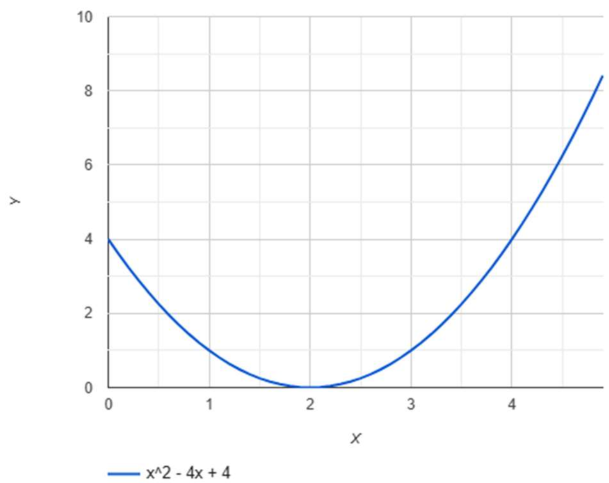
Newton’s method-case study:

Task:

- find the global minimum of the non-linear quadratic function
- Newton’s method converge faster than gradient descent
- takes larger steps when function is shallow and smaller when function is steep

$$f(x) = x^2 - 4x + 4 \tag{1}$$



Newton’s method formula:

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

$$x_{n+1} = x_n - H(x_n)^{-1} . \nabla f(x_n) \tag{2}$$

Hessian matrix:

- matrix with second order derivatives
- Hessian matrix is always square
- example of Hessian matrix of the function **f(x)** with 2 variables **x**

$$H(x_n) = \begin{bmatrix} \boldsymbol{\frac{\partial^2 f(x)}{\partial x_1^2}} & \frac{\partial^2 f(x)}{\partial x_1 x_2} \\ \frac{\partial^2 f(x)}{\partial x_2 x_1} & \boldsymbol{\frac{\partial^2 f(x)}{\partial x_2^2}} \end{bmatrix} \tag{3}$$

- example of inversion Hessian matrix:

$$H(x_n)^{-1} = \begin{bmatrix} \boldsymbol{\frac{1}{\left(\frac{\partial^2 f(x)}{\partial x_1^2}\right)}} & \frac{1}{\left(\frac{\partial^2 f(x)}{\partial x_1 x_2}\right)} \\ \frac{1}{\left(\frac{\partial^2 f(x)}{\partial x_2 x_1}\right)} & \boldsymbol{\frac{1}{\left(\frac{\partial^2 f(x)}{\partial x_2^2}\right)}} \end{bmatrix} \tag{4}$$

Compute the Gradient and Hessian:

- calculate gradient

$$\nabla f(x) = \frac{\partial f(x)}{\partial x} = \frac{\partial(x^2 - 4x + 4)}{\partial x} = 2x - 4 \tag{5}$$

- calculate Hessian matrix

$$H(x) = \left[\frac{\partial^2 f(x)}{\partial x^2}\right] = \left[\frac{\partial^2(x^2 - 4x + 4)}{\partial x^2}\right] = 2 \tag{6}$$

- calculate inverse Hessian matrix

$$H(x)^{-1} = \left[\frac{1}{\left(\frac{\partial^2 f(x)}{\partial x^2}\right)}\right] = \frac{1}{2} \tag{7}$$

Set up the Newton's method equation:

- calculate Hessian matrix

$$x_{n+1} = x_n - H(x_n)^{-1}.\nabla f(x_n)$$

$$x_{n+1} = x_n - \frac{1}{2}(2x_n - 4) \tag{8}$$

- initialize with x_n=0

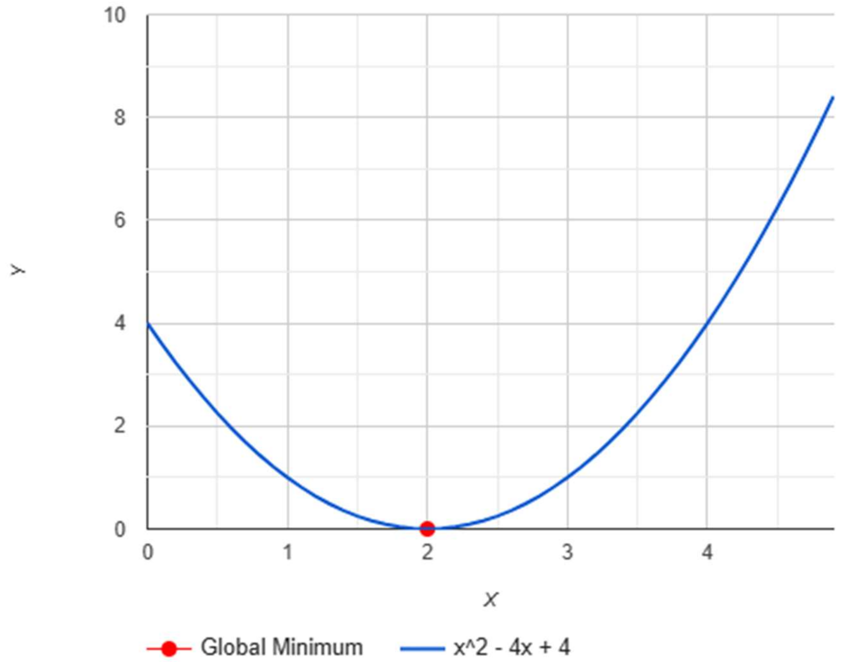
$$x_1 = x_0 - H(x_0)^{-1}.\nabla f(x_0)$$

$$x_1 = 0 - \frac{1}{2}(2 * 0 - 4) \tag{9}$$

$$x_1 = 2$$

Conclusion:

- we can see that after one iteration we found the solution – after one iteration thanks to the second derivatives
- for quadratic functions Newton's method is faster than Gradient descent



10
8
6
4
2
0
y
0
1
2
3
4
x
Global Minimum
x^2 - 4x + 4