Newton's method

The idea behind this method is as follows:

Given a starting point, we construct a quadratic approximation to the cost function that matches the 1st and 2nd derivative values at that point. we then minimize the quadratic (approximation) function and repeat the procedure.

Let $f: \mathbb{R}^n \to \mathbb{R}$ be the cost function which is twice continuously differentiable. Taylor expansion around the point $x^{(k)}$ is

$$f(x^k) + (x - x^k)^T \nabla f(x^k) + \frac{1}{2}(x - x^k) H(x^k)(x - x^k) = q(x)$$

$$f(x) \simeq f(x^k) + (x - x^k)^T \nabla f(x^k) + \frac{1}{2}(x - x^k) H(x^k)(x - x^k)^T = q(x)$$

Let $x^* \in \mathbb{R}^n$ be the local minimum point.

Then $\nabla q(x^*) = 0$

Applying FONC to $q(x)$ we get

$$0 = \nabla q(x) = \nabla f(x^k) + H(x^k)(x - x^k)$$

$$\Rightarrow x = x^k - [H(x^k)]^{-1} \nabla f(x^k)$$

So it is of the form $\boxed{x^{k+1} = x^k - [H(x^k)]^{-1} \nabla f(x^k)}$

Newton's Algorithm

① Initialize the tolerance ε and the initial point $x^{(0)}$. Set $k = 0$

② If $\|\nabla f(x^k)\| > \varepsilon$, then

ⓐ Solve $\boxed{H(x^k)\, d^k = -\nabla f(x^k)}$ for d^k

or $\boxed{d^k = -\left[[H(x^k)]^{-1} \nabla f(x^k) \right]}$

ⓑ $\boxed{\alpha_k = 1}$

ⓒ $\boxed{x^{k+1} = x^k + \alpha_k d^k}$

ⓓ $k = (k+1)$

end if output $x^* = x^k$.

Drawback

① Evalution of $F(x^k)$ for $\boxed{\text{large } n}$ can be computationally $\boxed{\text{expansive.}}$

⑪ $\boxed{x^{(0)}}$ is $\boxed{\text{require}}$ to be $\boxed{\text{very close}}$ to the solution x^*.

Note Despite these drawbacks, Newton's method has superior convergence property when starting point $x^{(0)}$ is closer to x^*.

The convergence analysis of Newton's method when the cost function $f(x)$ is quadratic

If the cost function is quadratic, then the Newton's method yields the true minimizer in one step. That is the Newton's method reaches the point x^* such that $\boxed{\nabla f(x^*) = 0}$ in just one step.

Let $Q = Q^T$ be invertible and $\boxed{f(x) = \frac{1}{2} x^T Q(x) - x^T b}$

$\left[Q = \begin{bmatrix} a_1 & b_1 \\ b_1 & c_1 \end{bmatrix} \right.$

$b = [b' \; b'']$

Then $\nabla f(x) = Qx - b$

$\nabla f(x^*) = 0 \implies \boxed{x^* = Q^{-1} b}$

Now $H(x) = \nabla^2 f(x) = Q$

Hence, for given any initial point $x(0)$ we get by Newton's method

$$\boxed{x^{(1)}} = x^{(0)} + d^1$$

$$= x^{(0)} - [H(x^0)]^{-1} \nabla f(x^0)$$

$$= x^{(0)} - Q^{-1} [Q x^{(0)} - b]$$

$$= x^{(0)} - x^{(0)} + Q^{-1} b$$

$$= \boxed{x^*}$$

Problem 1 Let $f(x) = 7x - \ln(x)$, $x > 0$

Find the iterative seq^n $\{x^k\}$ by Newton's method. by taking initial points $x^{(0)} = 1, 0.0, 0.1$ and 0.01.

$\square$ $\dfrac{df}{dx} = 7 - \dfrac{1}{x}$ and $\dfrac{d^2f}{dx^2} = \dfrac{1}{x^2} > 0$ at $x = \dfrac{1}{7}$

$\therefore$ at $x = \dfrac{1}{7}$, the functions attains its local minima.

Newton's will generate the iterative seq^n $\{x^k\}$ where $x^{k+1} = 2x^k - 7(x^k)^2$

For the starting point $x^{(0)} = 0$, all the iterates are coming 0. For the starting point $x^{(0)} = 1$, the method diverges to $-\infty$.

For the starting point $x^{(0)} = 0.1$, the Newton

$$x^{k+1} = x^k - \frac{\left(7 - \frac{1}{x^k}\right)}{\left(\frac{1}{x^k}\right)^2}$$

$$= x^k - 7x^{k^2} + x^k$$

$$= 2x^k - 7x^{k^2}$$

$x^{(0)} = 0.1$, the method converges to the sol^n after third iteration.

Conjugate Direction method

The class of conjugate direction method can be viewed as intermediate between the Steepest descent and Newton method. In fact, it has the following properties.

1. solve the $\boxed{\text{quadratic function}}$ of n variable in $\boxed{n \text{ steps}}$

2. The implementation of this method does $\boxed{\text{not require}}$ calculating the $\boxed{\text{Hessian matrix.}}$

3. $\boxed{\text{No matrix inversion}}$ and no storage of the matrix is required.

Note It is applicable to the quadratic cost function. $\boxed{f(x) = \frac{1}{2} x Q x^T - x^T b,}$

$$x \in \mathbb{R}^n, \quad Q = Q^T > 0$$

Defn Let Q be a symmetric positive $\boxed{\text{semi definite}}$ matrix of $\boxed{\text{order } n.}$ The directions $d^0, d^1, \ldots d^{n-1} \in \mathbb{R}^n$ are said to be $\boxed{Q \text{ conjugate}}$ if for all $i \neq j$, we have

$$\boxed{(d^i)^T Q d^j = 0}$$

Conjugate Direction algorithm

We now present the conjugate direction algorithm for minimizing the quadratic function $\boxed{f(x) = \frac{1}{2} x^T Q x - x^T b}$, $x \in \mathbb{R}^n$ and $Q = Q^T > 0$

Algorithm

Given a starting point $x^{(0)}$ and Q conjugate directions $d^0, d^1, \ldots d^{n-1}$

(i) calculate $\boxed{g^k = \nabla f(x^k) = Qx^k - b}$, $k \geq 0$

(ii) $$\boxed{\alpha_k = -\frac{g^k d^k}{(d^k)^T Q d^k}}$$

(iii) $$\boxed{x^{k+1} = x^k + \alpha_k d^k}$$

Theorem For any point $x^{(0)}$, the conjugate direction algorithm converges to x^* in $\boxed{n \text{ steps}}$, that is $\boxed{x^{(n)} = x^*}$.

Prob 1 Find the minimum point of

$$f(x_1, x_2) = \frac{1}{2}[x_1 \; x_2]\begin{bmatrix}4 & 2\\ 2 & 2\end{bmatrix}\begin{bmatrix}x_1\\ x_2\end{bmatrix} - [x_1 \; x_2]\begin{bmatrix}-1\\ 1\end{bmatrix}$$

using conjugate direction method with initial point $x^0 = \begin{bmatrix}0\\0\end{bmatrix}$ and Q-conjugate directions are $d^0 = \begin{bmatrix}1\\0\end{bmatrix}$ and $d^{(1)} = \begin{bmatrix}-\frac{3}{8}\\ \frac{3}{4}\end{bmatrix}$

$$\alpha_0 = \frac{-g^0 d^{(0)}}{(d^0)^T Q d^0}$$

$$g^{(0)} = \nabla f(x^0)$$

$$= Q(x^0) - b$$

$$g^{(0)} = \begin{bmatrix}4 & 2\\ 2 & 2\end{bmatrix}\begin{bmatrix}0\\0\end{bmatrix} - \begin{bmatrix}-1\\1\end{bmatrix}$$

$$g^{(0)} = \begin{bmatrix}1\\-1\end{bmatrix}$$

$$g^0 d^{(0)} = [1 \; -1]\begin{bmatrix}1\\0\end{bmatrix} = 1$$

$$\alpha_0 = \frac{-1}{[1 \; 0]\begin{bmatrix}4\\2\end{bmatrix}} = \frac{-1}{4}$$

$$\left[\begin{pmatrix}4 & 2\\ 2 & 2\end{pmatrix}\begin{bmatrix}1\\0\end{bmatrix}\right] = \frac{4}{2}$$

$$\alpha_0 = -\frac{1}{4}$$

$$x^{(1)} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \frac{1}{4} \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$= \begin{bmatrix} -\frac{1}{4} \\ 0 \end{bmatrix}$$

$x^{(2)}$??