mercoledì 27 settembre 2023 11:32

Ex $f(x,y) = (x+2y^3)^2$ $\qquad \nabla f(x,y) = \left[\frac{\partial f(x,y)}{\partial x} \quad \frac{\partial f}{\partial y}(x,y)\right]$

$$= \left[2(x+2y^3) \quad 2(x+2y^3)6y^2\right] = \left[2(x+2y^3) \quad 12y^2(x+2y^3)\right]$$

GRADIENT DESCENT

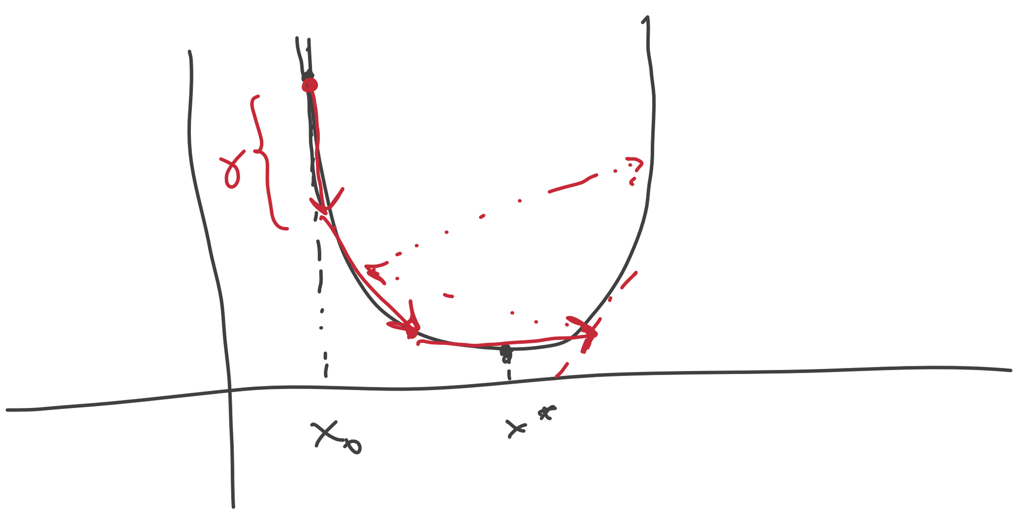
If we want to find a local optimum $f(x^*)$ of a function $f:\mathbb{R}^n \to \mathbb{R}$ we start with an initial guess x_0 of the ~~parameters~~ we wish to optimize and then iterate according to

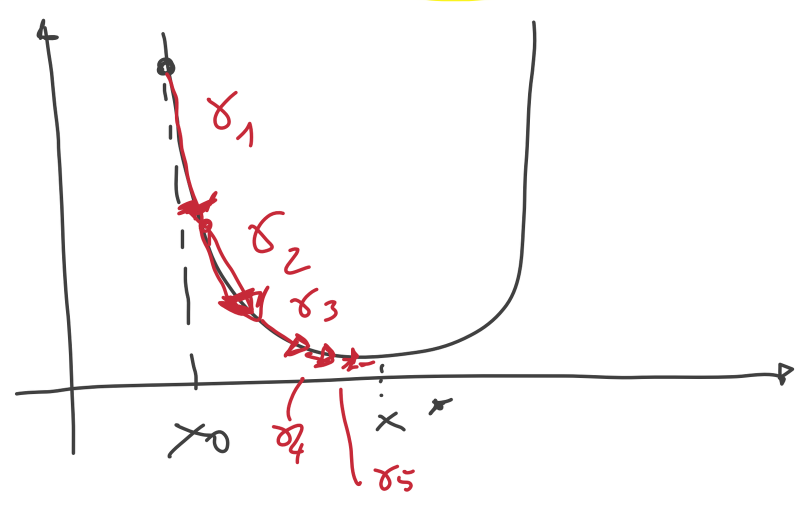
$$\boxed{x_{i+1} = x_i - \gamma \nabla f(x_i)} \quad (*)$$

where γ dictates how far to move along the gradient descent curve.

The choice of γ is critical for the performance. If γ is too small the process might take too long and if γ is too large we are in danger of overshooting.

In (*) γ is a fixed step size but it may also change at each iteration i.e. $x_{i+1} = x_i - \gamma_i \nabla f(x_i)$.





Ex (*) x_0 , $x_{i+1} = x_i - \gamma \nabla f(x_i)$

i = 0 $\quad x_1 = x_0 - \gamma \nabla f(x_0)$

i = 1 $\quad x_2 = x_1 - \gamma \nabla f(x_1)$ $\quad \longrightarrow$ Matlab code

i = 2 $\quad x_3$ ...

i = 3 $\quad x_4$ ...

i = 0, 1, 2, 3 ......

if $\nabla f(x_{i+1}) = 0$ ① ✓✗

? if $\|x_{i+1}\| > \|x_i\|$ ② NO

stop

$x_{fin} = x_{i+1}$

INPUT: x_0, ∇f, γ, itMax

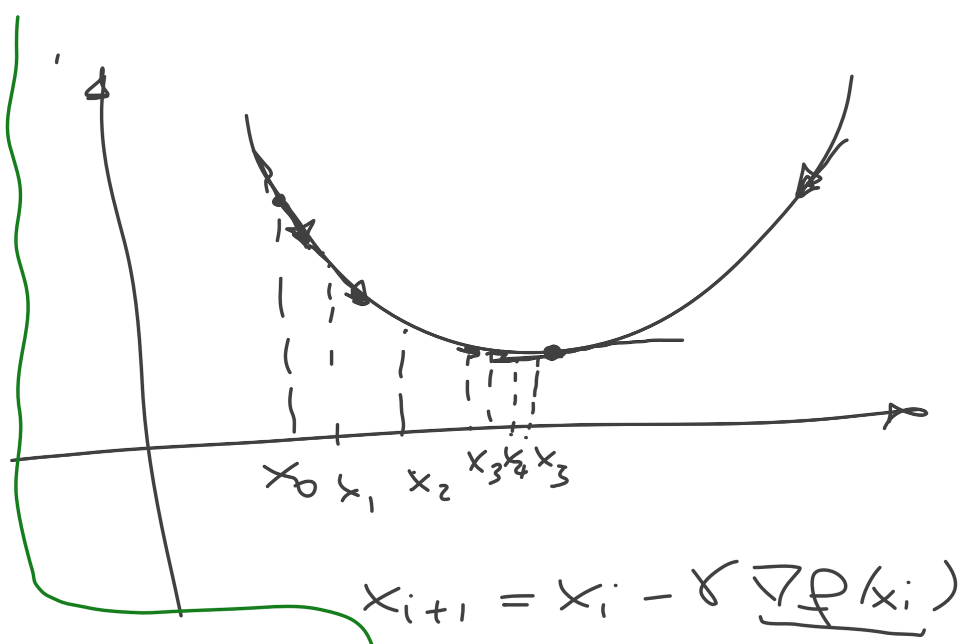
OUTPUT: x_{fin} ; $x_{fin} \simeq x^*$

STOPPING CRITERIA $\quad \varepsilon$ tolerance (10^{-5})

1) $\|\nabla f(x_{i+1})\| \le \varepsilon$ ~~= 0~~

2) $\|x_{i+1} - x_i\| < \varepsilon$ ✓

3) $i <$ itMax ✓



$x_i = x_0$; $x_{i+1} = x_i - \gamma * \nabla f(x_i)$; $i = 0$;

while ($\|x_{i+1} - x_i\| > \varepsilon$) && ($i <$ itMax) && $\|\nabla f(x_{i+1})\| > \varepsilon$

$\quad i = i+1$; $x_i = x_{i+1}$;

$\quad x_{i+1} = x_i - \gamma * \nabla f(x_i)$

end

$x_{i+1} = x_i - \gamma \underbrace{\nabla f(x_i)}_{small}$

$x_{i+1} \sim x_i$

$x_0 \to x_1 \to x_2 \to x_3$

i = 0

while

$x_1 = x$

$x_2 = x_1 - \gamma$ ..

Example. $f(x,y) = x^2 + 2y^2$, $x_0 = [4 \quad 3]$, $\gamma = 0.1$ $\quad (x^* = [0 \quad 0])$

$x_1 = x_0 - \gamma \nabla f(x_0) = [4 \quad 3] - 0.1\,[8 \quad 12] = [3.2 \quad 1.8]$

$\nabla f(x,y) = [2x \quad 4y]$

$\nabla f(x_0) = [2x_0(1) \quad 4x_0(2)]$

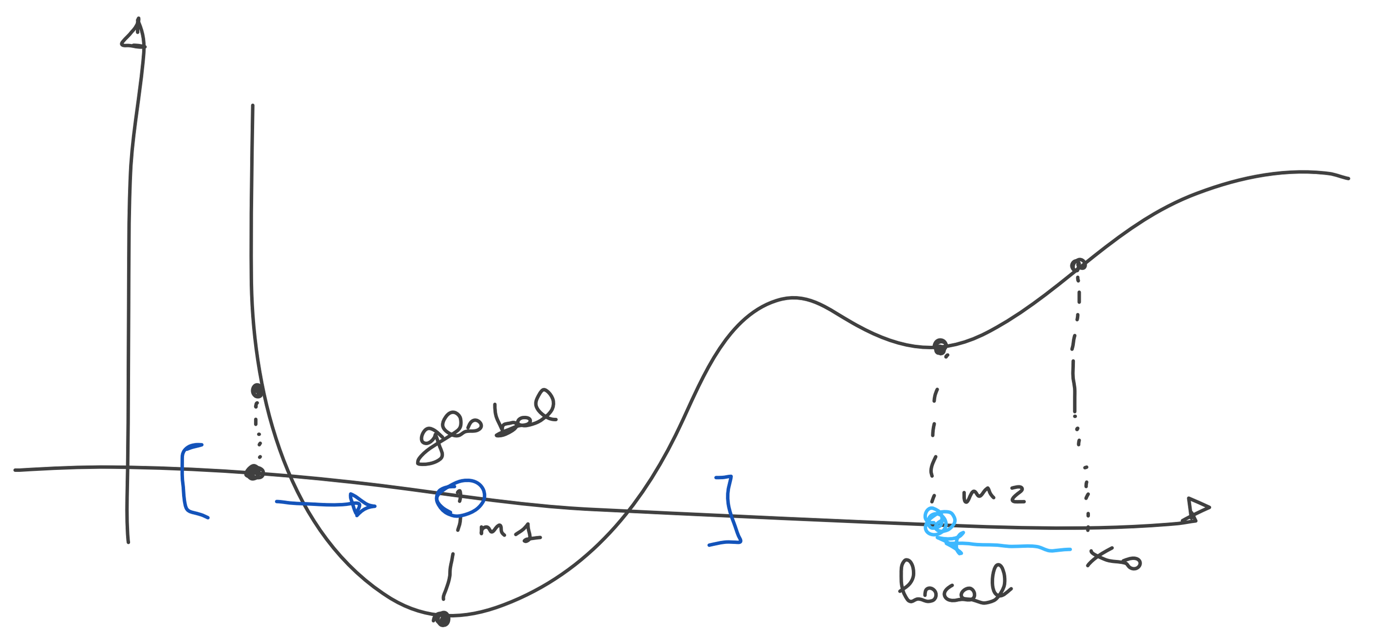
$x_2 = x_1 - \gamma \nabla f(x_1) = [3.2 \quad 1.8] - 0.1\,[6.4 \quad 7.2] = [2.56 \quad 1.08]$

⋮

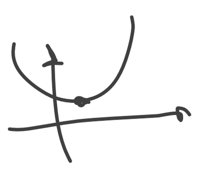
$x_{38} = [8 \cdot 10^{-4} \quad 1 \cdot 10^{-8}]$

• $\gamma = 0.5 \longrightarrow$ no convergence

Rk



Depending on the starting point x_0 we may reach m_1 or m_2

For CONVEX functions  all local minima are global minima

So these functions do not exhibit the tricky dependence on the starting point of the optimization algorithm.