



مجموعة جامعة
الجميع الذكية
للتعليم القابضة

MIDOCEAN
UNIVERSITY
BY EVERYONE'S SMART UNIVERSITY

Mathematical Foundations For Data Science and AI

Lecture 06

AI102

Calculus; Optimization in Neural Networks – Part 2

By
Dr. Khaled ELMenshawy

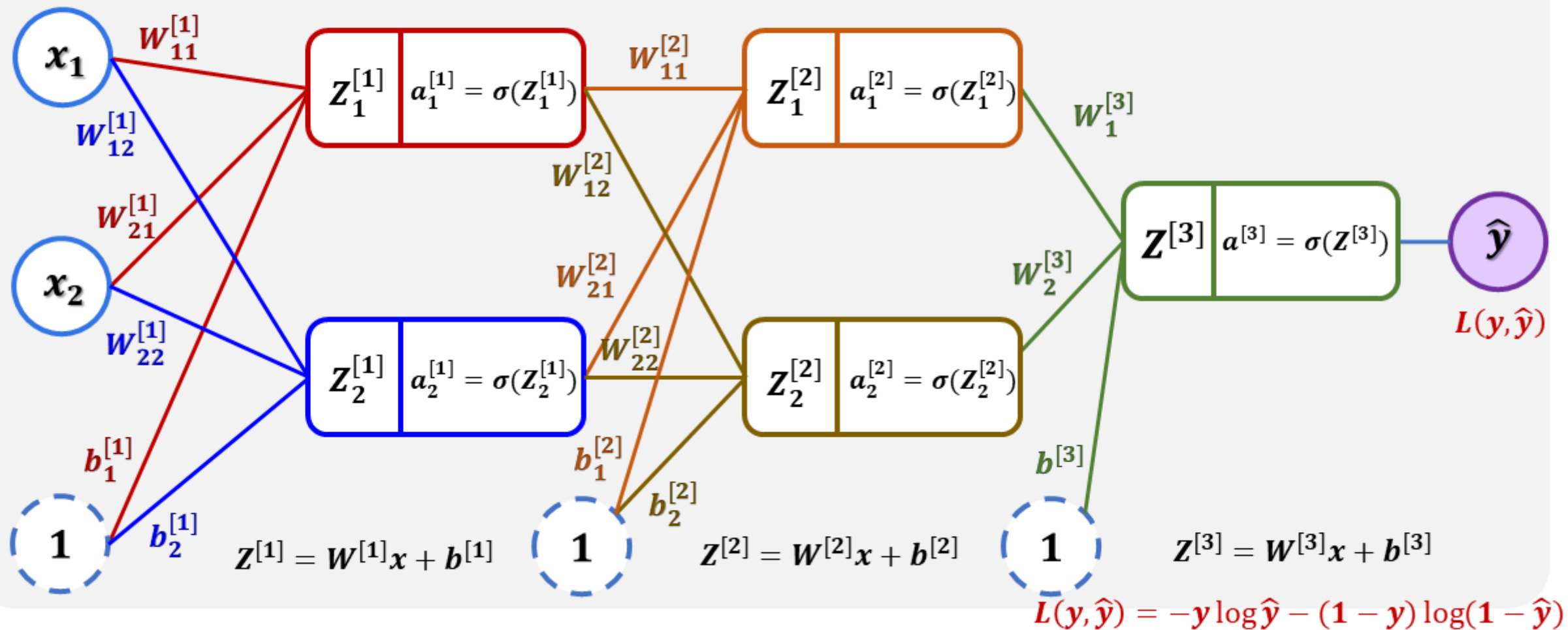
www.midocean.ae

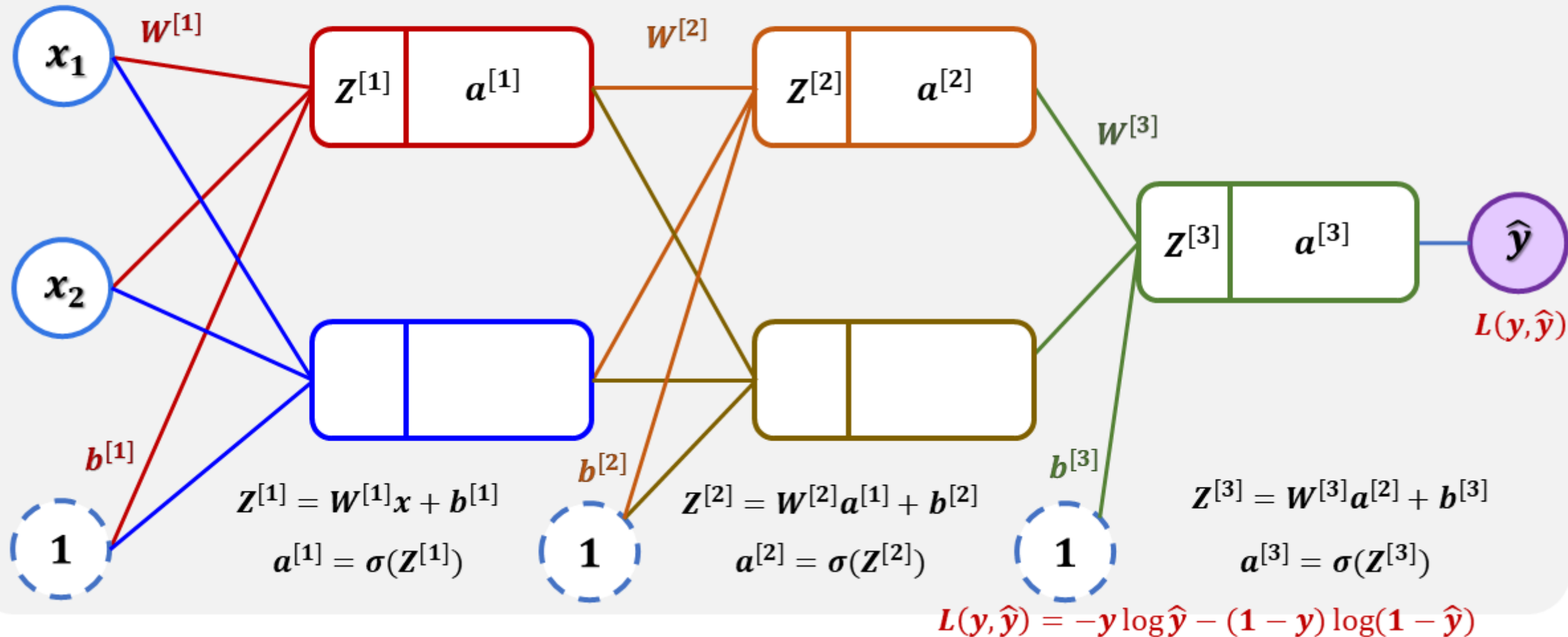
Lecture

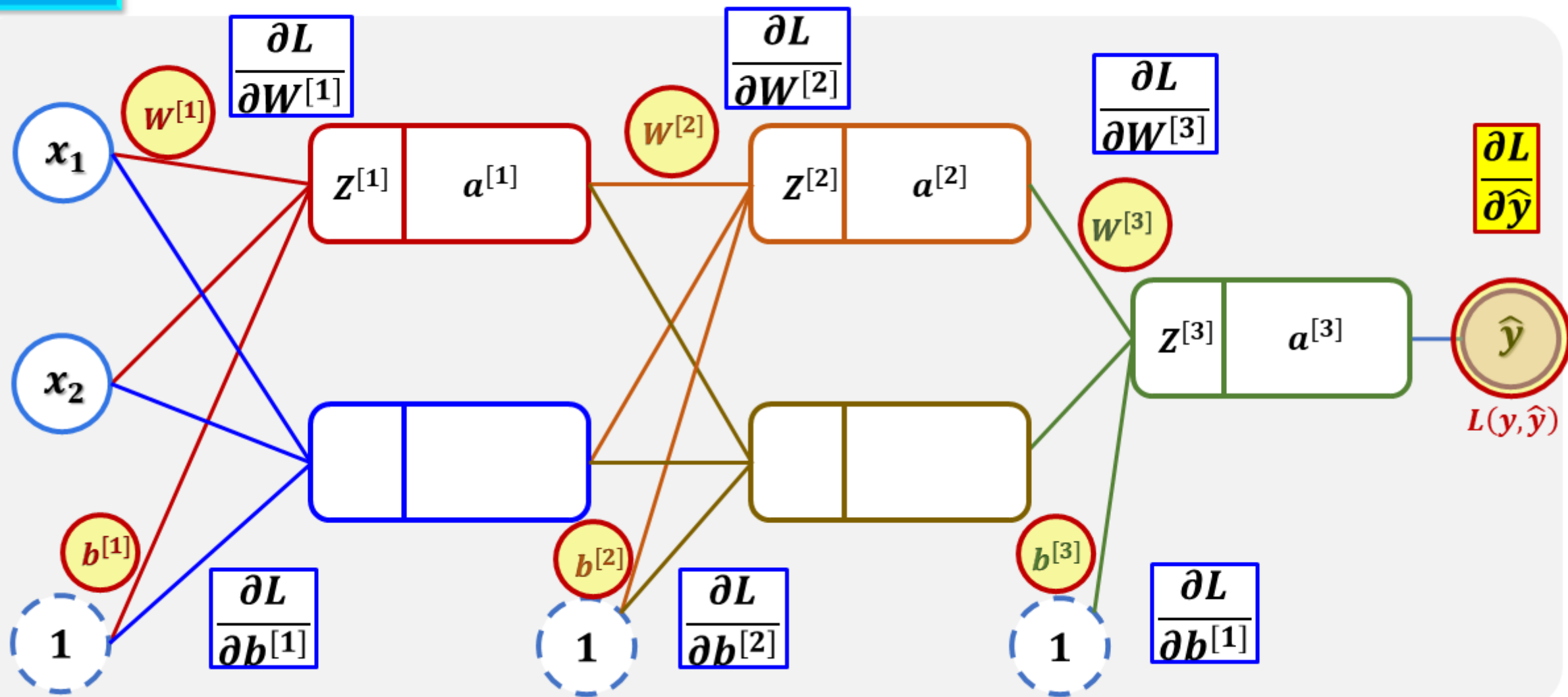
06

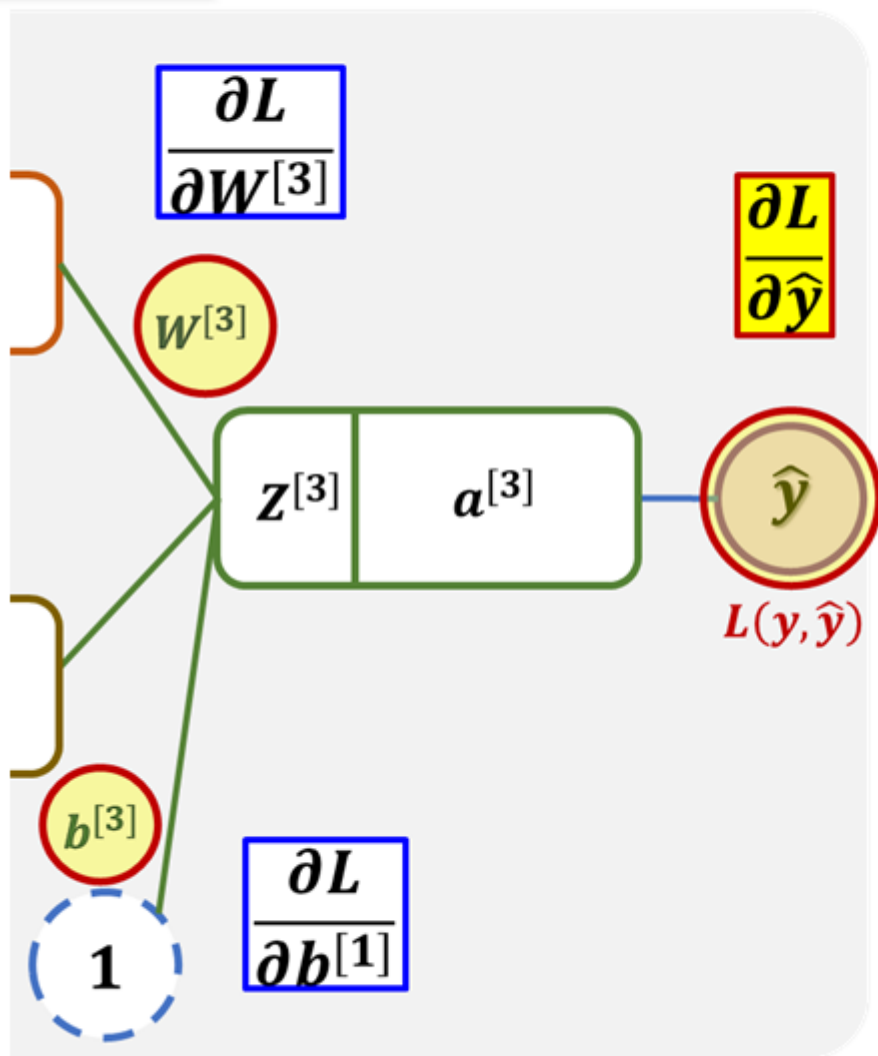
LEARNING OUTCOMES

- Understanding Backpropagation.
- Understanding The Second Derivative.
- Understanding Newtons Method
- Understanding The Hessian.
- Understanding Newton's Method for Two Variables.







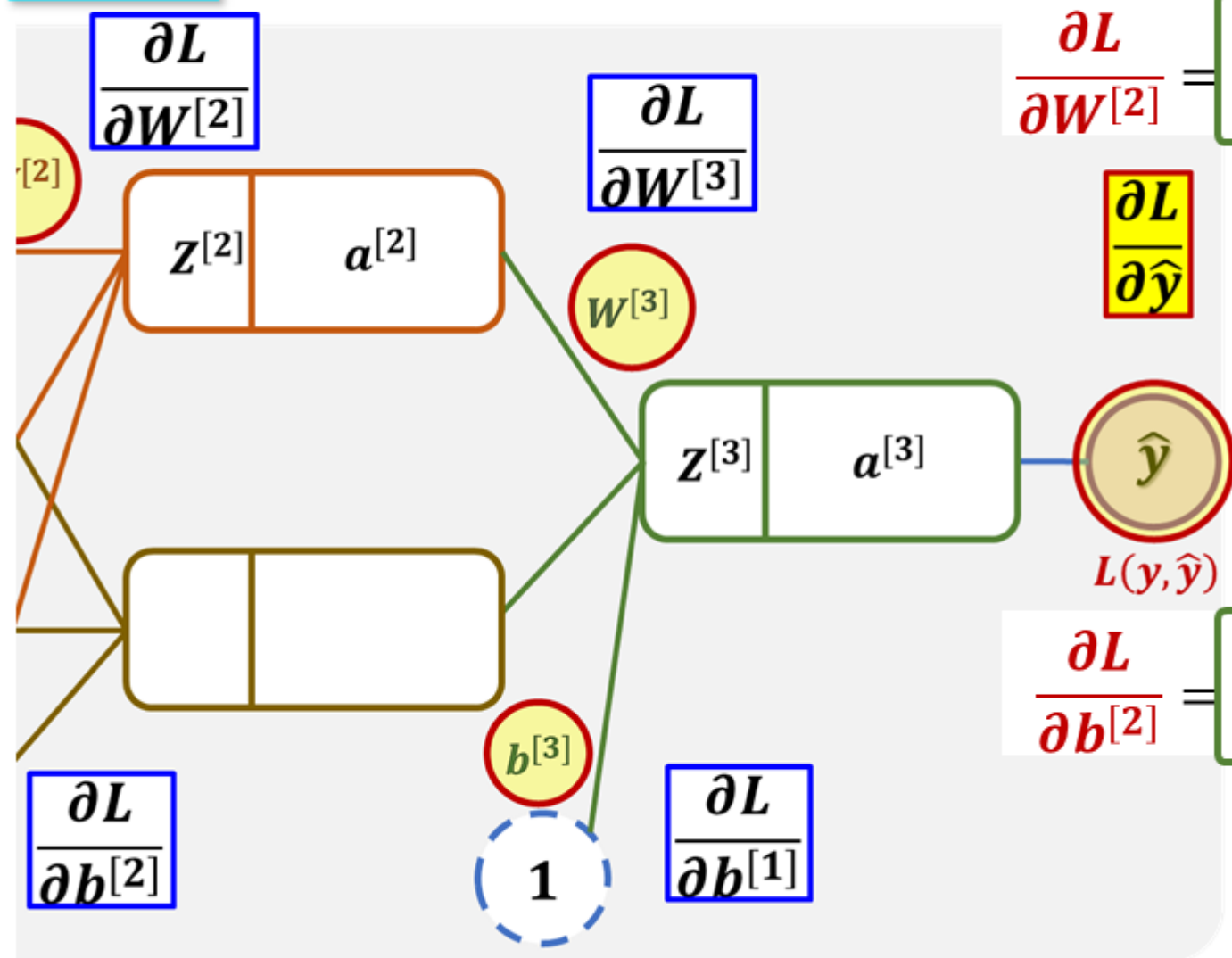


$$\frac{\partial L}{\partial W^{[3]}} = \frac{\partial L}{\partial a^{[3]}} \times \frac{\partial a^{[3]}}{\partial Z^{[3]}} \times \frac{\partial Z^{[3]}}{\partial W^{[3]}}$$

$$\frac{\partial L}{\partial b^{[3]}} = \frac{\partial L}{\partial a^{[3]}} \times \frac{\partial a^{[3]}}{\partial Z^{[3]}} \times \frac{\partial Z^{[3]}}{\partial b^{[3]}}$$

$$\frac{\partial L}{\partial a^{[3]}}$$

$$\frac{\partial a^{[3]}}{\partial Z^{[3]}}$$



$$\frac{\partial L}{\partial W^{[2]}} = \frac{\partial L}{\partial a^{[3]}} \times \frac{\partial a^{[3]}}{\partial z^{[3]}} \times \frac{\partial z^{[3]}}{\partial a^{[2]}} \times \frac{\partial a^{[2]}}{\partial z^{[2]}} \times \frac{\partial z^{[2]}}{\partial W^{[2]}}$$

$$\frac{\partial L}{\partial \hat{y}}$$

$$\frac{\partial L}{\partial b^{[2]}} = \frac{\partial L}{\partial a^{[3]}} \times \frac{\partial a^{[3]}}{\partial z^{[3]}} \times \frac{\partial z^{[3]}}{\partial a^{[2]}} \times \frac{\partial a^{[2]}}{\partial z^{[2]}} \times \frac{\partial z^{[2]}}{\partial b^{[2]}}$$

$$\frac{\partial z^{[3]}}{\partial a^{[2]}}$$

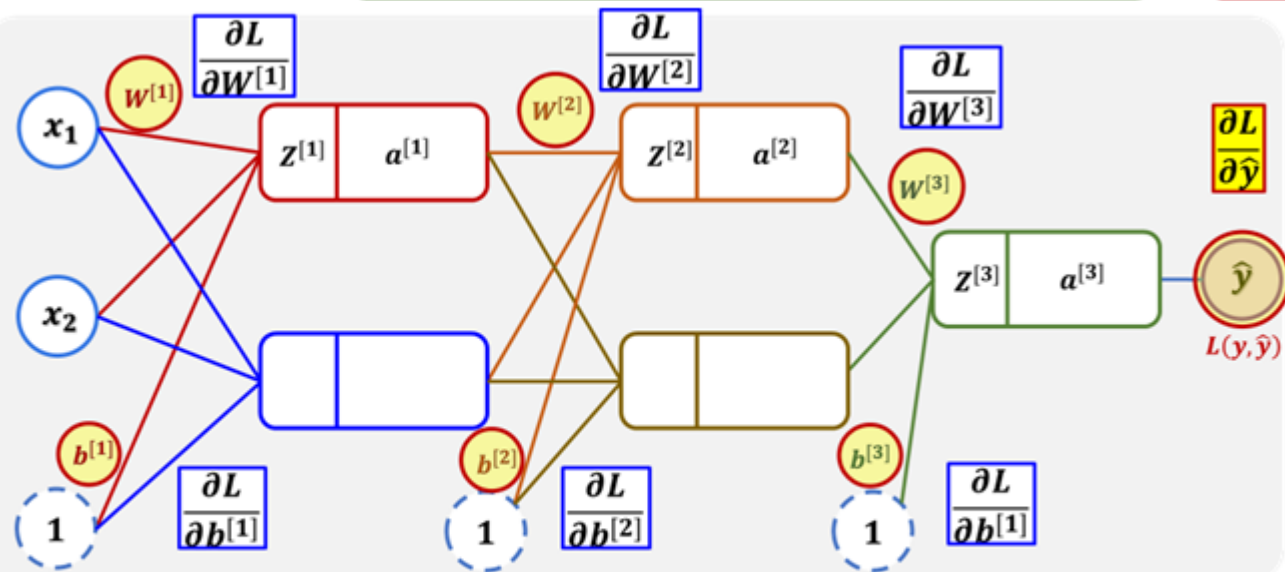
$$\frac{\partial a^{[2]}}{\partial z^{[2]}}$$

$$\frac{\partial L}{\partial a^{[3]}}$$

$$\frac{\partial a^{[3]}}{\partial z^{[3]}}$$

$$\frac{\partial L}{\partial W^{[1]}} = \frac{\partial L}{\partial a^{[3]}} \times \frac{\partial a^{[3]}}{\partial Z^{[3]}} \times \frac{\partial Z^{[3]}}{\partial a^{[2]}} \times \frac{\partial a^{[2]}}{\partial Z^{[2]}} \times \frac{\partial Z^{[2]}}{\partial a^{[1]}} \times \frac{\partial a^{[1]}}{\partial Z^{[1]}} \times \frac{\partial Z^{[1]}}{\partial W^{[1]}}$$

$$\frac{\partial L}{\partial b^{[1]}} = \frac{\partial L}{\partial a^{[3]}} \times \frac{\partial a^{[3]}}{\partial Z^{[3]}} \times \frac{\partial Z^{[3]}}{\partial a^{[2]}} \times \frac{\partial a^{[2]}}{\partial Z^{[2]}} \times \frac{\partial Z^{[2]}}{\partial a^{[1]}} \times \frac{\partial a^{[1]}}{\partial Z^{[1]}} \times \frac{\partial Z^{[1]}}{\partial b^{[1]}}$$



$$\frac{\partial Z^{[3]}}{\partial a^{[2]}}$$

$$\frac{\partial a^{[2]}}{\partial Z^{[2]}}$$

$$\frac{\partial L}{\partial a^{[3]}}$$

$$\frac{\partial a^{[3]}}{\partial Z^{[3]}}$$

*The **velocity** (v) is the rate of change of the distance in respect to the time, if x is the distance and t is the time,*

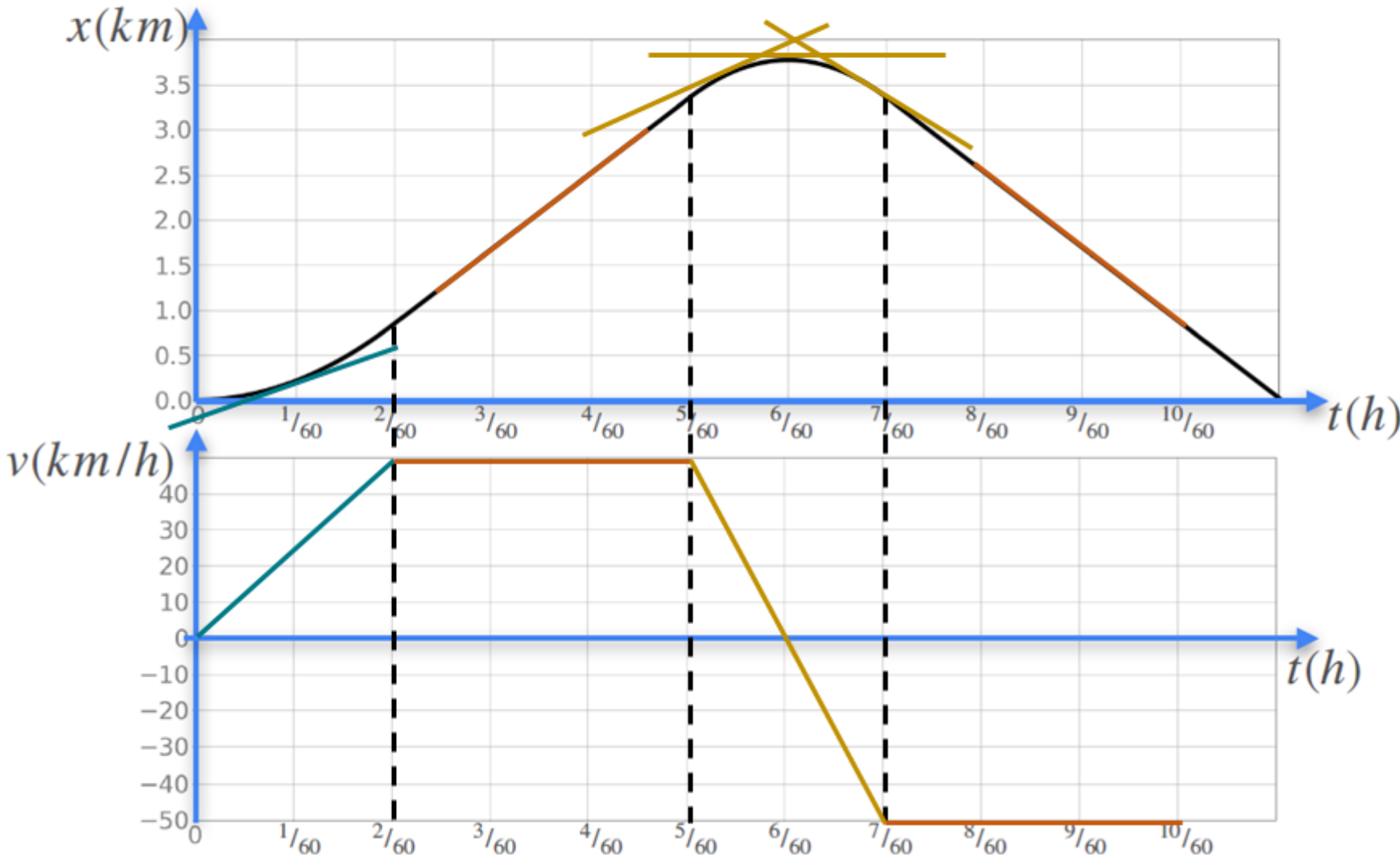
then the velocity (speed) = $\frac{dx}{dt} = f'(x)$



*The **acceleration** (a) is the rate of change of the velocity in respect to the time, if v is the velocity and t is the*

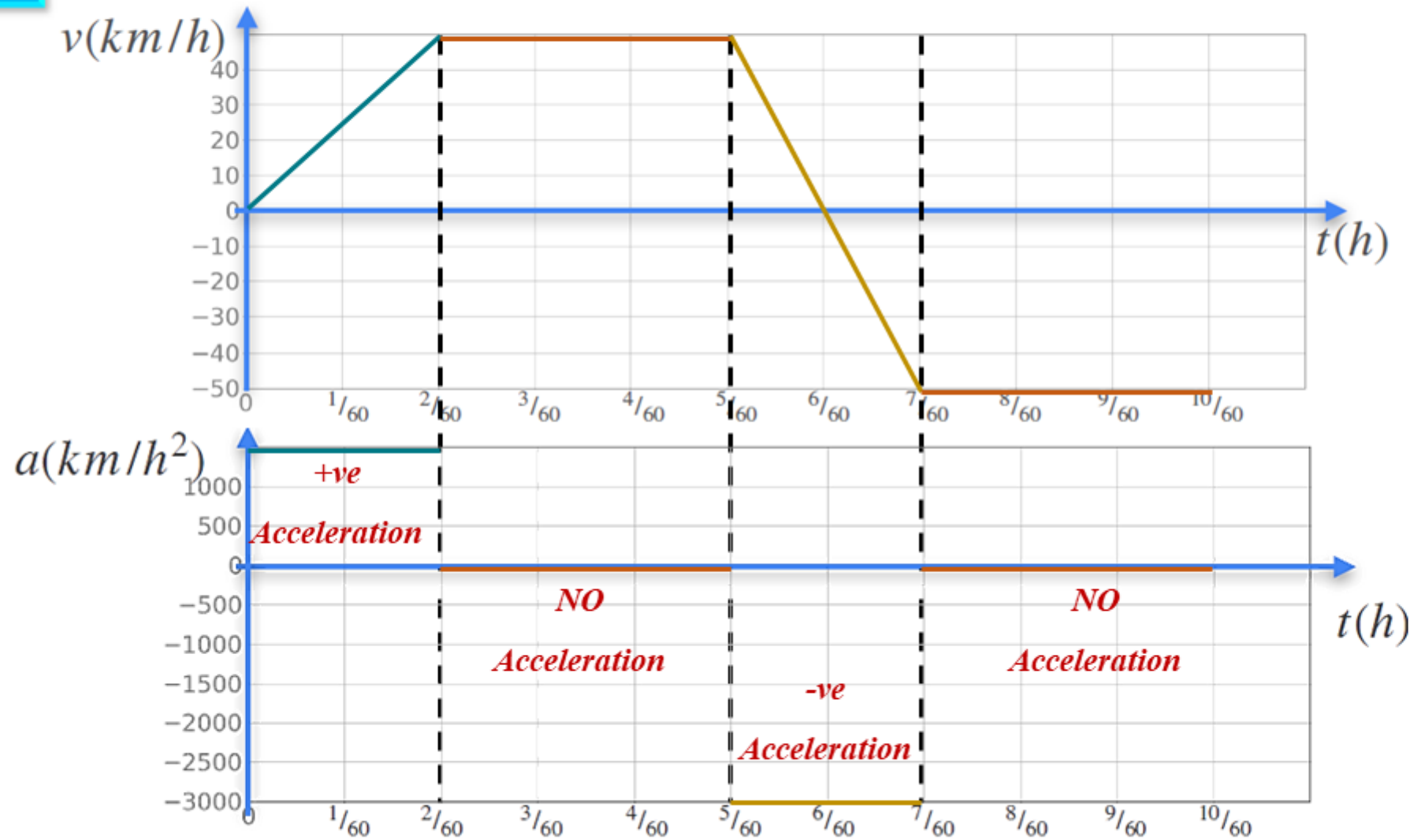
time, then the acceleration (a) = $\frac{dv}{dt}$, since $v = \frac{dx}{dt}$, the $a = \frac{d}{dt} \left[\frac{dx}{dt} \right] = \frac{d^2x}{d^2t} = f''(x)$ the second derivative of (x)

The Second Derivative

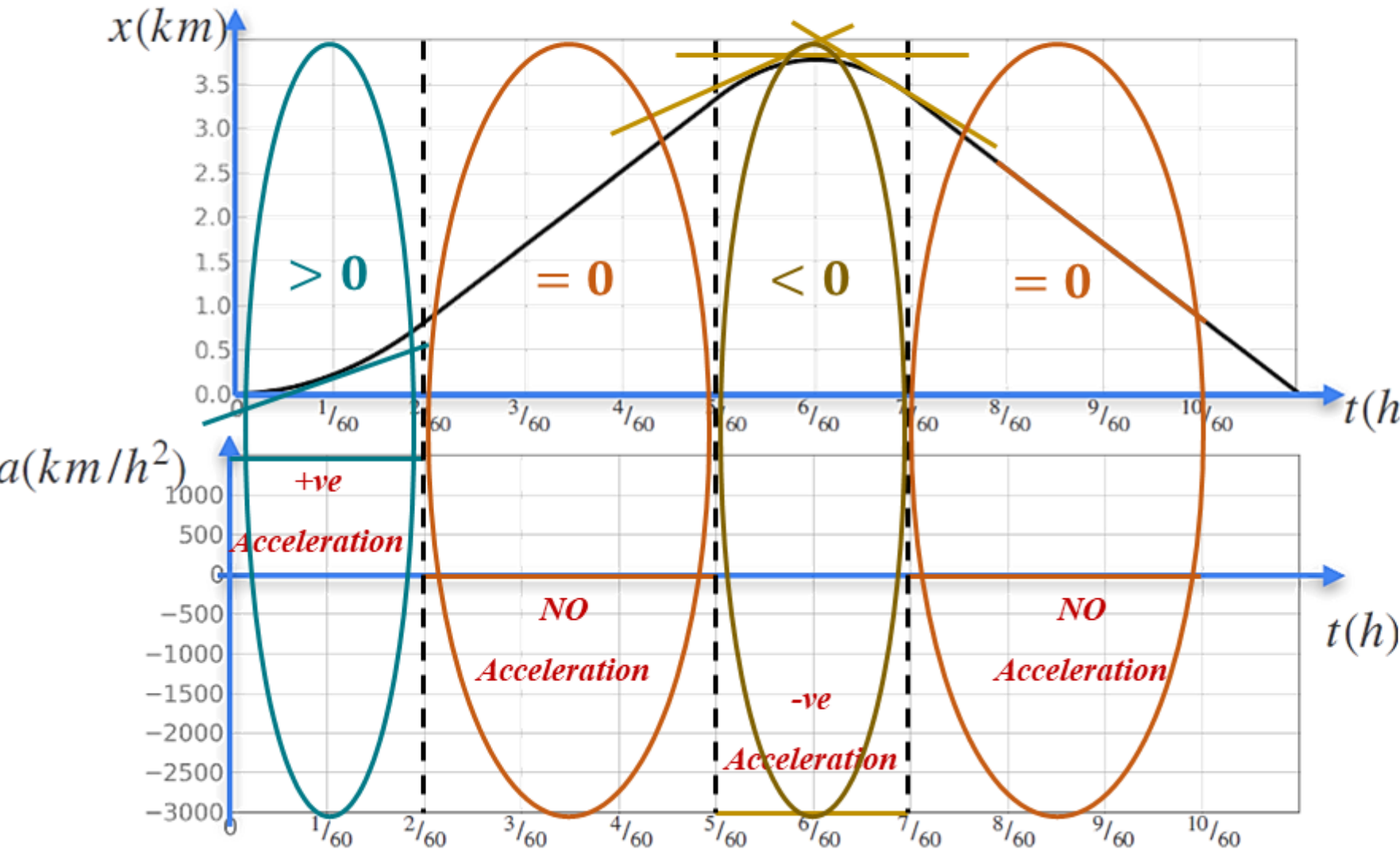


***Acceleration** is any process where the velocity changes. Since velocity is a speed and a direction, there are only two ways to accelerate: change in speed, change in direction, or change both.*

The Second Derivative



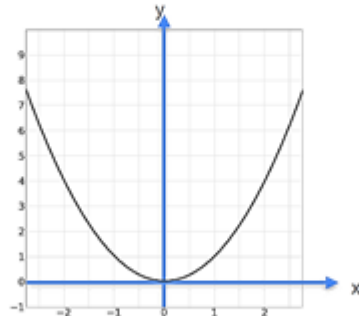
The Second Derivative



*The Second Derivative gives a measure of the amount by which a curve deviates from being a straight line. This is called the **curvature**.*

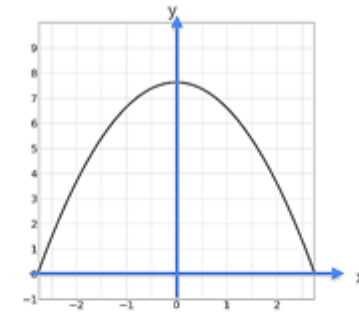
The Second Derivative

$$\frac{d^2x}{d^2t} > 0$$



Concave Up.

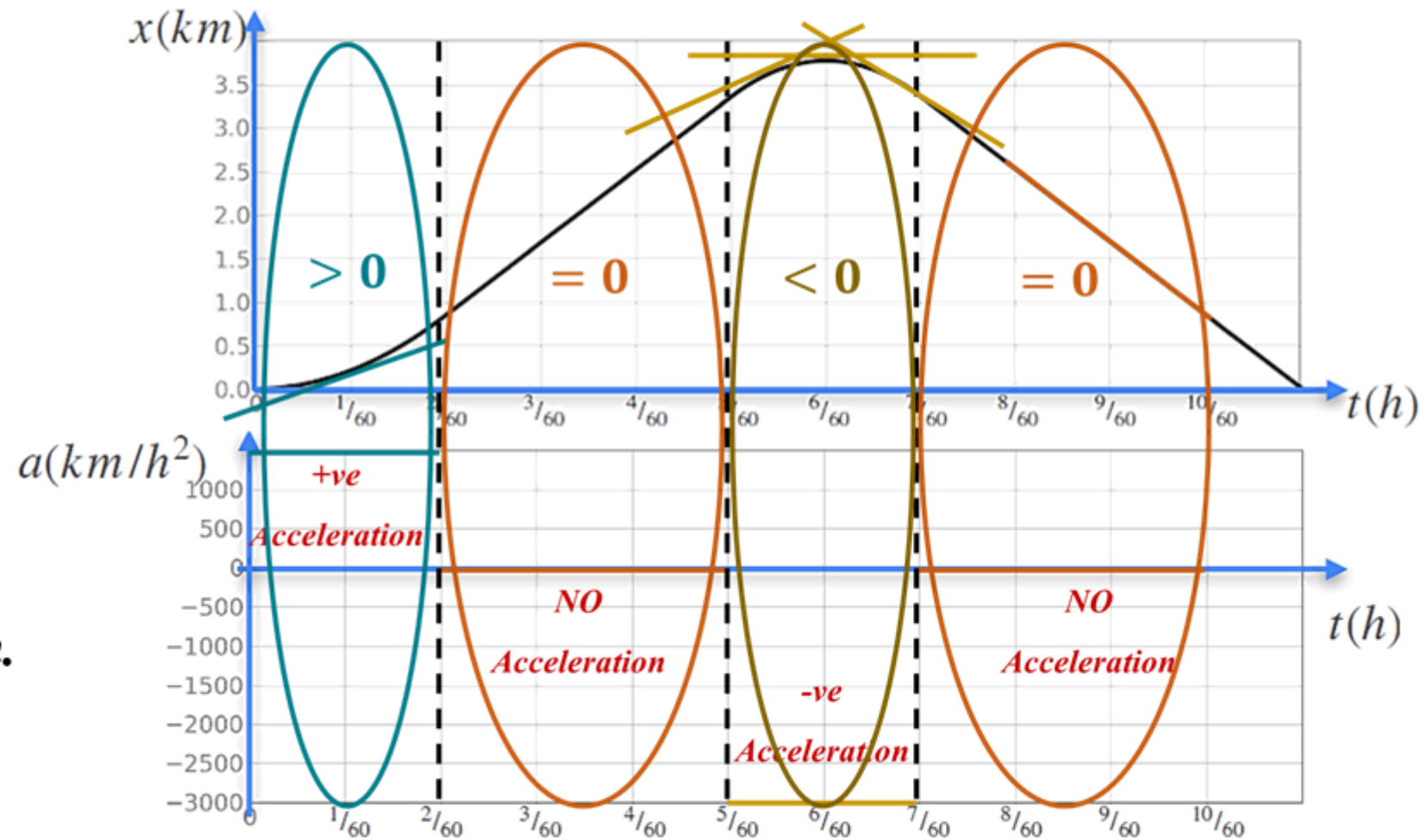
$$\frac{d^2x}{d^2t} < 0$$



Concave Down.

$$\frac{d^2x}{d^2t} = 0$$

Need more information

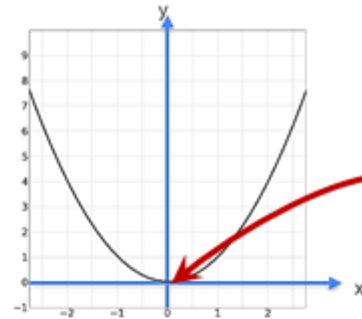


Why the second derivative is important in optimization?

The second derivative determine the maximum and minimum of the points, the points of derivative zero are the candidates for maximum or minimum, but we don't know if their maximum or minimum, but the second derivative does.

$$\frac{d^2x}{d^2t} > 0$$

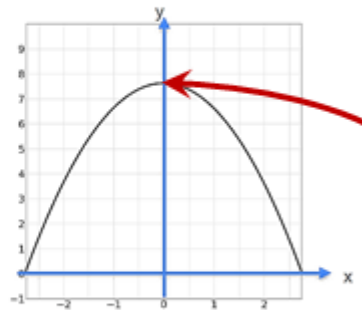
Concave Up.



Local Minimum

$$\frac{d^2x}{d^2t} < 0$$

Concave Down.



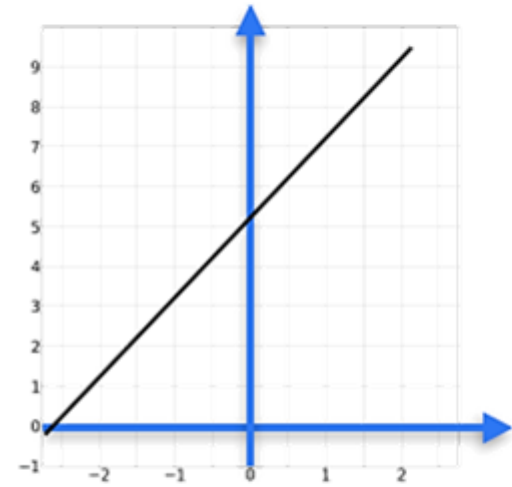
Local Maximum

$$\frac{d^2x}{d^2t} = 0$$

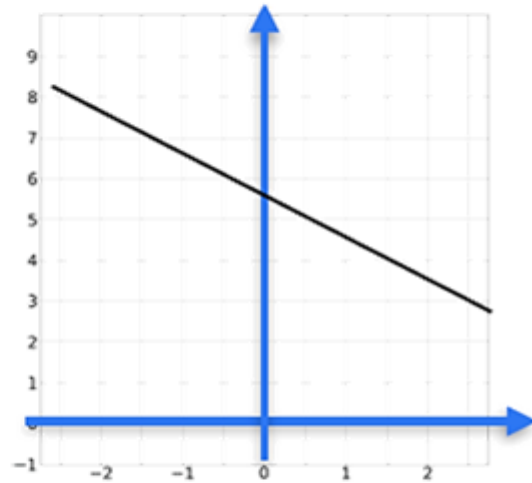
Inconclusive

The Second Derivative

First Derivative

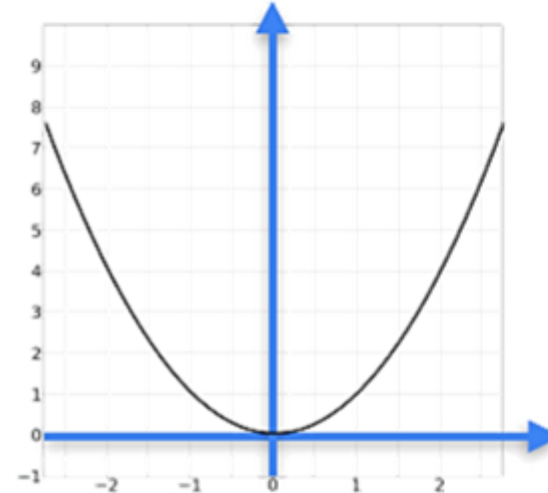
*Increasing*

$$f'(x) > 0$$

*decreasing*

$$f'(x) < 0$$

Second Derivative

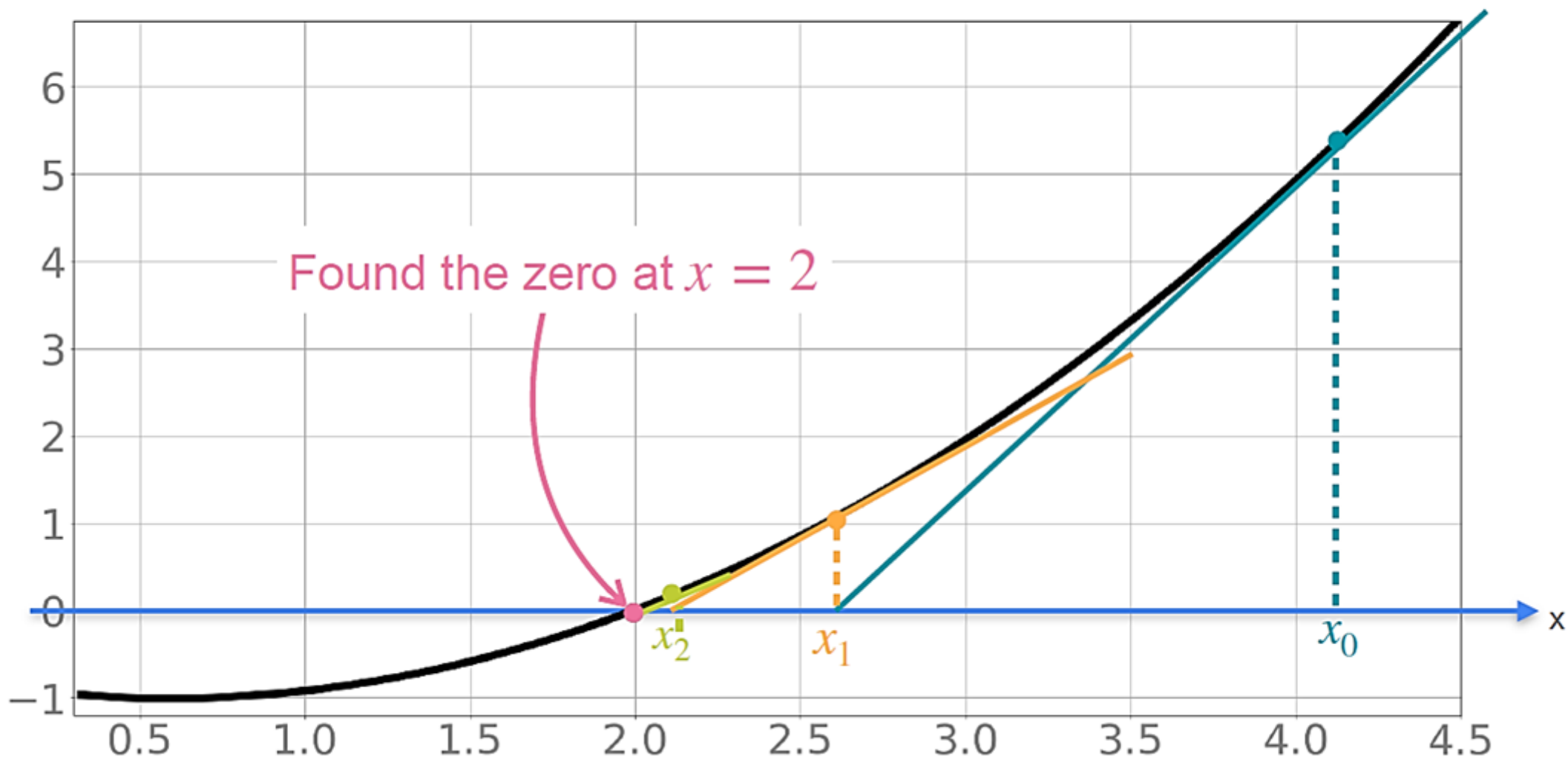
*Concave Up*

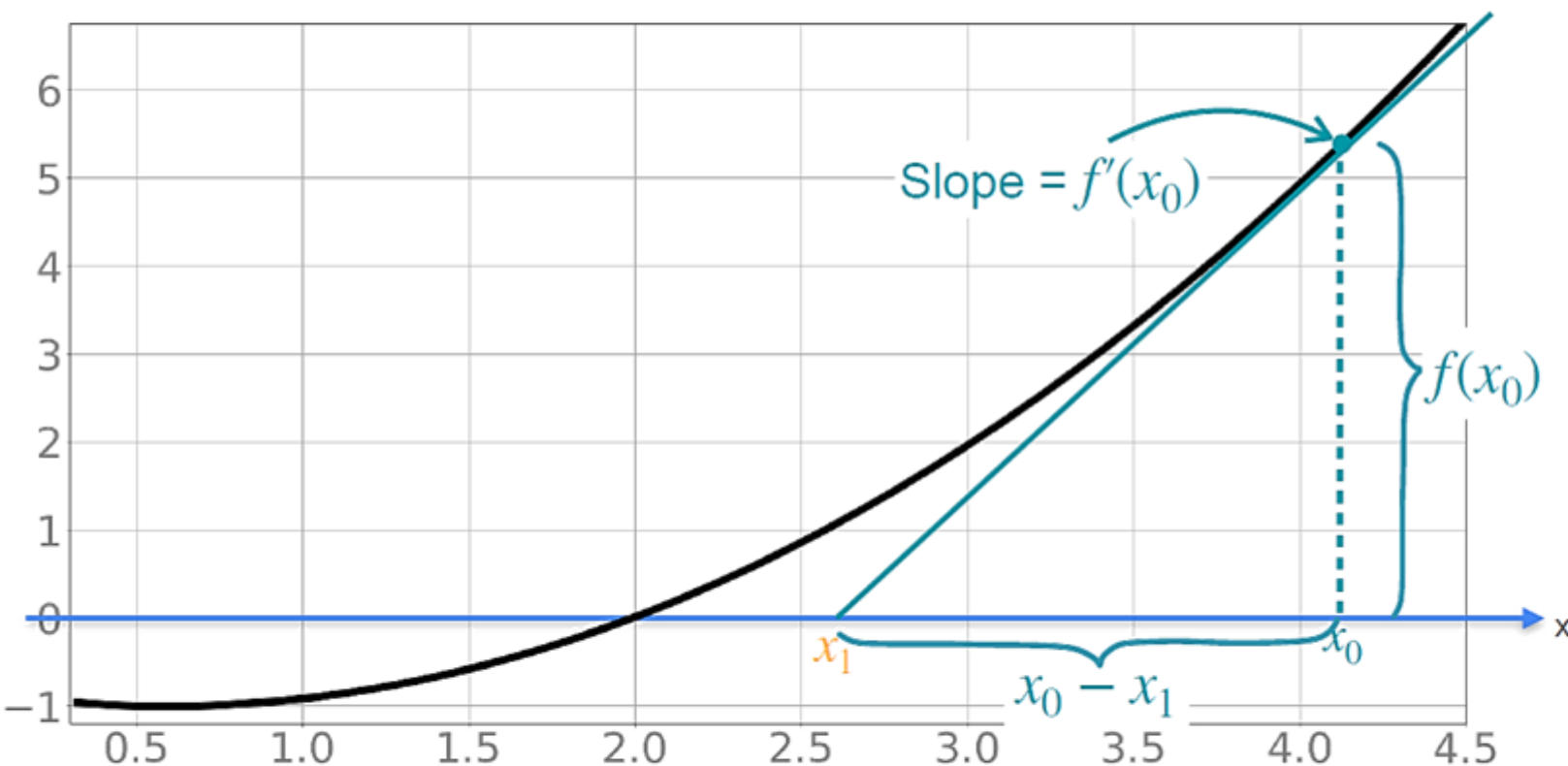
$$f''(x) > 0$$

*Concave Down*

$$f''(x) < 0$$

Newton's Method

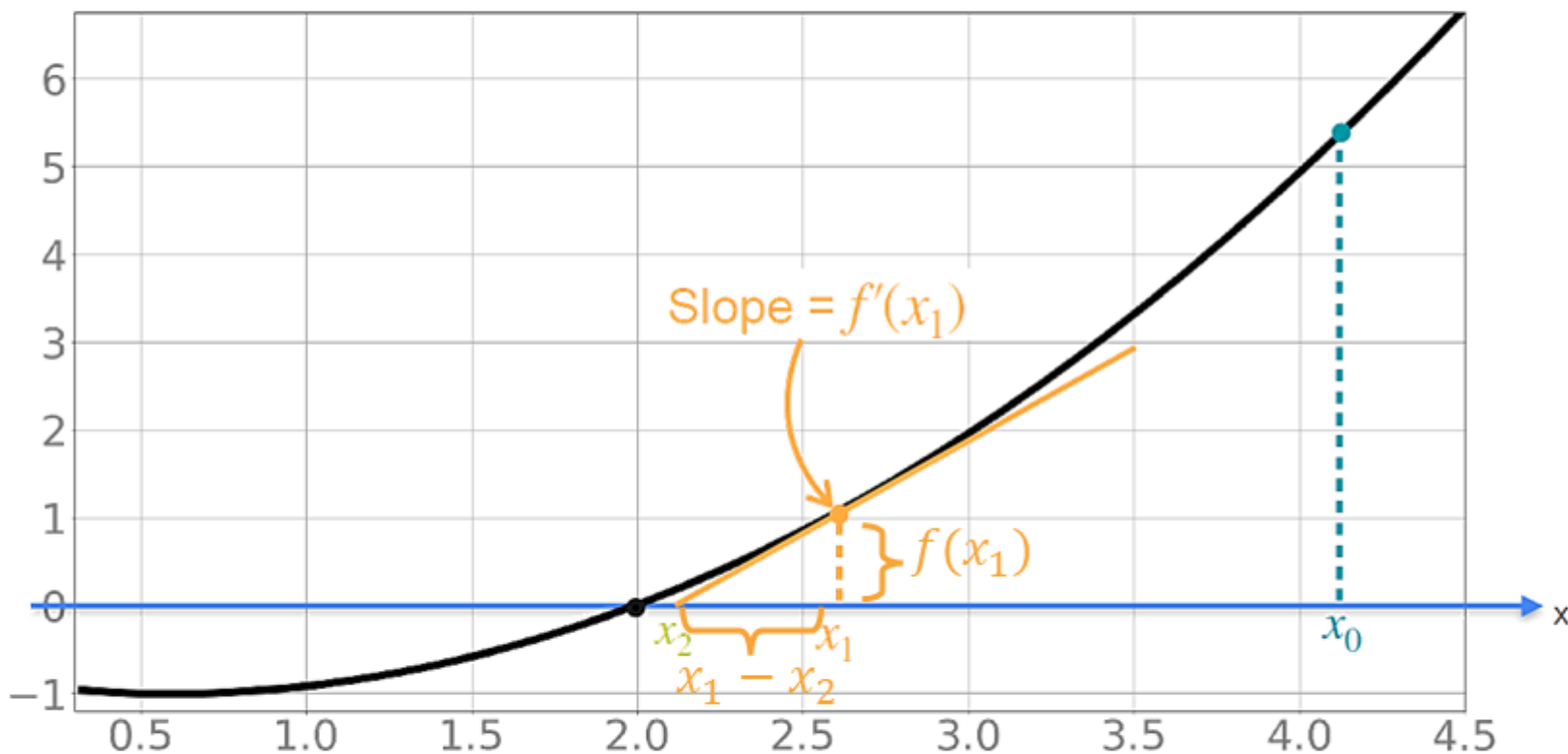




$$f'(x_0) = \frac{f(x_0)}{x_0 - x_1}$$

$$\frac{f(x_0)}{f'(x_0)} = x_0 - x_1$$

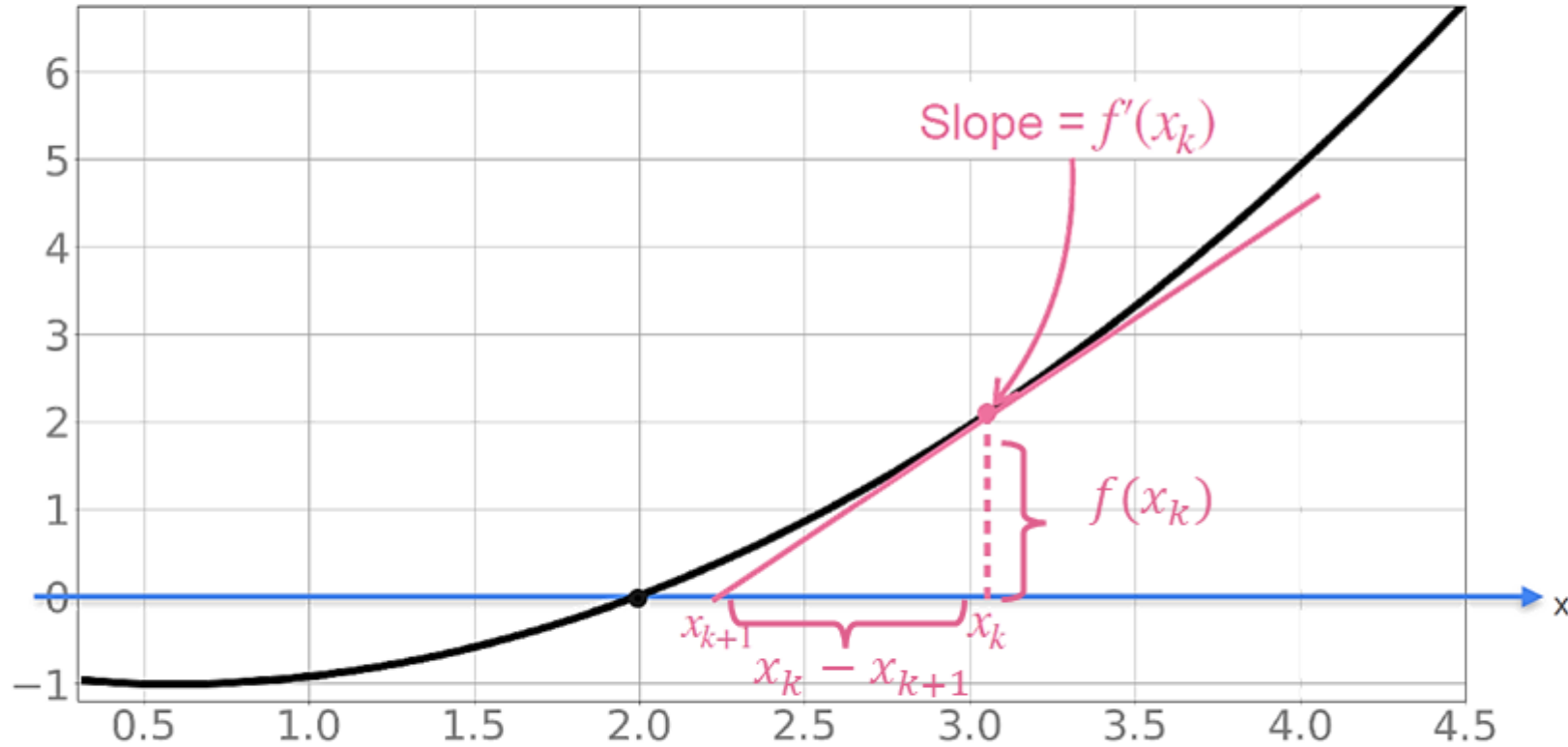
$$x_1 = x_0 - \frac{f(x_0)}{f'(x_0)}$$



$$f'(x_1) = \frac{f(x_1)}{x_1 - x_2}$$

$$\frac{f(x_1)}{f'(x_1)} = x_1 - x_2$$

$$x_2 = x_1 - \frac{f(x_1)}{f'(x_1)}$$



$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

Newton's Method

Goal: Find the Zero of $f(x)$

1. Start with initial value of x_k
2. Update to x_{k+1} using:

$$x_{k+1} = x_k - \frac{f(x_k)}{f'(x_k)}$$

3. Repeat step (2) until finding the root

Newton's Method for Optimization

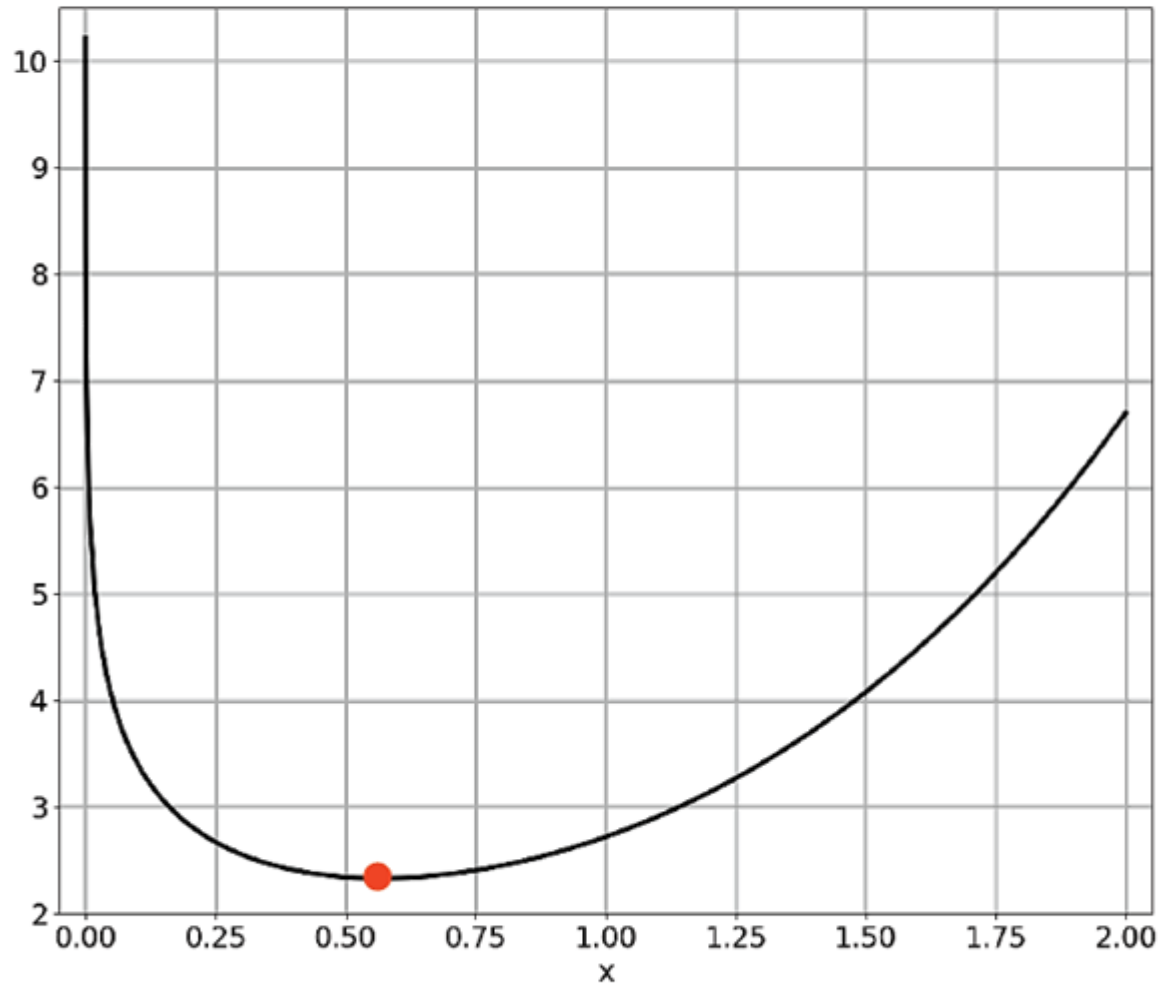
Goal: Find the Zero of $g'(x)$

$$f(x) \rightarrow g'(x) \qquad f'(x) \rightarrow g''(x)$$

1. Start with initial value of x_k
2. Update to x_{k+1} using:

$$x_{k+1} = x_k - \frac{g'(x_k)}{g''(x_k)}$$

3. Repeat step (2) until finding the root

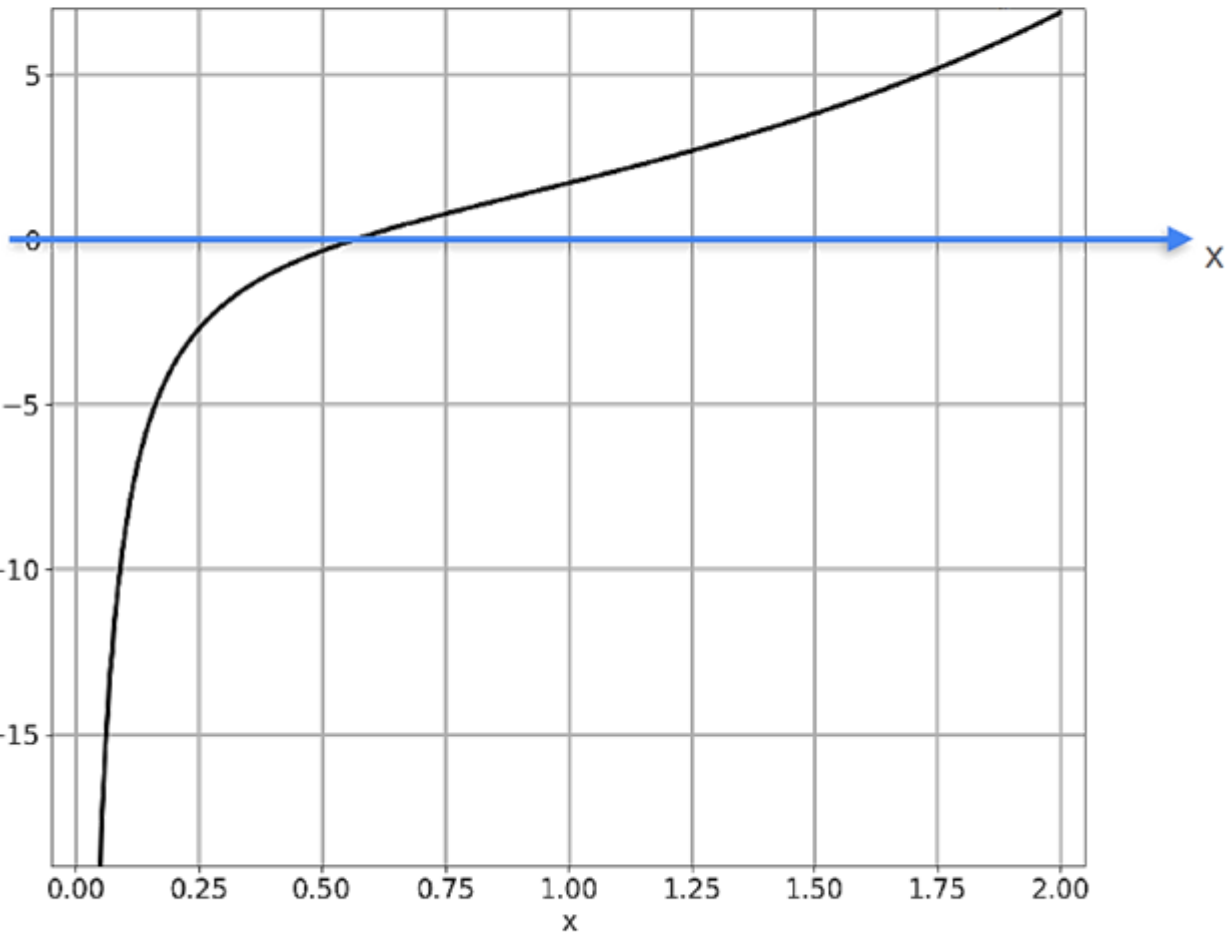


$$f(x) = e^x - \log(x) \rightarrow f'(x) = e^x - \frac{1}{x}$$

Minimum: $x^* = 0.5671$

Known as the Omega constant (Ω)

Minimum: $x^* = 0.5671$



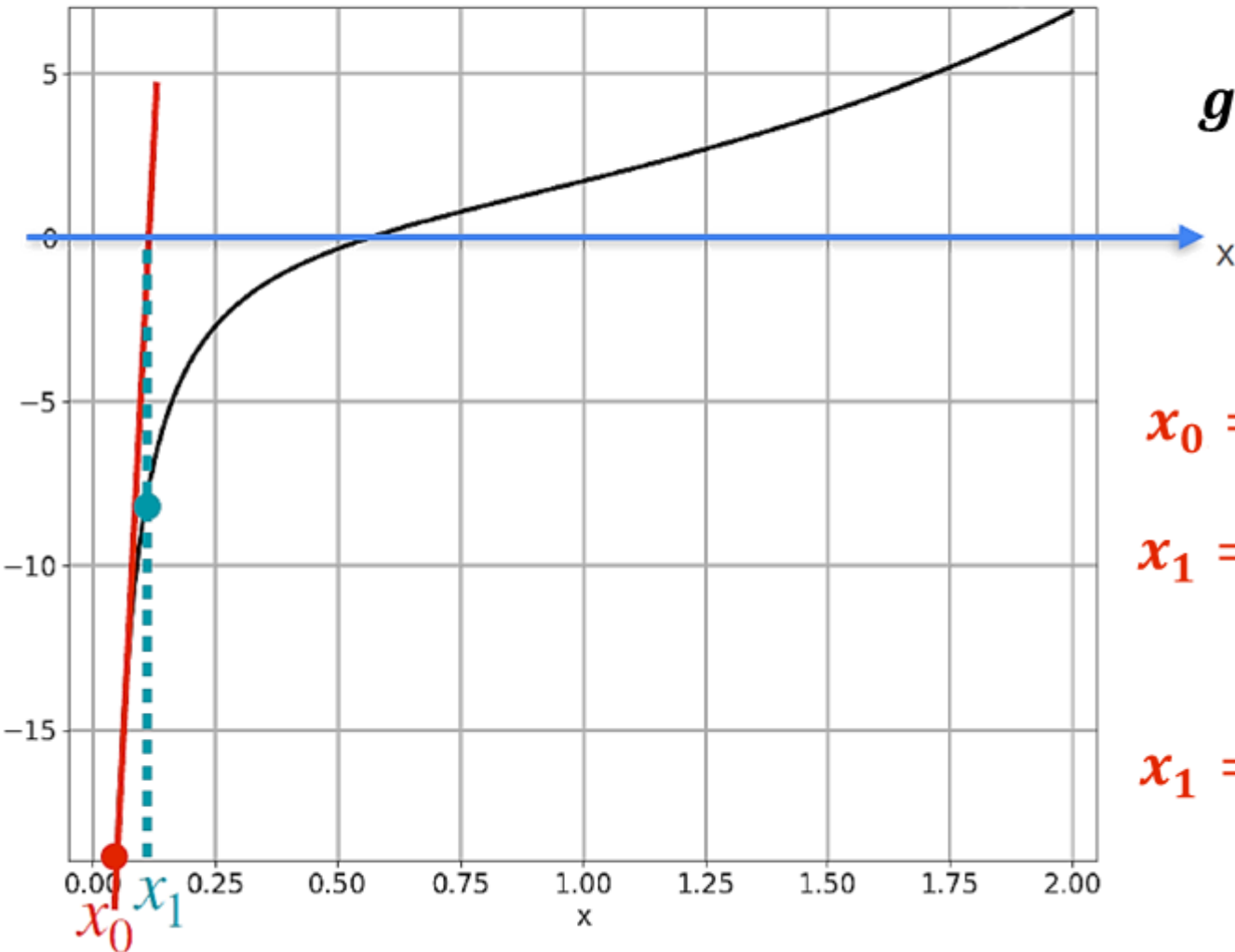
$$g(x) = e^x - \log(x) \rightarrow g'(x) = e^x - \frac{1}{x}$$

The derivative $g'(x)$ is shown with a blue bracket above it labeled $f(x)$.

$$g''(x) = e^x + \frac{1}{x^2}$$

The second derivative $g''(x)$ is shown with a blue bracket below it labeled $f'(x)$.

Newton's Method

Minimum: $x^* = 0.5671$ 

$$g(x) = e^x - \log(x) \rightarrow g'(x) = e^x - \frac{1}{x}$$

$$g''(x) = e^x + \frac{1}{x^2}$$

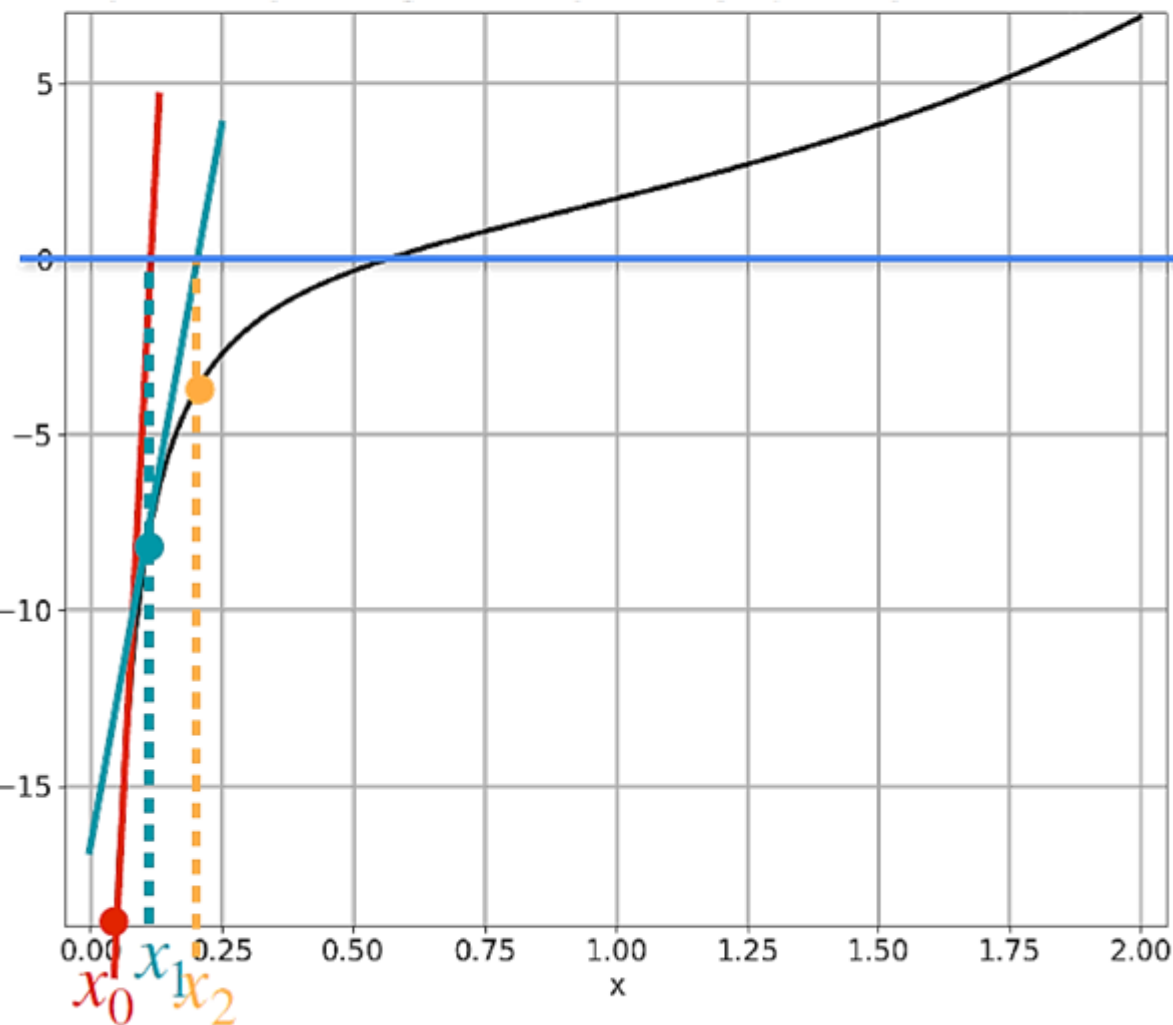
$$f'(x)$$

$$x_0 = 0.05$$

$$x_1 = x_0 - \frac{g'(x_0)}{g''(x_0)}$$

$$x_1 = 0.05 - \frac{\left(e^{0.05} - \frac{1}{0.05}\right)}{\left(e^{0.05} + \frac{1}{0.05^2}\right)} = 0.097$$

Newton's Method

Minimum: $x^* = 0.5671$ 

$$g(x) = e^x - \log(x) \rightarrow g'(x) = e^x - \frac{1}{x}$$

$$g''(x) = e^x + \frac{1}{x^2}$$

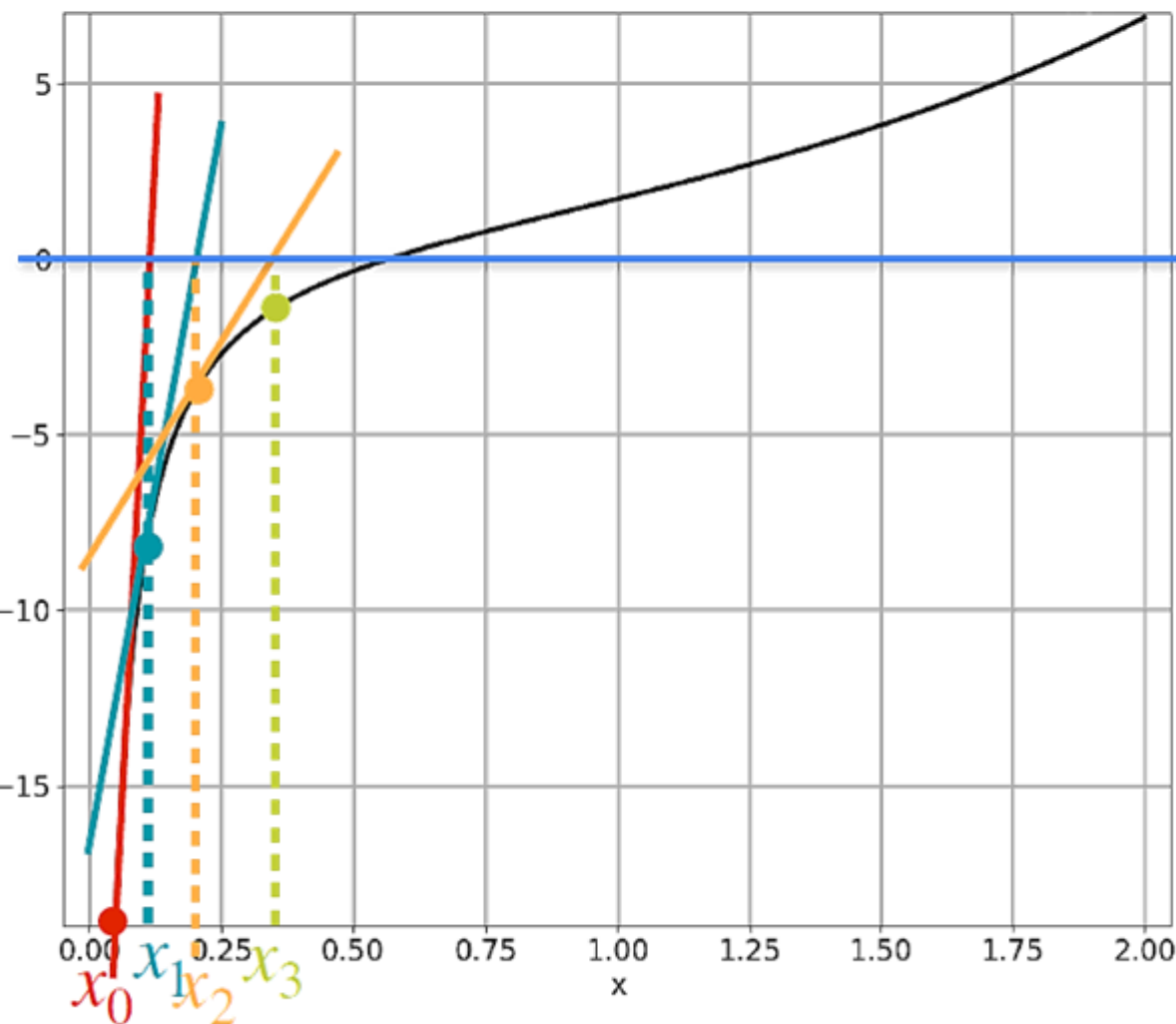
$$f'(x)$$

$$x_1 = 0.097$$

$$x_2 = x_1 - \frac{g'(x_1)}{g''(x_1)}$$

$$x_2 = 0.097 - \frac{\left(e^{0.097} - \frac{1}{0.097}\right)}{\left(e^{0.097} + \frac{1}{0.097^2}\right)} = 0.183$$

Newton's Method

Minimum: $x^* = 0.5671$ 

$$g(x) = e^x - \log(x) \rightarrow g'(x) = e^x - \frac{1}{x}$$

$$g''(x) = e^x + \frac{1}{x^2}$$

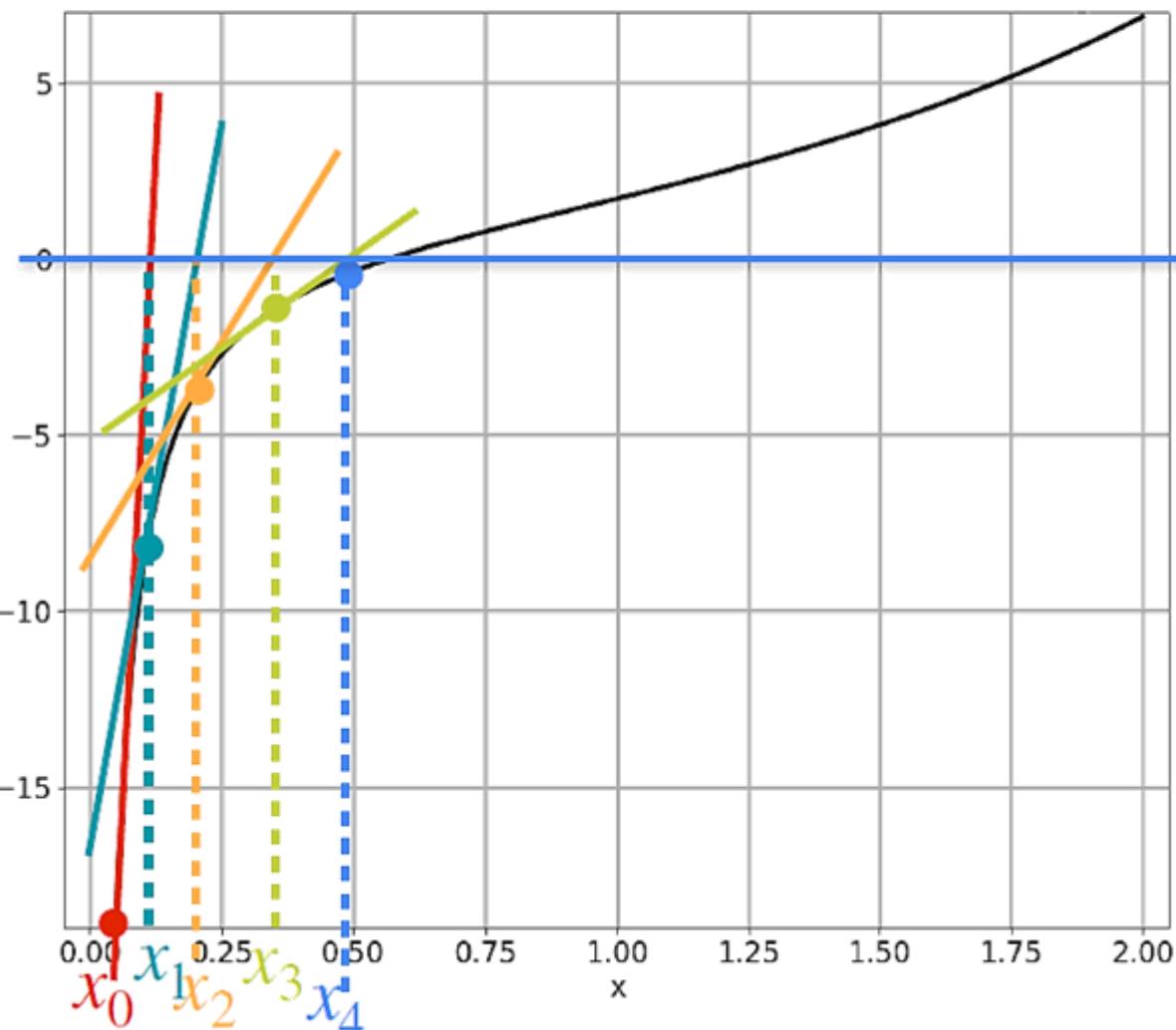
$$f'(x)$$

$$x_2 = 0.183$$

$$x_3 = x_2 - \frac{g'(x_2)}{g''(x_2)}$$

$$x_3 = 0.183 - \frac{\left(e^{0.183} - \frac{1}{0.183}\right)}{\left(e^{0.183} + \frac{1}{0.183^2}\right)} = 0.320$$

Newton's Method

Minimum: $x^* = 0.5671$ 

$$g(x) = e^x - \log(x) \rightarrow g'(x) = e^x - \frac{1}{x}$$

$$g''(x) = e^x + \frac{1}{x^2}$$

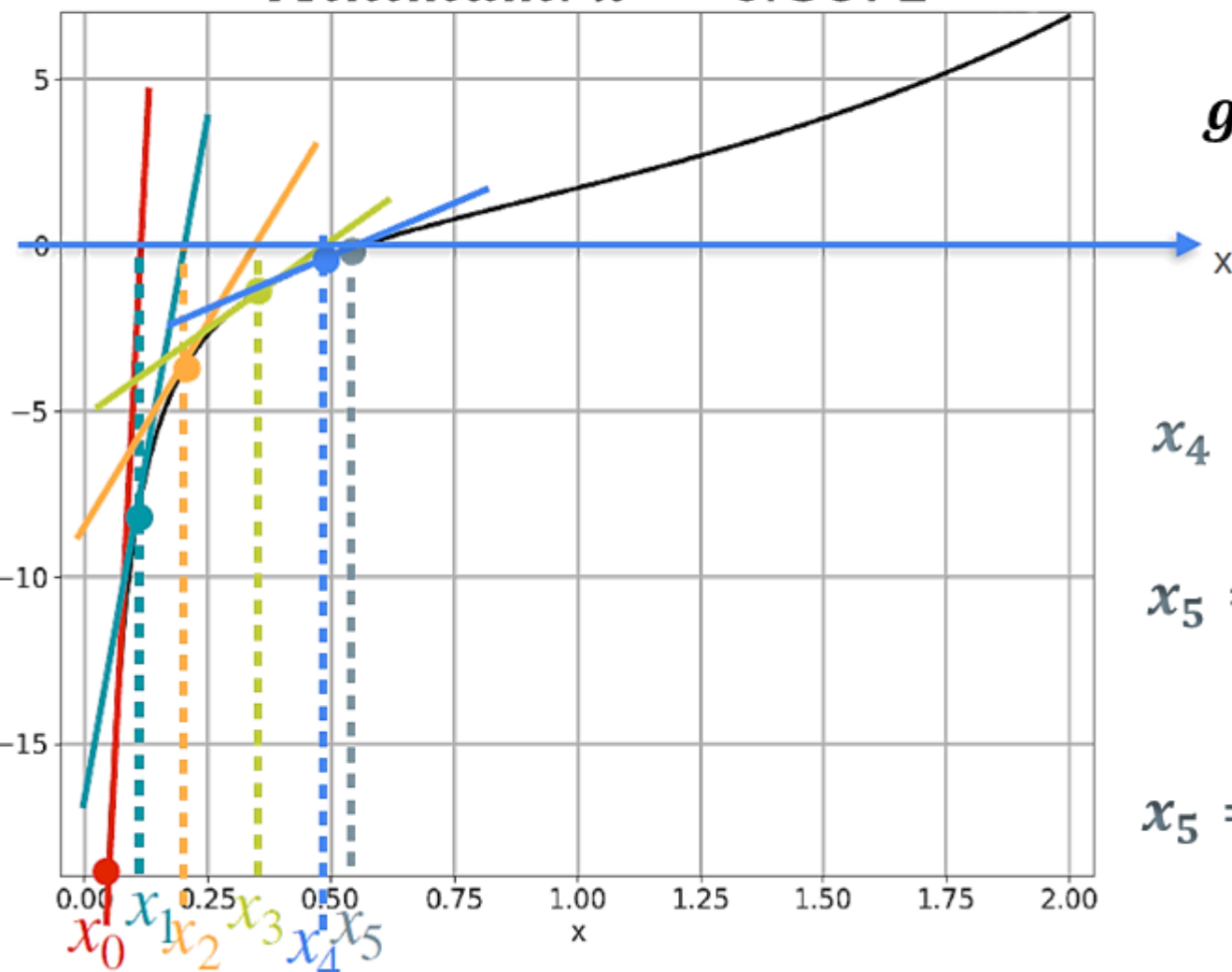
$$f'(x)$$

$$x_3 = 0.320$$

$$x_4 = x_3 - \frac{g'(x_3)}{g''(x_3)}$$

$$x_4 = 0.320 - \frac{\left(e^{0.320} - \frac{1}{0.320}\right)}{\left(e^{0.320} + \frac{1}{0.320^2}\right)} = 0.477$$

Newton's Method

Minimum: $x^* = 0.5671$ 

$$g(x) = e^x - \log(x) \rightarrow g'(x) = e^x - \frac{1}{x}$$

$$g''(x) = e^x + \frac{1}{x^2}$$

$$f(x)$$

$$f'(x)$$

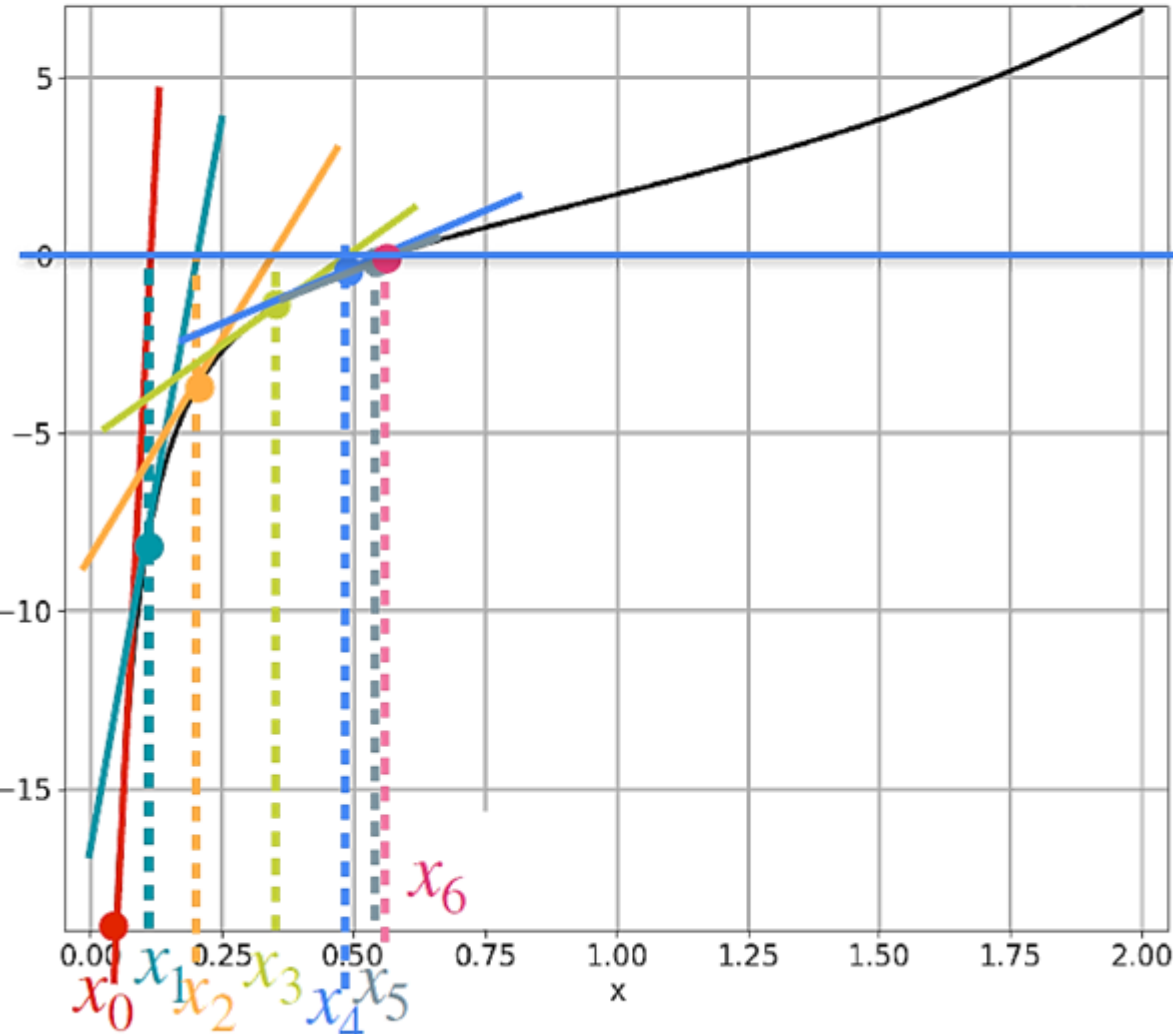
$$x_4 = 0.477$$

$$x_5 = x_4 - \frac{g'(x_4)}{g''(x_4)}$$

$$x_5 = 0.477 - \frac{\left(e^{0.477} - \frac{1}{0.477}\right)}{\left(e^{0.320} + \frac{1}{0.477^2}\right)} = 0.558$$

Newton's Method

Minimum: $x^* = 0.5671$



$$g(x) = e^x - \log(x) \rightarrow g'(x) = e^x - \frac{1}{x}$$

$$g''(x) = e^x + \frac{1}{x^2}$$

$$f'(x)$$

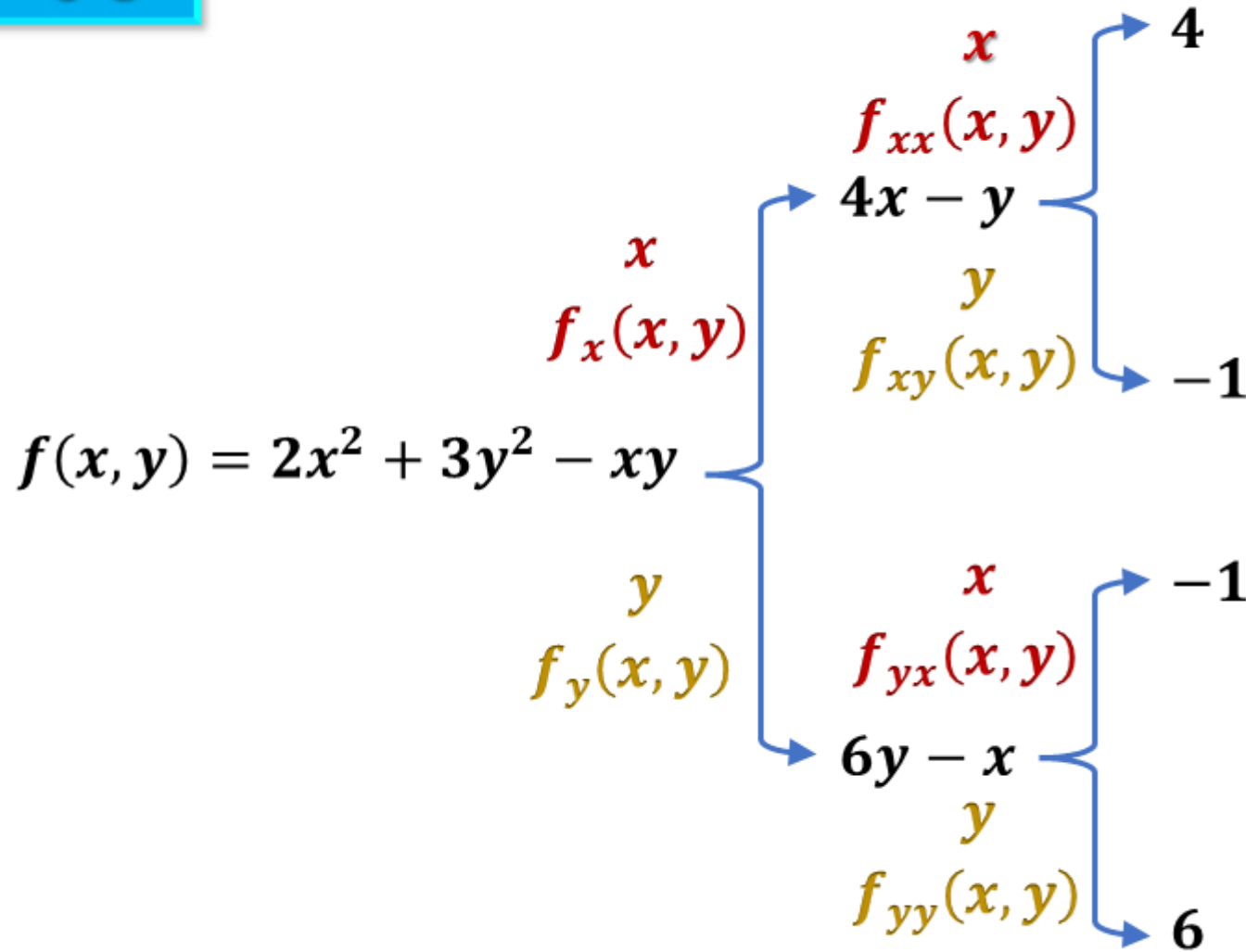
$$x_5 = 0.558$$

Minimum (x^*)

$$x_6 = x_5 - \frac{g'(x_5)}{g''(x_5)}$$

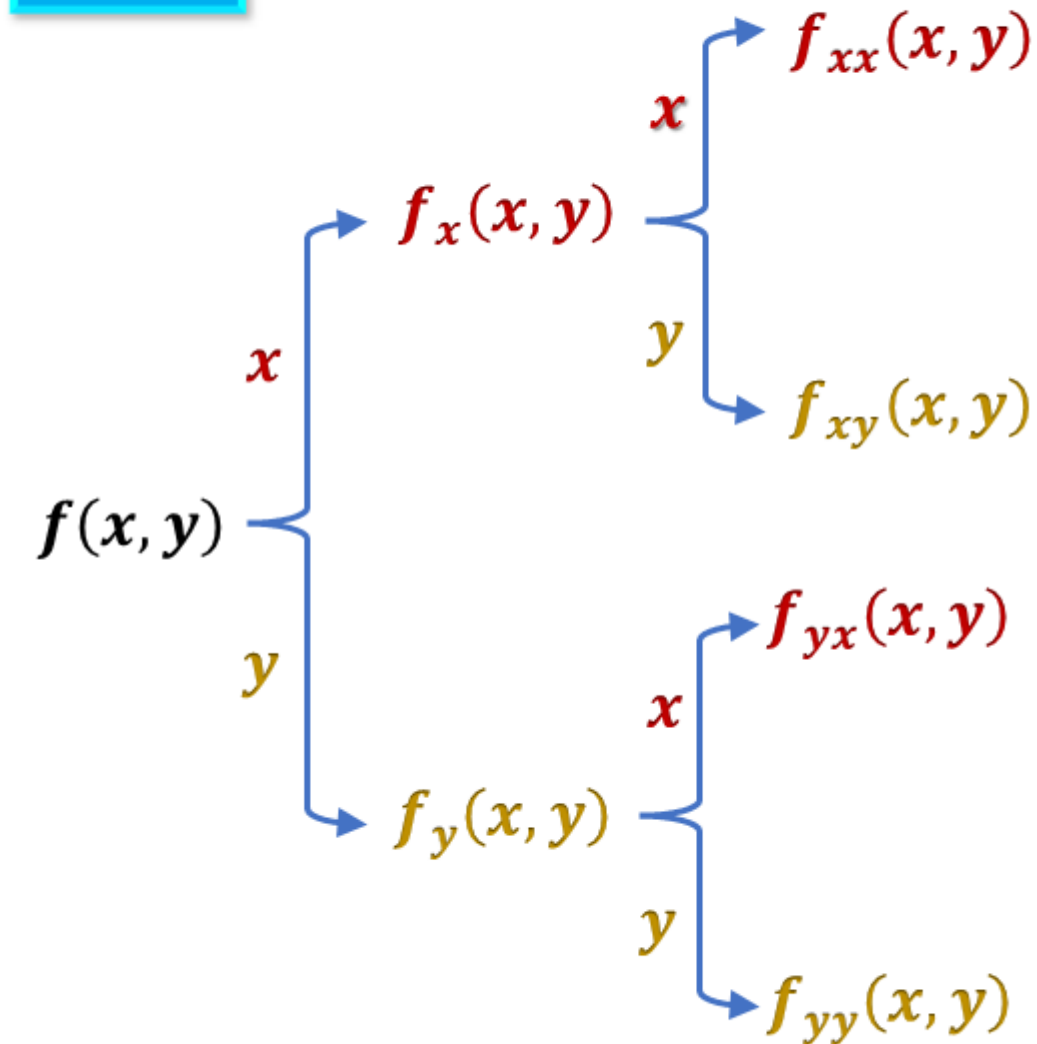
$$x_6 = 0.558 - \frac{\left(e^{0.558} - \frac{1}{0.558}\right)}{\left(e^{0.558} + \frac{1}{0.558^2}\right)} = 0.567$$

The Hessian



	Leibniz's notation	Lagrange's notation
Rate of change of $f'_x(x, y)$ with respect to x	$\frac{\partial^2 f}{\partial x^2}$	$f_{xx}(x, y)$
Rate of change of $f'_y(x, y)$ with respect to y	$\frac{\partial^2 f}{\partial y^2}$	$f_{yy}(x, y)$
Rate of change of $f'_x(x, y)$ with respect to y	$\frac{\partial^2 f}{\partial x \partial y}$	$f_{xy}(x, y)$
Rate of change of $f'_y(x, y)$ with respect to x	$\frac{\partial^2 f}{\partial y \partial x}$	$f_{yx}(x, y)$

The Hessian



$$\nabla f_x = \begin{bmatrix} f_{xx}(x, y) \\ f_{xy}(x, y) \end{bmatrix} \quad \nabla f_y = \begin{bmatrix} f_{yx}(x, y) \\ f_{yy}(x, y) \end{bmatrix}$$

$$H = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$$

Hessian Matrix

The Hessian

$$f(x, y) = 2x^2 + 3y^2 - xy$$

$f_x(x, y)$ (red x)

- $f_{xx}(x, y)$ (red x) → 4
- $f_{xy}(x, y)$ (yellow y) → -1

$f_y(x, y)$ (yellow y)

- $f_{yx}(x, y)$ (red x) → -1
- $f_{yy}(x, y)$ (yellow y) → 6

$$H = \begin{bmatrix} 4 & -1 \\ -1 & 6 \end{bmatrix}$$

Hessian Matrix

	One Variable	Two Variable
Function	$f(x)$	$f(x, y)$
First Derivative	$f'(x)$ Rate of change of $f(x)$	$f_x(x, y)$ Rate of change w.r.t. x $f_y(x, y)$ Rate of change w.r.t. y $\nabla f = \begin{bmatrix} f_x(x, y) \\ f_y(x, y) \end{bmatrix}$
Second Derivative	$f''(x)$ Rate of change of the rate of change $f(x)$	$H(x, y) = \begin{bmatrix} f_{xx}(x, y) & f_{xy}(x, y) \\ f_{yx}(x, y) & f_{yy}(x, y) \end{bmatrix}$

Newton's Method for Two Variables

Newton's Method for One Variables

$$x_{k+1} = x_k - \frac{g'(x_k)}{g''(x_k)}$$

$$x_{k+1} = x_k - \boxed{g''(x_k)^{-1}} \boxed{g'(x_k)}$$

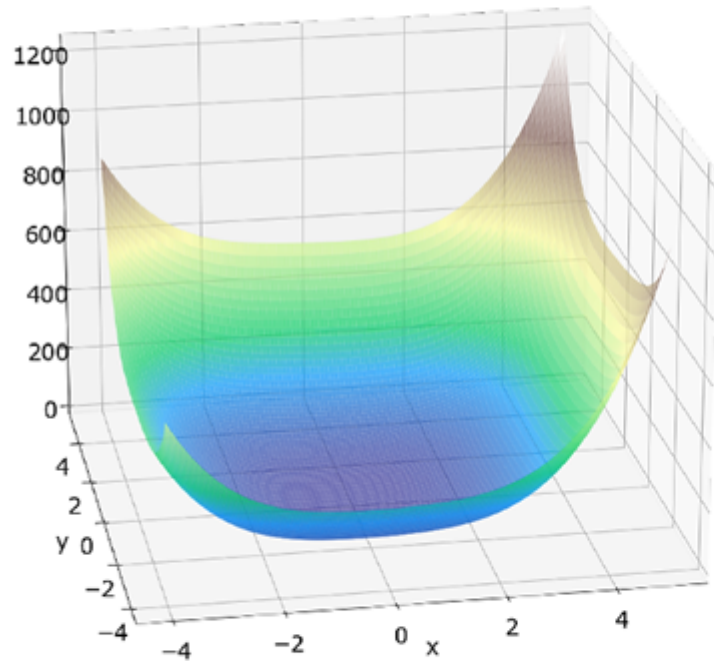
Newton's Method for Two Variables

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - H^{-1}(x_k, y_k) \times \nabla f(x_k, y_k)$$

$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \end{bmatrix} = \begin{bmatrix} x_k \\ y_k \end{bmatrix} - \begin{bmatrix} f_{xx}(x_k, y_k) & f_{xy}(x_k, y_k) \\ f_{yx}(x_k, y_k) & f_{yy}(x_k, y_k) \end{bmatrix}^{-1} \times \begin{bmatrix} f_x(x_k, y_k) \\ f_y(x_k, y_k) \end{bmatrix}$$

Newton's Method for Two Variables

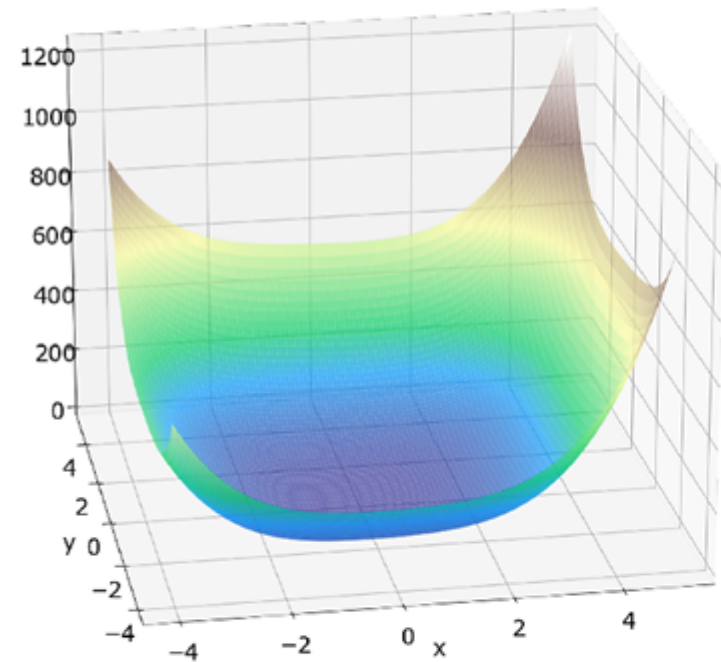
$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$



$$\begin{array}{l}
 \begin{array}{l} \textcolor{red}{x} \\ \textcolor{brown}{y} \end{array} \left\{ \begin{array}{l} 4x^3 + 8x - y - 0.4xy \\ 3.2y^3 + 4y - x - 0.2x^2 \end{array} \right. \\
 f(x, y) \\
 \begin{array}{l} \textcolor{red}{x} \\ \textcolor{brown}{y} \end{array} \left\{ \begin{array}{l} 12x^2 + 8 - 0.4y \\ -1 - 0.4x \end{array} \right. \\
 \begin{array}{l} \textcolor{red}{x} \\ \textcolor{brown}{y} \end{array} \left\{ \begin{array}{l} -1 - 0.4x \\ 9.6y^2 + 4 \end{array} \right.
 \end{array}$$

Newton's Method for Two Variables

$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y$$

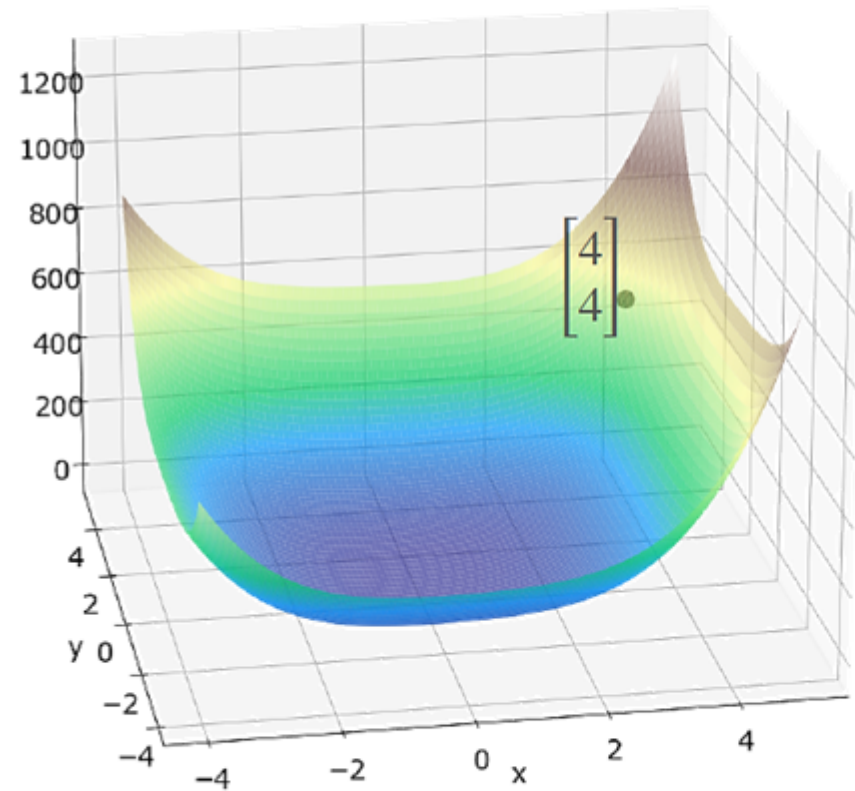


$$\nabla f(x, y) = \begin{bmatrix} 4x^3 + 8x - y - 0.4xy \\ 3.2y^3 + 4y - x - 0.2x^2 \end{bmatrix}$$

$$H(x, y) = \begin{bmatrix} 12x^2 + 8 - 0.4y & -1 - 0.4x \\ -1 - 0.4x & 9.6y^2 + 4 \end{bmatrix}$$

Newton's Method for Two Variables

$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y \quad \text{Start at } \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix}$$



$$\nabla f(4, 4) = \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix}$$

$$H(4, 4) = \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 4 \\ 4 \end{bmatrix} - \begin{bmatrix} 198.4 & -2.6 \\ -2.6 & 157.6 \end{bmatrix}^{-1} \times \begin{bmatrix} 277.6 \\ 213.6 \end{bmatrix}$$

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

Newton's Method for Two Variables

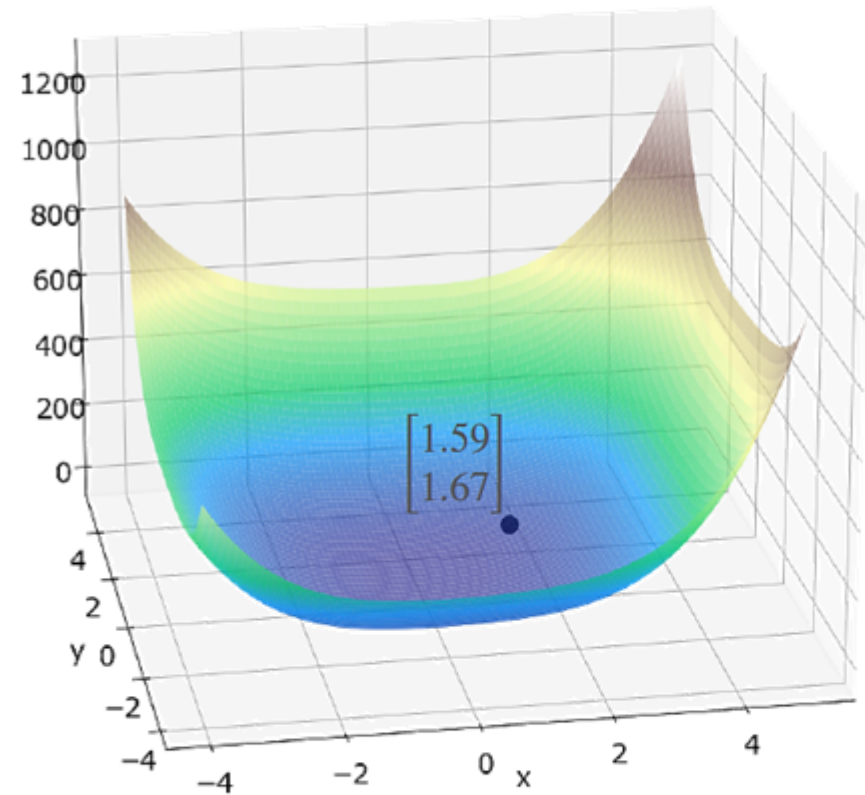
$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y \quad \text{Next } \begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix}$$

$$\nabla f(2.58, 2.61) = \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix}$$

$$H(2.58, 2.61) = \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}$$

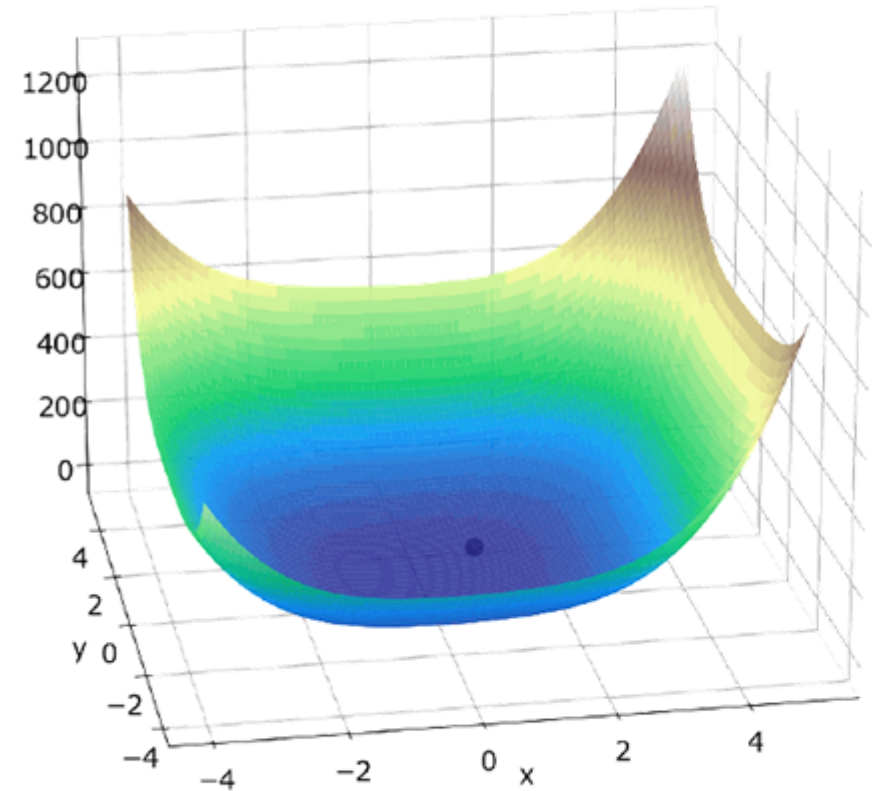
$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 2.58 \\ 2.61 \end{bmatrix} - \begin{bmatrix} 86.83 & -2.032 \\ -2.032 & 69.39 \end{bmatrix}^{-1} \times \begin{bmatrix} 84.25 \\ 63.4 \end{bmatrix}$$

$$\begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1.59 \\ 1.67 \end{bmatrix}$$



Newton's Method
for Two Variables

$$f(x, y) = x^4 + 0.8y^4 + 4x^2 + 2y^2 - xy - 0.2x^2y \quad \text{Next } \begin{bmatrix} x_2 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1.59 \\ 1.67 \end{bmatrix}$$



Repeat until you are close enough to the actual zero

It will take up eight steps to reach the minimum value

$$\begin{bmatrix} x_8 \\ y_8 \end{bmatrix} = \begin{bmatrix} 4.15 \cdot 10^{-17} \\ -2.05 \cdot 10^{-17} \end{bmatrix} \cong \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

