, then $g'(p) = 0$ is good since it means fast convergence.)

So, putting it altogether

Thm If p is a simple root of f, then if Newton's method is initialized sufficiently close to p, it will converge, and at a quadratic rate.

Note:

For methods that can be cast as fixed-pt. iterations $p_{n+1} = g(p_n)$,

where $g(x) = x - \phi(x) f(x)$, need $\phi(p) \neq 0$,

and for quadratic convergence, need $g'(p) = 0$, which is true iff $\underbrace{\phi(p) = 1/f'(p)}$
(or any superlinear convergence)

Newton's method defines $\phi(x) = 1/f'(x)$ $\forall x$ to ensure this.