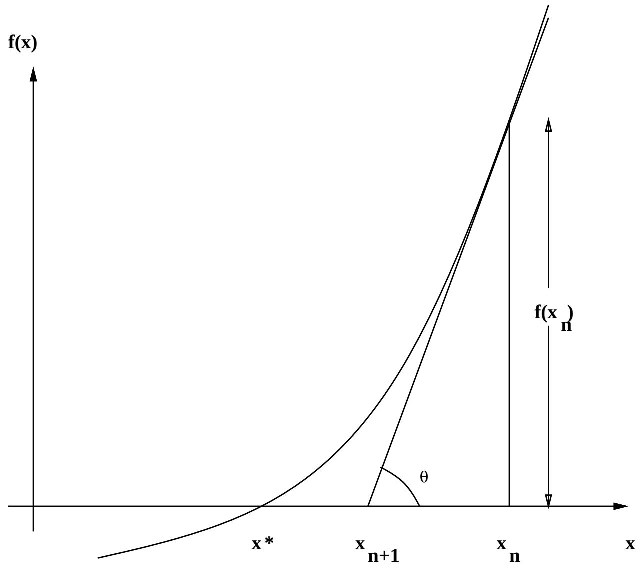
Mathematics 16: Solving Algebraic Equations: Newton-Raphson

The *Newton-Raphson* Algorithm is one of the most effective ways of solving algebraic equations. The problem we require to solve may be reduced to the finding of solutions to $f(x) = 0$, or in higher dimensions $\mathbf{f}(\mathbf{x}) = \mathbf{0}$ both function and variable being vectors. The idea is readily depicted:



and we can immediately see, that based upon the function value and slope at a point x_n, a better estimate for where the function $f(x)$ might vanish, can be obtained using:

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

which constitutes the Newton-Raphson Algorithm in one-dimension.

A more mathematical derivation comes from Taylor's Theorem, from which we find:

$$0 = f(x^*) = f(x_n + h) = f(x_n) + f'(x_n)h + \frac{1}{2}f''(x_n)h^2 + O(h^3)$$

where $f'(x) = \frac{df}{dx}$ and $f''(x) = \frac{d^2f}{dx^2}$. The derived estimate for x^* is therefore:

$$h = -\frac{f(x_n)}{f'(x_n)} - \frac{1}{2}\frac{f''(x_n)}{f'(x_n)}h^2 + O(h^3)$$

and so:

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)} = x_n + h + \frac{1}{2}\frac{f''(x_n)}{f'(x_n)}h^2 + O(h^3) = x^* + \frac{1}{2}\frac{f''(x_n)}{f'(x_n)}h^2 + O(h^3)$$

and if h is small, an error of order h is replaced by an error of order h^2. This algorithm is said to be *quadratically convergent.* This convergence is very impressive! The improvement in accuracy is a *doubling* of the number of correct decimal places.

Extensions into higher-dimensions are straightforward, but we have to be very aware of the Tensor nature of the objects we are dealing with:

$$\mathbf{0} = \mathbf{f}(\mathbf{x}^*) = \mathbf{f}(\mathbf{x}_n + \mathbf{h}) = \mathbf{f}(\mathbf{x}_n) + (\mathbf{h}.\nabla)\mathbf{f}(\mathbf{x}_n) + \frac{1}{2}(\mathbf{h}.\nabla)^2\mathbf{f}(\mathbf{x}_n) + O(\mid \mathbf{h} \mid^3)$$

the natural analogue from high-dimensional Taylor's Theorem. We may use the same formal algebra, to provide:

$$\mathbf{x}_{n+1} = \mathbf{x}_n - [\nabla\mathbf{f}(\mathbf{x}_n)]^{-1}\,\mathbf{f}(\mathbf{x}_n) + O(\mid \mathbf{h} \mid^2)$$

but we must be aware that the object $[\nabla\mathbf{f}(\mathbf{x}]$ is a matrix with components:

$$[\nabla\mathbf{f}(\mathbf{x})]_{ij} = \frac{\partial f_i}{\partial x_j}(\mathbf{x}_n)$$

In two dimensions we seek solutions to two equations:

$$f_1(x_1, x_2) = 0 \qquad\qquad f_2(x_1, x_2) = 0$$

and the matrix we need to employ is:

$$[\nabla\mathbf{f}(\mathbf{x})]_{ij} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \frac{\partial f_1}{\partial x_2} \\ \frac{\partial f_2}{\partial x_1} & \frac{\partial f_2}{\partial x_2} \end{bmatrix}$$

leading to the algorithm:

$$x_1 \mapsto x_1 - \frac{\frac{\partial f_2}{\partial x_2} f_1 - \frac{\partial f_1}{\partial x_2} f_2}{\frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_2} - \frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_1}}$$

$$x_2 \mapsto x_2 - \frac{\frac{\partial f_1}{\partial x_1} f_2 - \frac{\partial f_2}{\partial x_1} f_1}{\frac{\partial f_1}{\partial x_1}\frac{\partial f_2}{\partial x_2} - \frac{\partial f_1}{\partial x_2}\frac{\partial f_2}{\partial x_1}}$$