Newton's method

It approximates f(x) with a quadratic function in the neiborhood of the current point using the Taylor-series expansion of f

then optimizes the approximated quadratic function to obtain the new iterate point.

As in the single-variable case the optimality conditions can be derived from the Taylor-series expansion

$$f(\boldsymbol{x}_k + \Delta x) \approx f(\boldsymbol{x}_k) + \nabla f(\boldsymbol{x}_k)\ \Delta x + \Delta x\ H(\boldsymbol{x}_k)\ \Delta x$$

Note that $\boldsymbol{x}_k$ is the known current point (therefore, also $\nabla f(\boldsymbol{x}_k)$ and $H(\boldsymbol{x}_k)$ are known.

The objective is now to determine Δx which optimizes $f(\boldsymbol{x}_k + \Delta x)$. Then we solve:

$$\frac{\partial f(x_k+\Delta x)}{\partial \Delta x} = 0$$

Newton's method

$$\frac{\partial f(x_k+\Delta x)}{\partial \Delta x} = 0 \qquad \Rightarrow \qquad H(\boldsymbol{x}_k)\,\Delta x = -\,\nabla f(\boldsymbol{x}_k)$$

$$\Delta x = -\,H(\boldsymbol{x}_k)^{-1}\,\nabla f(\boldsymbol{x}_k)$$

Newton step, it moves to a stationary point of the second order approximation derived from the Taylor-series expansion

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H(\boldsymbol{x}_k)^{-1}\,\nabla f(\boldsymbol{x}_k)$$

Newton's method

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H(\boldsymbol{x}_k)^{-1}\nabla f(\boldsymbol{x}_k)$$

If $H(\boldsymbol{x}_k)$ is definite positive than only one iteration is required for a quadratic function to reach the optimum point, from any starting point

Positive definite. Matrix $\boldsymbol{A}$ is said to be positive definite if its quadratic form $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ is positive for any $\boldsymbol{x} \neq \mathbf{0}$.

Newton's Method Steps

1. K=0
2. Choose a starting point, $\boldsymbol{x}_k$
3. Calculate $\nabla f(\boldsymbol{x}_k)$ and $H(\boldsymbol{x}_k)$
4. Calculate the next $\boldsymbol{x}_{k+1}$ using the equation

$$\boldsymbol{x}_{k+1} = \boldsymbol{x}_k - H(\boldsymbol{x}_k)^{-1} \nabla f(\boldsymbol{x}_k)$$

5. Use either of the convergence criteria discussed earlier to determine convergence. If it hasn't converged, return to step 2.

Comments on Newton's Method

- We can see that unlike the gradient descend, Newton's method uses both the gradient and the Hessian

- This usually reduces the number of iterations needed, but increases the computation needed for each iteration

- So, for very complex functions, a simpler method is usually faster

Newton's Method Example

For an example, we will use the same problem as before:

Minimize $f(x_1, x_2, x_3) = (x_1)^2 + x_1(1 - x_2) + (x_2)^2 - x_2x_3 + (x_3)^2 + x_3$

$$\nabla f(\mathbf{x}) = \begin{bmatrix} 2x_1 - x_2 + 1 & -x_1 + 2x_2 - x_3 & -x_2 + 2x_3 + 1 \end{bmatrix}$$

Newton's Method Example

The Hessian is:

$$H(x_k) = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}$$

And we will need the inverse of the Hessian:

$$H(x_k)^{-1} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}^{-1} = \begin{bmatrix} 3/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 3/4 \end{bmatrix}$$

Newton's Method Example

So, pick $x_0 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$

Calculate the gradient for the 1st iteration:

$$\nabla f(\boldsymbol{x}_0) = \begin{bmatrix} 0-0+1 & -0+0-0 & -0+0+1 \end{bmatrix}$$

$$\Rightarrow \nabla f(\boldsymbol{x}_0) = \begin{bmatrix} 1 & 0 & 1 \end{bmatrix}$$

Newton's Method Example

So, the new $\mathbf{x}$ is:

$$\boldsymbol{x}_1 = \boldsymbol{x}_0 - H(\boldsymbol{x}_0)^{-1}\,\nabla f(\boldsymbol{x}_0)$$

$$= \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 3/4 & 1/2 & 1/4 \\ 1/2 & 1 & 1/2 \\ 1/4 & 1/2 & 3/4 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}$$

$$\boldsymbol{x}_1 = \begin{bmatrix} -1 \\ -1 \\ -1 \end{bmatrix}$$

Newton's Method Example

Now calculate the new gradient:

$$\nabla f(\boldsymbol{x}_1) = \begin{bmatrix} -2+1+1 & 1-2+1 & 1-2+1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \end{bmatrix}$$

Since the gradient is zero, the method has converged

Comments on Example

- Because it uses the 2nd derivative, Newton's Method models quadratic functions exactly and can find the optimum point in one iteration.

- If the function had been a higher order, the Hessian would not have been constant and it would have been much more work to calculate the Hessian and take the inverse for each iteration.