Data Science for Civil Engineering

Gradient Descent Method and Newtown's Method

Jintao Ke

Department of Civil Engineering
Faculty of Engineering
The University of Hong Kong, Pokfulam, Hong Kong
Email: kejintao@hku.hk

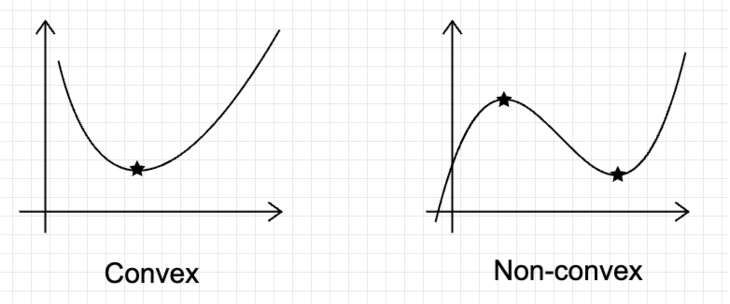
Table of Contents

Optimization Problems

To obtain the minimum value of a problem is to find $\mathbf{x}^*$, s.t. $f(\mathbf{x}) > f(\mathbf{x}^*), \forall \mathbf{x} \in \Omega^n$.

Gradient Descent

- For solving global minima/maxima for convex functions[1]
- For solving local minima/maxima for non-convex functions



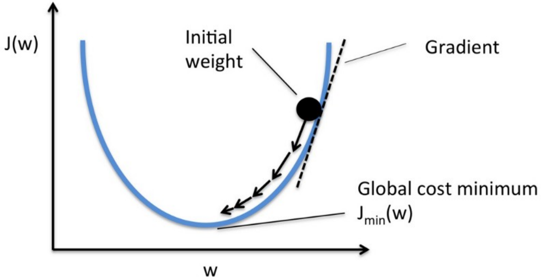
[1] A convex function $f(x)$ is that $\mu f(x_1) + (1-\mu) f(x_2) \geq f(\mu x_1 + (1-\mu) x_2)$, $\forall x_1, x_2 \in \Omega$.

Gradient Descent

- Update rules

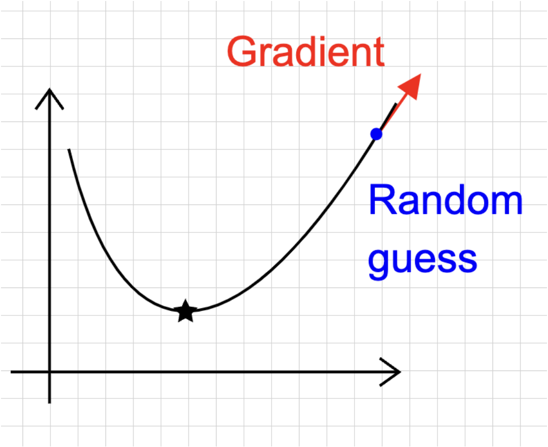
$$\mathbf{x}_{i+1} \leftarrow \mathbf{x}_i - \eta \nabla f(\mathbf{x}_i), i = 1, \cdots, n \tag{1}$$

- where η is the learning rate to control the speed of gradient descent.
- ∇ denote the gradient operator. Note that gradient always points to the direction that the function value increases.



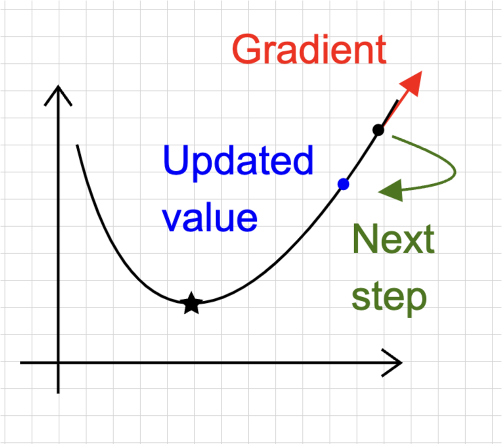
Gradient Descent

- Step 1, initialization



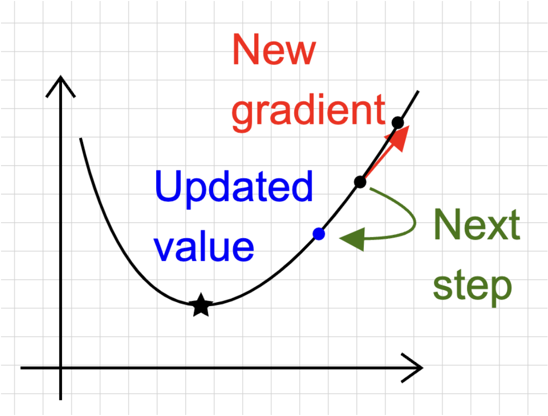
Gradient Descent

- Step 2, obtain gradient. Note that gradient always points to the direction that the function value increases.
- Step 3, update



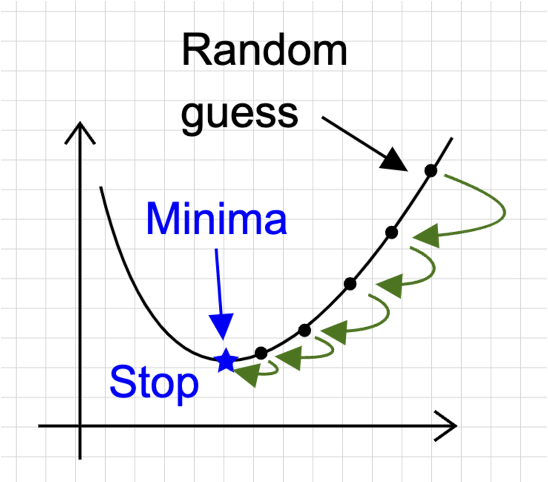
Gradient Descent

- obtain new gradient, update



Gradient Descent

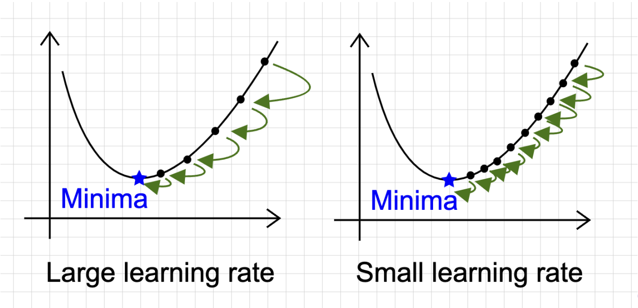
- repeat until "close" to minimum



Gradient Descent

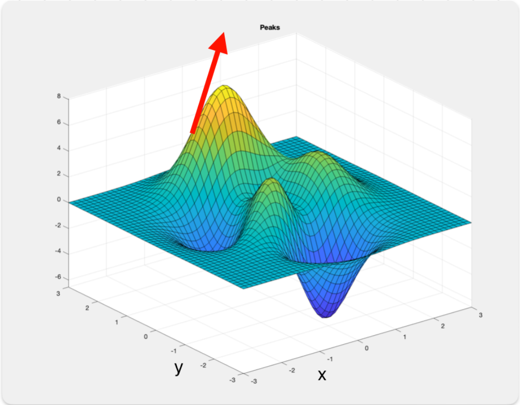
Summary

- Iterative. 1) random initialization 2) obtain gradient 3) update value according to gradient and learning rate 4) repeat until close to minima.
- The update direction is opposite to the gradient. That's why it is called gradient DESCENT.
- Different learning rate may result different descent speed.



Gradient Descent - Multi-variable

Consider function $f(x, y)$.

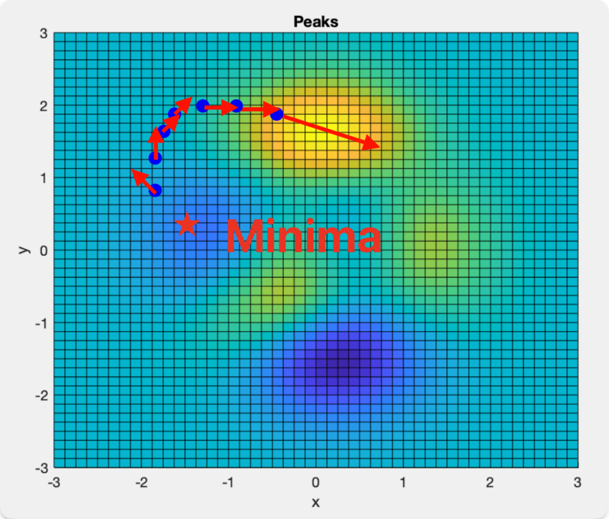


The gradient of a multi-variable function, $f(\cdot)$, is represented as $\nabla f(\cdot)$. In 2-variable case,

$$\nabla f(x, y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} \tag{2}$$

Gradient Descent - Multi-variable

Consider function $f(x, y)$.



Gradient Descent - Multi-variable - Example

Consider function $f(x,y) = (x-2)^2 + (y-2)^2$, $x, y \in \mathbb{R}$. Denote $x_{(i)}$ and $y_{(i)}$ the i^{th} iteration result.

Random Initialization

$x_0 = 1, y_0 = 1$. Learning rate $\eta = 0.1$

Obtain gradient

$\nabla f(x,y) = \begin{bmatrix} \frac{\partial f(x,y)}{\partial x} \\ \frac{\partial f(x,y)}{\partial y} \end{bmatrix} = \begin{bmatrix} 2x-4 \\ 2y-4 \end{bmatrix}$. Then $\nabla f(x_0, y_0) = \begin{bmatrix} -2 \\ -2 \end{bmatrix}$.

Update value according to gradient and learning rate

$$\begin{bmatrix} x_1 \\ y_1 \end{bmatrix} = \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} - \eta \times \nabla f(x,y) = \begin{bmatrix} 1 \\ 1 \end{bmatrix} - 0.1 \times \begin{bmatrix} -2 \\ -2 \end{bmatrix} = \begin{bmatrix} 1.2 \\ 1.2 \end{bmatrix}.$$

Repeat the above steps.

Newton's Method

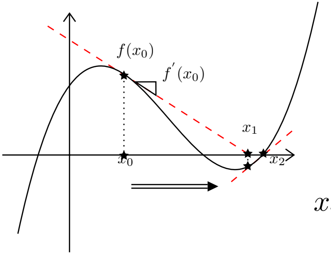
- Suppose we are solving an optimization problem

$$\min g(x) \tag{3}$$

- The optimal solution x^* should satisfy $g'(x^*) = 0$.
- Then the question becomes: how to find x^* such that $g'(x^*) = 0$?
- In other words, how to obtain one zero point of a differentiable function numerically?
- *Newton's method* could be useful.

Newton's Method

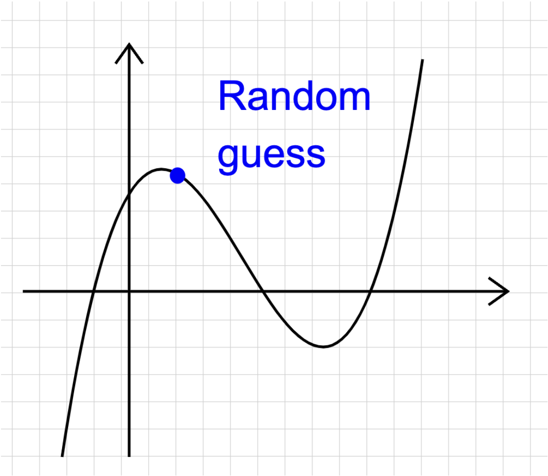
- Let $f(x) = g'(x)$, then we try using Newton's method to find the point x^* that satisfies $f(x^*) = g'(x^*) = 0$



$$
\begin{aligned}
x_1 &= x_0 - \frac{f(x_0)}{f'(x_0)} = x_0 - \frac{g'(x_0)}{g''(x_0)} \\
x_2 &= x_1 - \frac{f(x_1)}{f'(x_1)} = x_1 - \frac{g'(x_1)}{g''(x_1)} \\
&\vdots \\
x_{n+1} &= x_n - \frac{f(x_n)}{f'(x_n)} = x_n - \frac{g'(x_n)}{g''(x_n)}
\end{aligned}
$$

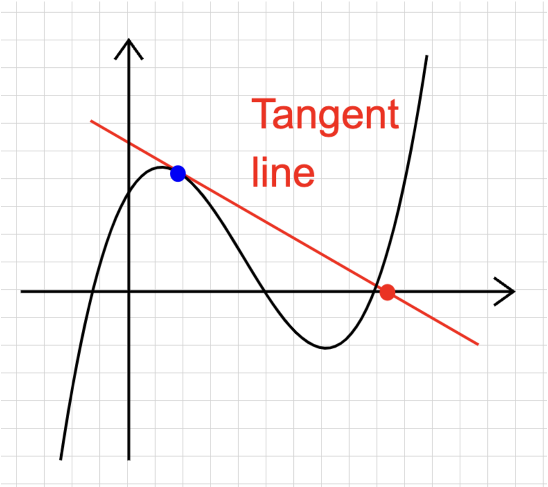
Newton's Method

- Step 1, random initialization



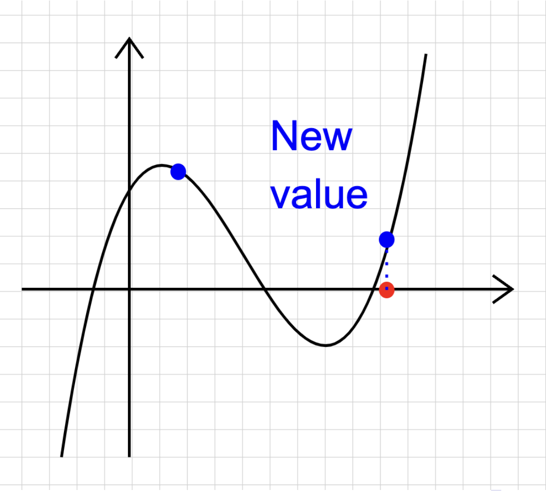
Newton's Method

- Step 2, take the derivative, obtain the zero point of the tangent line



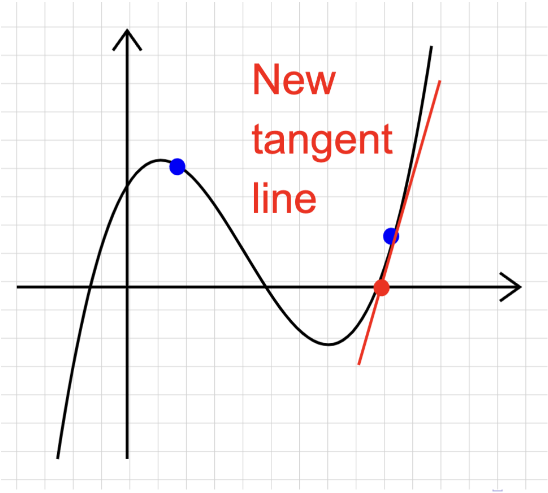
Newton's Method

- Step 3, get the new value, which is the function value of the corresponding zero point



Newton's Method

- Step 4, repeat: take the derivative again and obtain the new tangent line and zero point.



Newton's Method

Summary

- Iterative. 1) random initialization 2) derivative, get zero point of the tangent line 3) get new value 4) repeat until "close" to the zero.
- Fast. And no pre-specified learning rate needed.
- Can only solve for one zero. Different initials may result to different zeros.