Review

$f : \mathbb{R}^m \to \mathbb{R}$ $\quad x^* = \underset{x \in \mathbb{R}^m}{\arg\min} \, f(x)$

$\therefore$ $\nabla f(x^*) = 0$ and $-[\nabla f(x_0)]^T$ direction of steepest descent

$$\nabla f(x) := \left[\frac{\partial f}{\partial x_1}(x) \quad \frac{\partial f}{\partial x_2}(x) \quad \cdots \quad \frac{\partial f}{\partial x_m}(x) \right]_{1 \times m}$$

$$\frac{\partial f}{\partial x_i}(x_0) \approx \frac{f(x_0 + h e_i) - f(x_0 - h e_i)}{2h}$$ symmetric

$$\frac{\partial f}{\partial x_i}(x_0) \approx \frac{f(x_0 + h e_i) - f(x_0)}{h}$$ forward

$$\boxed{f(x) \approx \underbrace{f(x_0)}_{1 \times 1} + \underbrace{\nabla f(x_k)}_{1 \times m} (x - x_0)_{m \times 1}}$$

Gradient descent pseudo code

$k = 0$, tol = my value

x_0 given $\quad s > 0$ given

While $(\|\nabla f(x_k)\| > tol) \,\&\, (k < k_{max})$

$\quad x_k = x_k - s\left[\nabla f(x_k)\right]^T$

$\quad k = k + 1$

End

Why it works: $f(x_{k+1}) \approx f(x_k) + \nabla f(x_k) \underbrace{(x_{k+1} - x_k)}_{\Delta x_k}$

trust the approx.

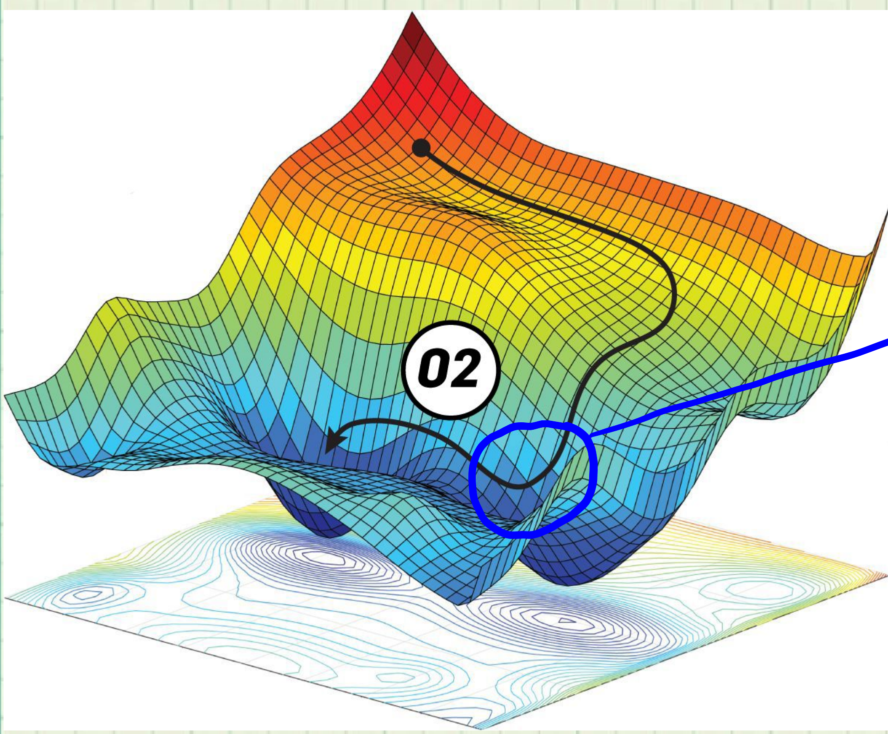
$$f(x_{k+1}) - f(x_k) < 0 \iff \nabla f(x_k)\,\Delta x_k < 0$$

$$\therefore \nabla f(x_k) \neq 0 \quad \text{and} \quad \Delta x_k = -s\left[\nabla f(x_k)\right]^T$$

$s > 0 \quad \Rightarrow \quad \nabla f(x_k)\,\Delta x_k = -s \underbrace{\nabla f(x_k)}_{1\times m} \underbrace{\left[\nabla f(x_k)\right]^T}_{m\times 1}$

$$= -s \left\| \left[\nabla f(x_k)\right]^T \right\|^2 < 0$$

Why it works? Linear Algebra!



Wishful thinking here? ☺

Today: Optimization as a Root Finding Problem!

Remind ourselves Newton-Raphson

$$g: \mathbb{R}^m \to \mathbb{R}^m$$

Seek $g(x^*) = 0$

$$\frac{\partial g(x_k)}{\partial x} \Delta x_k = -g(x_k)$$

solve for Δx_k

$$x_{k+1} = x_k + \Delta x_k$$

$$\frac{\partial g}{\partial x}(x_0) = \text{Jacobian at } x_0$$

$$\frac{\partial g}{\partial x}(x_0) = \left[\frac{\partial g}{\partial x_1}(x_0) \; \cdots \; \frac{\partial g}{\partial x_m}(x_0) \right]_{m \times m}$$

$$\frac{\partial g}{\partial x_i}(x_0) \approx \frac{g(x_0 + h e_i) - g(x_0 - h e_i)}{2h}$$

$$x^* = \underset{x \in \mathbb{R}^m}{\arg\min}\, f(x) \implies \nabla f(x^*) = 0_{1 \times m}$$

$$\iff [\nabla f(x^*)]^T = 0_{m \times 1}$$

Do root finding on $\underbrace{[\nabla f]^T}_{g} : \mathbb{R}^m \to \mathbb{R}^m$

We do the math

$$\frac{\partial}{\partial x}\left[\nabla f(x_k)\right]^T \Delta x_k = -\left[\nabla f(x_k)\right]^T$$

$$x_{k+1} = x_k + \Delta x_k$$

What is $\frac{\partial}{\partial x}\left[\nabla f(x)\right]^T$? Looks terrifying!

Hessian: $\frac{\partial}{\partial x}\left[\nabla f(x)\right]^T =: \underbrace{\nabla^2 f(x)}_{m\times m}$

$$\left[\nabla^2 f(x)\right]_{ij} = \frac{\left[f(x+he_i+\delta e_j) - f(x+he_i-\delta e_j) + -f(x-he_i+\delta e_j) + f(x-he_i-\delta e_j)\right]}{4h\delta}$$

Double for loop in Julia

Pseudo code, $s>0$

while $(\|\nabla f(x_k)\| > tol)$ & $(k < k_{max})$

$$\underbrace{\nabla^2 f(x_k)}_{m\times m}\,\underbrace{\Delta x_k}_{m\times 1} = -s\underbrace{\left[\nabla f(x_k)\right]^T}_{m\times 1}$$ solve for Δx_k

$X_{k+1} = X_k + \Delta X_k$

End

Global vs Local?

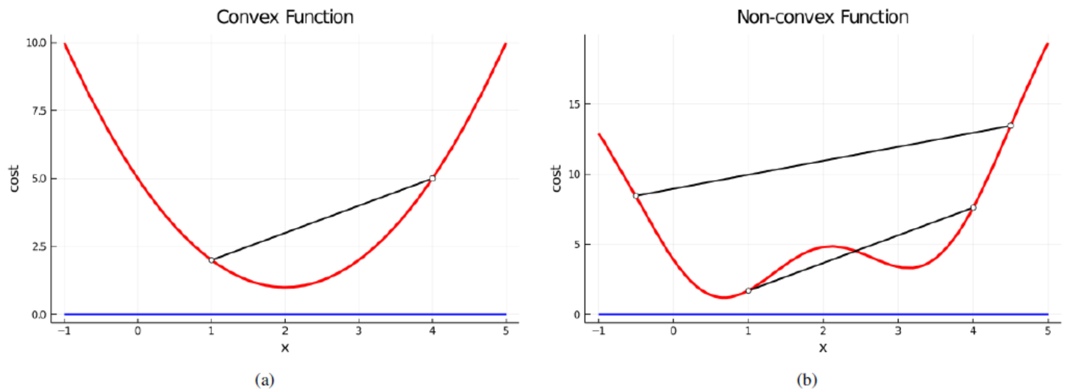


Figure 6: (a) This is a graph of our "simple" cost function with a global minimum at $x^* = 2.$, while (b) a graph of our "less simple" cost function, where there are two local minima, one at $x^* \approx 0.68$ and one at $x^* \approx 3.14$, as well as a local maximum at ≈ 2.1. In each case, we have overlaid line segments that are used to check for the mathematical property called **convexity**. A convex function only always has global minima. It does not have any local minima.

Convexity: Whole new area that is super hot called Convex optimization