Gradient descent algorithm: Since finding max and minima using differentiation is very hard computationaly, therefore we use gradient descent algo.

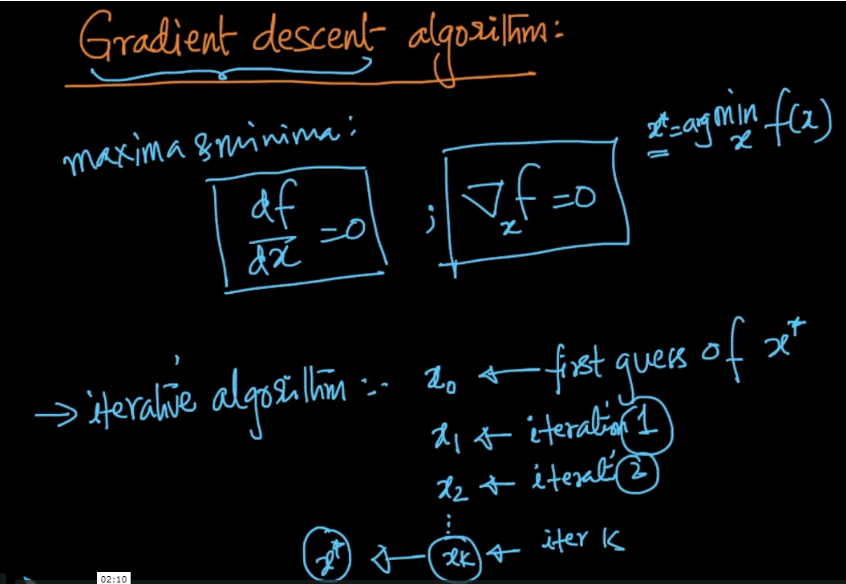
Let’s say we need to find x where f(x) is min.

Then Algo is as follows:

Make a guess for first value of x let’ say x0, calculate f(x0).

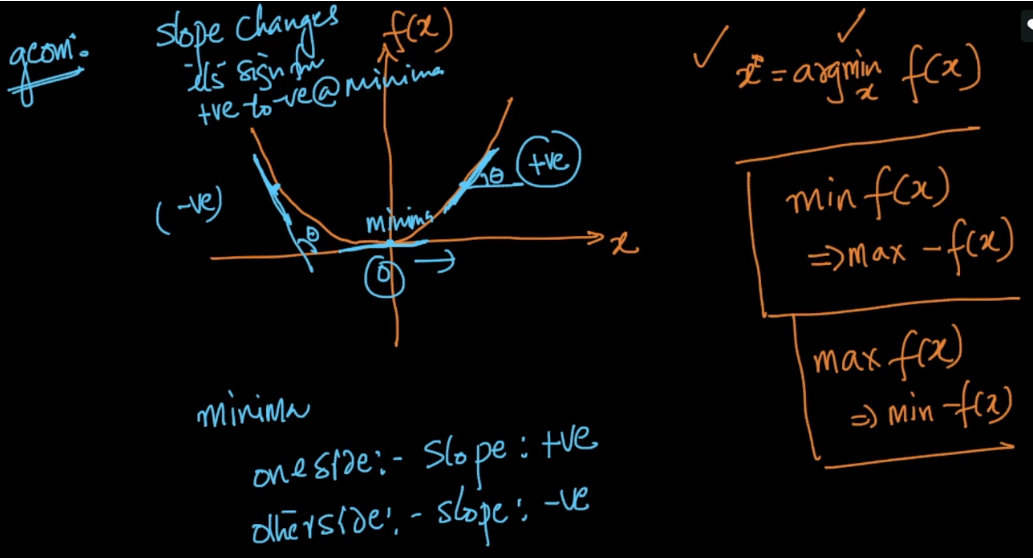
Now we’ll find x1 using x0, and calculate f(x1).

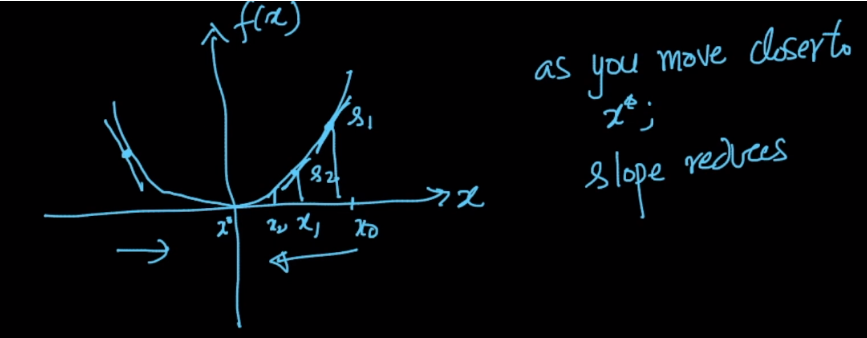
Repeat this process.



We can observe that:

* On one side of minima slope is +ve
* And on other side of minima slope is -ve.





**How do we find new x value to calculate f(x):**

X1 = x0 – r (df/dx)x=0

Using above formula we’ll find each new point, where  **r** is step size or learning rate and we are multiplying it with slope obtained at x=0.

Therefore, according to first guess value of x there can be two case.

Case 1: Let’s say x is guessed which is on right side of minima(as shown in below images).

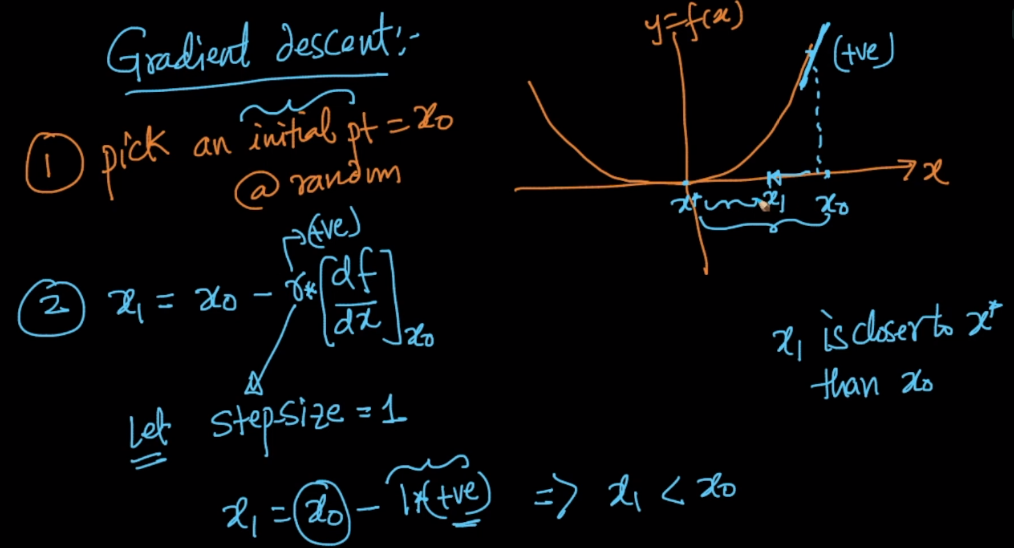
Slope will be +ve, therefore the new x point will be lesser than older one.

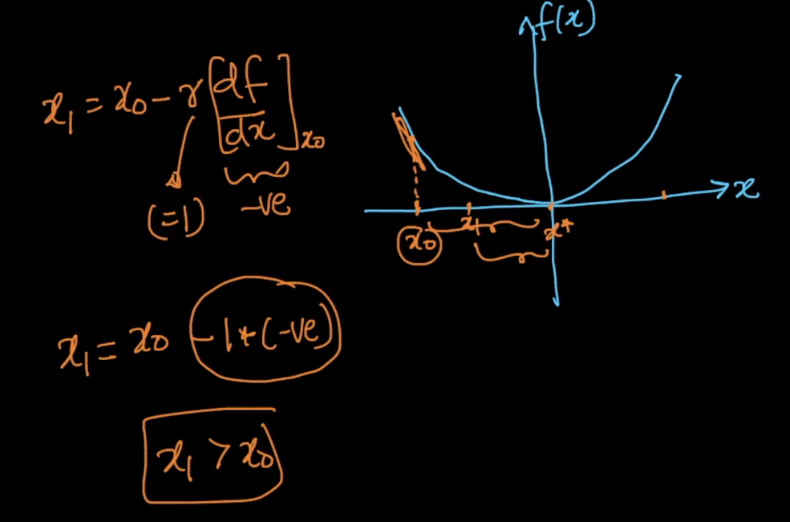
Case 2: Let’s say x is guessed which is on left side of minima(as shown in below images).

Slope will be -ve, therefore the new x point will be bigger than older one.

**At what point we’ll stop finding new x:**

Whenever diff between new x and old x becomes very small, ie xi+1 – xi = very small, then we stop iterating over new x, and call this as minima.





One more thing we can note here is that as we are moving towards minima, difference of jumps we make in moving from old x to new x becoming smaller. Why?

Let’s say x0 is on right side of minima, now each time we pick new x, the slope for that x gonna decrease as angle gonna decrease which eventually leads to smaller difference between new x and old x.

