CSC721 Algorithms Fall 2017

Newton's Method

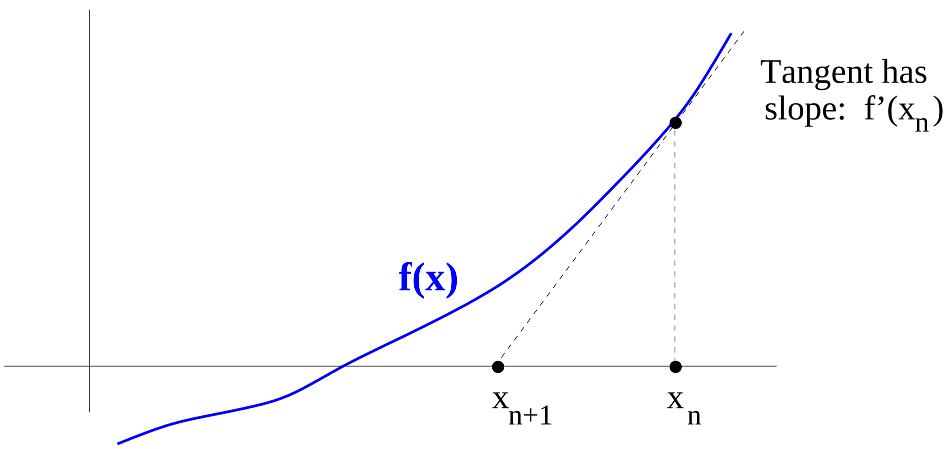
From Wikipedia

The idea of the method is as follows: one starts with an initial guess which is reasonably close to the true root, then the function is approximated by its tangent line (which can be computed using the tools of calculus), and one computes the x-intercept of this tangent line (which is easily done with elementary algebra). This x-intercept will typically be a better approximation to the function's root than the original guess, and the method can be iterated.

Suppose $f : [a, b] \to \mathcal{R}$ is a differentiable function defined on the interval $[a, b]$ with values in the real numbers. We seek to find x such that:

$$f(x) = 0$$

The formula for converging on the root can be easily derived. Suppose we have some current approximation x_n. Then we can derive the formula for a better approximation, x_{n+1} by referring to the diagram on below.



The equation of the tangent line to the curve $y = f(x)$ at the point $x = x_n$ is

$$y = f'(x_n)(x - x_n) + f(x_n)$$

where f' denotes the derivative of the function f.

The x-intercept of this line (the value of x such that $y = 0$) is then used as the next approximation to the root, x_{n+1}. In other words, setting y to zero and x to x_{n+1} gives:

$$0 = f'(x_n)\,(x_{n+1} - x_n) + f(x_n)$$

Solving for x_{n+1} yields:

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

We start the process off with some arbitrary initial value x_0. The method will usually converge, provided this initial guess is close enough to the solution, and $f'(x) \neq 0$ in a neighborhood of the solution. Furthermore, for a zero of multiplicity 1, the convergence is at least quadratic in a neighborhood of the zero. Intuitively, quadratic convergence means that the number of correct digits roughly doubles in every step.

Newton's method was first published in 1685 in *A Treatise of Algebra both Historical and Practical* by John Wallis.

Proof of quadratic convergence for Newton's iterative method

According to Taylor's theorem, any function $f(x)$ which has a continuous second derivative can be represented by an expansion about a point that is close to a root of $f(x)$. Let α denote a root, i.e., $f(\alpha) = 0$. Then the Taylor expansion of $f(\alpha)$ about x_n is:

$$f(\alpha) \;=\; f(x_n) \;+\; f'(x_n)(\alpha - x_n) \;+\; R_1 \tag{1}$$

where the Lagrange form of the Taylor series remainder is

$$R_1 = \frac{1}{2!} f''(\xi_n)(\alpha - x_n)^2 \qquad \text{and } \xi_n \text{ is between } x_n \text{ and } \alpha.$$

Since α is the root, equation (1) becomes:

$$0 \;=\; f(\alpha) \;=\; f(x_n) \;+\; f'(x_n)(\alpha - x_n) \;+\; \frac{1}{2} f''(\xi_n)(\alpha - x_n)^2 \tag{2}$$

Dividing equation (2) by $f'(x_n)$ and rearranging gives:

$$\frac{f(x_n)}{f'(x_n)} \;+\; (\alpha - x_n) \;=\; \frac{-f''(\xi_n)}{2f'(x_n)}\,(\alpha - x_n)^2 \tag{3}$$

Remembering that x_{n+1} is defined by

$$x_{n+1} \;=\; x_n - \frac{f(x_n)}{f'(x_n)} \tag{4}$$

one finds that:

$$\underbrace{\alpha - x_{n+1}}_{\varepsilon_{n+1}} = \frac{-f''(\xi_n)}{2f'(x_n)}(\underbrace{\alpha - x_n}_{\varepsilon_n})^2$$

That is:

$$\varepsilon_{n+1} \;=\; \frac{-f''(\xi_n)}{2f'(x_n)}\,{\varepsilon_n}^2\,. \tag{5}$$

Taking absolute value of both sides gives:

$$|\varepsilon_{n+1}| \;=\; \frac{|f''(\xi_n)|}{2\,|f'(x_n)|}\,{\varepsilon_n}^2 \tag{6}$$

Equation (6) shows that the rate of convergence is quadratic if the following conditions are satisfied:

1. $f'(x) \neq 0$; for all $x \in I$, where I is an interval $[\alpha r, \alpha + r]$ for some $r \geq |\alpha x)0|$;
2. $f'(x)$ is continuous, for all $x \in I$;
3. x_0 is sufficiently close to the root α.

Finally, equation (6) can be expressed in the following way:

$$|\varepsilon_{n+1}| \leq M{\varepsilon_n}^2$$

where M is the supremum of the variable coefficient of ξ_n^2 on the interval I defined in condition 1, that is:

$$M = \sup_{x \in I} \frac{1}{2} \left| \frac{f''(x)}{f'(x)} \right|$$

Multi-Variate Newton's Method

Let $\mathbf{x}$ denote an $n \times 1$ column vector and let $\mathbf{f} : \mathcal{R}^n \to \mathcal{R}^n$ be a continuous differentiable function. We seek to solve

$$\mathbf{f}(\mathbf{x}) = \mathbf{0}$$

The n-dimensional Newton's iteration is defined by:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - J(\mathbf{x}_i)^{-1} f(\mathbf{x}_i) \tag{7}$$

where $J(\mathbf{x})$ is the Jacobian matrix for $f(\mathbf{x})$.

The formula above can be derived analogous to the 1-D derivation using multi-variate Taylor theorem.

Multi-variate Newton's method can also be used for optimization problems involving a functional $K : \mathcal{R}^n \to R$. I.e., the problem is to find $\hat{\mathbf{x}}$ such that:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \left\{ K(\mathbf{x}) \right\}$$

Using standard calculus techniques, we find a critical point by solving $\nabla K(\mathbf{x}) = \mathbf{0}$. We introduce the notation for the vector $\mathbf{x}$:

$$\mathbf{x} = [x_1, x_2, x_3, ..., x_n]^T$$

Recall the gradient is given by:

$$\nabla K(x_1, x_2, x_3, ..., x_n) = \begin{bmatrix} \frac{\partial K(x_1, x_2, x_3, ..., x_n)}{\partial x_1} \\ \frac{\partial K(x_1, x_2, x_3, ..., x_n)}{\partial x_2} \\ \vdots \\ \frac{\partial K(x_1, x_2, x_3, ..., x_n)}{\partial x_n} \end{bmatrix} \tag{8}$$

Observe from equation (8) that ∇K can be considered a vector-valued function. I.e., $\nabla K : \mathcal{R}^n \to \mathcal{R}^n$. Therefore, we can solve $\nabla K(\mathbf{x}) = \mathbf{0}$ using multi-variate Newton's method. The iteration becomes:

$$\mathbf{x}_{i+1} = \mathbf{x}_i - H(\mathbf{x}_i)^{-1} \nabla K(\mathbf{x}_i) \tag{9}$$

where $H(\mathbf{x})$ is the Hessian matrix of $K(\mathbf{x})$. I.e., the matrix of second-order partial derivatives:

$$H_{i,j}(x_1, x_2, x_3, ..., x_n) = \frac{\partial^2 K(x_1, x_2, x_3, ..., x_n)}{\partial x_i \partial x_j}$$