Newton's Method

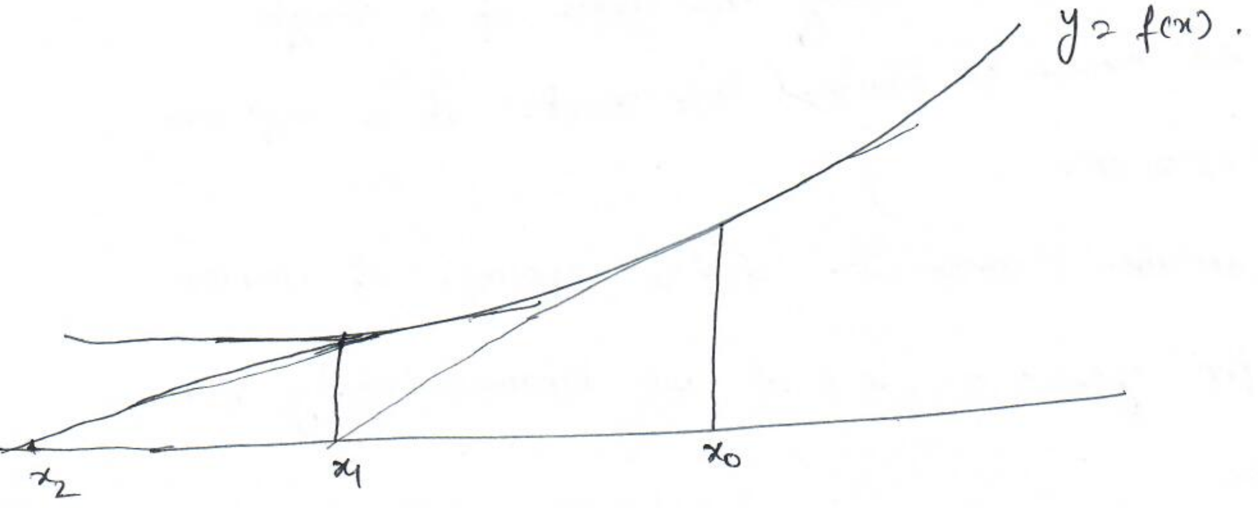
Newton's Method for finding real roots of the eqn.

$g(x) = 0$, $x \in \mathbb{R}$ is well known.

The basic scheme here is

$$y_{k+1} = y_k - \frac{g(y_k)}{g'(y_k)},$$ where y_k is the

current iterate or the current approximation.



Let us assume our problem is on unconstrained non-linear optimization problem of the form

$$\text{Min } f(x)$$
$$\text{s.t. } x \in \mathbb{R}^n$$

We are aiming at finding a point $x^* \in \mathbb{R}^n$ such that $\nabla f(x^*) = 0$. Therefore, the basic problem of root finding enters here very naturally except that rather than finding the roots of a single equation we have to find the roots of a system, i.e. $\nabla f(x) = 0$.

Looking at the standard basic scheme of Newton's method for $y(x) = 0$, $x \in \mathbb{R}$, we immediately get the scheme

$$x^{(k+1)} = x^{(k)} - [H_f(x^{(k)})]^{-1} \nabla f(x^{(k)}),$$

for finding a solution of the system $\nabla f(x) = 0$.

Another way to see this expression as

Let us approximate the given function f

(Note that, $f: \mathbb{R}^n \longrightarrow \mathbb{R}$ is the given function in problem

$$\begin{array}{l} \text{Min } f(x) \\ \text{s.t. } x \in \mathbb{R}^n. \end{array}$$

which is to be optimized.)

in a neighborhood of the current approximate $x^{(K)}$ by the truncated Taylor series to get

$$f(x) \simeq f(x^{(K)}) + (x - x^{(K)}) \nabla f(x^{(K)}) + \tfrac{1}{2}(x - x^{(K)})^T H_f(x^{(K)}) (x - x^{(K)}).$$

Therefore, if we wish to minimize $f(x)$, it makes sense to minimize its quadratic approximation $q(x)$ where

$$q(x) = f(x^{(K)}) + (x - x^{(K)}) \nabla f(x^{(K)}) + \tfrac{1}{2}(x - (x^{(K)})^T H_f(x^{(K)}) \times (x - x^{(K)})$$

Let this minimization be done exactly and hence

$$\nabla q(x) = 0$$

i.e. $\nabla f(x^{(K)}) + \frac{1}{2}(x - x^{(K)}) . 2 . H_f(x^{(K)}) = 0$

i.e. $\boxed{x^{(K+1)} = x^{(K)} - (H_f(x^{(K)}))^{-1} \nabla f(x^{(K)}).}$

Remark:- Since Newton's method for solving NLP essentially finds the roots of the system $\nabla f(x) = 0$, it follows that for minimizing a positive definite quadratic form of n variables, it will take exactly one iteration. It is like finding roots of a linear eqn. by Newton's ~~eq~~ method.

Remark:- Newton's Method ~~has~~ has order of convergence $p = 2$. (quadratic rate of convergence)

Remark:- If we are minimizing a positive definite quadratic form by Newton's Method, then not only we can start from any arbitrary point $x^{(0)} \in \mathbb{R}^n$, we also know that we will get the optimal solution in exactly one iteration. i.e. $x^{(1)}$ has to be the optimal solution $x^* = x^{(1)}$.

However, if the function $f(x)$ is not positive definite quadratic form then there are major problems with Newton's Method. In this situatation we can not start from an arbitrary point $x^{(0)}$. $x^{(0)}$ must close to x^*.

It may also be reasonable to assume that $H_f(x^*)$ is invertable in a nbd of x^*.

<u>Example</u>:- Use Newton's method to minimize

$$f(x_1, x_2) = 8x_1^2 - 4x_1x_2 + 5x_2^2, \quad (x_1, x_2) \in \mathbb{R}^2.$$

$\Rightarrow$ As the function f is a positive definite quadratic form ~~in two~~ in two variables, we know for certain that we can start from any arbitrary point $x^{(0)} \in \mathbb{R}^2$ and use Newton's method to get $x^{(1)}$, then $x^{(1)}$ has to be the minimizing point.

Let us assume $x^{(0)} = (5, 2)^T$.

Then

$$\nabla f(x^{(0)}) = \begin{pmatrix} 16x_1 - 4x_2 \\ -4x_1 + 10x_2 \end{pmatrix}\Bigg|_{x = x^{(0)}}$$

$$= (72, 0)$$

$$H_f(x^{(0)}) = \begin{pmatrix} 16 & -4 \\ -4 & 10 \end{pmatrix}$$

and $$H_f^{-1}(x^{(0)}) = \frac{1}{144}\begin{pmatrix} 10 & 4 \\ 4 & 16 \end{pmatrix}$$

$$\therefore \quad x^{(1)} = x^{(0)} - (H_f(x^{(0)}))^{-1} \nabla f(x^{(0)})$$

$$= \begin{bmatrix} 5 \\ 2 \end{bmatrix} - \frac{1}{144}\begin{pmatrix} 10 & 4 \\ 4 & 16 \end{pmatrix}\begin{pmatrix} 72 \\ 0 \end{pmatrix}$$

$$= \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

Therefore, $x^* = (0, 0)$ is the minimizing point.

Modified Newton's Method

While discussing Newton's method in the last method we noted certain limitations as $H_f(x^{(K)})$ may not be invertible at the point $x^{(K)}$.

For this we check that $-(M_K(x^{(K)}))^{-1}\nabla f(x^{(K)})$ is always a direction of descent for any positive definite matrix M_K. Another difficulty with Newton's Method has been its lack of global convergence property. Keeping these things in mind, the following modification to Newton's method is suggested

$$x^{(K+1)} = x^{(K)} - \bar{\alpha}_K M_K (\nabla f(x^{(K)})) \quad \text{———} \quad (*)$$

where M_K is an appropriate positive definite matrix (obtained from $H_f(x^{(K)})$ as ~~e~~ and $\bar{\alpha}_K > 0$ is the step size which is choose as in the steepest descent method, i.e. $\bar{\alpha}_K > 0$ is chosen such that

$$h(\bar{\alpha}_K) = \min_{\alpha_K > 0} h(\alpha_K)$$

where $h(\alpha_K) = f(x^{(K)} + \alpha_K S^{(K)})$, $S^{(K)} = -M_K(\nabla f(x^{(K)}))$

How should we chose M_K given the matrix $H_f(x^{(k)})$?

$\Rightarrow$ We remember that $H_f(x^{(k)})$ may not be invertible at $x^{(k)}$. and if at some point $x^{(k)}$ it is invertible it is not necessary that its inverse is positive definite. But $H_f(x^{(k)})$ is certainly real symmetric and hence all its eigen values are real. What we shall do now is to add a suitable matrix of the form $\varepsilon_K I$ $(\varepsilon_K > 0)$ and take $F_K = (\varepsilon_K I + H_f(x^{(k)}))$ s.t. $M_K = (F_K)^{-1}$.

Since $\varepsilon_K > 0$ is to be chosen so that all eigenvalues of F_K become strictly positive and therefore F_K and M_K become positive definite.

- ε_K be the smallest non-negative constant for which all eigenvalues of the matrix $\varepsilon_K I + H_f(x^{(k)})$ are greater than or equal to δ (pre-fixed constant)

- This method has property of global convergence and its order of convergence $p = 2$.

- However, it is still not practical because to get M_K we need to compute all eigen values of $H_f(x^{(k)})$.

- The steepest descent method is suggested to solve MLP.