For convex functions defined on convex sets, local minima are global minima

Theorem: Consider the opt. problem

$$\min_{x \in \Omega} f(x)$$

where $f: \Omega \to \mathbb{R}$ is a convex function & Ω is a convex set

then if x^* is a local min it is also a global min.

Proof: Suppose x^* is a local minimizer that is not a global minimizer

$$\Rightarrow \exists w \in \Omega : \underbrace{f(w) < f(x^*)}_{(1)}$$

Since f is convex, $\forall \alpha \in [0,1]$

$$f(\alpha w + (1-\alpha)x^*) \leq \underbrace{\alpha f(w) + (1-\alpha) f(x^*)}$$

def'n of convex fts

$$\text{by } ① \longrightarrow \; < \alpha f(x^*) + (1-\alpha) f(x^*)$$

$$= f(x^*)$$

In short:

$$f(\alpha w + (1-\alpha)x^*) < f(x^*)$$

but this essentially says that every pt. z on the line segment connecting w to x^* has

$$f(z) < f(x^*)$$

In particular z can be arbitrarily close to x^*, so x^* cannot be a local minimizer.

This is a contradiction arising from our supposing that x^* was not a global min.

So we conclude that x^* is a global min.

$\boxtimes$

The above theorem guarantees us that local mini of convex functions on convex sets are global.

We still need algorithms to find them.

So we will need some concepts & result from vector calculus

Gradients, Hessians, Taylor's theorem

Defn:

For $f: \mathbb{R}^n \to \mathbb{R}$ the gradient

$$(\nabla f)(x) = \nabla f\big|_x = \nabla f(x)$$

is defined as

same dimensions as x, but often row vector

$$\boxed{\begin{aligned}\nabla f(x) &= \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}, \dots, \frac{\partial f}{\partial x_n}\right)\\ &= \left(\frac{\partial f}{\partial x_i}\right)_{i=1}^{n}\end{aligned}}$$

COL
UMN
VEC
TOR

and the Hessian is

$$\underbrace{(\nabla^2 f)(x)}_{n \times n \text{ matrix}} = \left(\frac{\partial^2 f}{\underbrace{\partial x_i \partial x_j}_{\text{its entries}}}\right)_{\substack{i=1,\dots,n \\ j=1,\dots,n}}$$

Example: For $f: \mathbb{R}^2 \to \mathbb{R}$

given by

$$f(x_1, x_2) = 2x_1^2 + x_2^4 - 2x_2^2 + 1$$

$$\nabla f(x_1, x_2) = (4x_1, 4x_2^3 - 4x_2)$$

$$\nabla^2 f(x_1, x_2) = \begin{pmatrix} 4 & 0 \\ 0 & 12x_2^2 - 4 \end{pmatrix}$$

Directional derivatives

Recall that a partial derivative of say $f: \mathbb{R}^3 \to \mathbb{R}$ is given by

$$\frac{\partial f}{\partial x_1} = \lim_{h \to 0} \frac{f(x_1 + h, x_2, x_3) - f(x_1, x_2, x_3)}{h}$$

$$= \lim_{h \to 0} \frac{f(\vec{x} + h\vec{e}) - f(\vec{x})}{h}$$

So $\frac{\partial f}{\partial x_1}$ is the rate of change of f in the direction of $\vec{i}$.

If on the other hand, we wanted the rate of change in the direction of a <u>unit</u> vector $\vec{v}$, we could compute

$$\boxed{\lim_{h \to 0} \frac{f(\vec{x} + h\vec{v}) - f(\vec{x})}{h}}$$

→ directional derivative of f in the direction of the <u>unit</u> vector $\vec{v}$.

Directional derivatives & gradients

Theorem: The directional derivative of a differentiable function $f: \mathbb{R}^n \to \mathbb{R}$ in the direction of the unit vector v is

$$\langle \nabla f, v \rangle = \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \cdot v_i$$

Example: $f(x_1, x_2, x_3) = x_1^2 e^{-x_2 x_3}$

Find the rate of change of f in the direction of $\vec{v} = \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$

at $(1,0,0)$

Sol'n: The question requires us to compute

$$\langle \nabla f, v \rangle \Big|_{(1,0,0)}$$

$$\nabla f = \left(2x_1 e^{-x_2 x_3}, -x_1^2 x_3 e^{-x_2 x_3}, -x_1^2 x_2 e^{-x_2 x_3}\right)$$

$$\Rightarrow \langle \nabla f, v \rangle \Big|_{(1,0,0)} = \left\langle (2,0,0), \left(\tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}, \tfrac{1}{\sqrt{3}}\right) \right\rangle$$

$$= \frac{2}{\sqrt{3}}.$$

Remark: Make sure to check that $\|v\| = 1$; otherwise you need to replace v by $\frac{v}{\|v\|}$.

$\boxed{\text{The gradient is in the direction of greatest increase of the function}}$

Why is this true?

Suppose $\vec{n}$ is such that $\|n\| = 1$ then the rate of change of f in the direction of n is

$$\langle \nabla f, n \rangle = \|\nabla f\| \underbrace{\|n\|}_{1} \cos\theta$$

angle between ∇f & n

So if $\nabla f \neq 0$, $\langle \nabla f, n \rangle$ is greatest when $\theta = 0$, i.e. when $\vec{n}$ & ∇f are in the same direction.

Example $f(x_1, x_2) = x_1^2 + x_2^3$

The direction of greatest increase of $f(x_1, x_2)$ at $p = (2,3)$ is

$$\nabla f|_p = \left(\frac{\partial f}{\partial x_1}, \frac{\partial f}{\partial x_2}\right)\Big|_p$$

$$= (2x_1, 3x_2^2)\Big|_{(2,3)}$$

$$= (4, 27)$$

a unit vector in that direction is $(4,27)/\sqrt{4^2+27^2}$

Calculus + Linear Algebra:

Often $f: \mathbb{R}^n \to \mathbb{R}^m$. $f(x) = [f_1(x), f_2(x), \dots, f_m(x)]$

$$\frac{\partial f}{\partial x} = \begin{bmatrix} \nabla f_1 \\ \vdots \\ \nabla f_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix}$$

Ex: $f(x) = Ax$.

$$\left.\begin{aligned} f_i(x) &= \sum_{j=1}^{n} A_{ij} x_j \\ \frac{\partial f_i}{\partial x_j} &= A_{ij} \end{aligned}\right\} \Longrightarrow \frac{\partial f}{\partial x} = A$$

Ex: $f(x) = x^T A y$ note: $x^T A y = y^T A^T x$ b/c its scalar

$$\nabla f = \frac{\partial}{\partial x}\left(y^T A^T x\right) = y^T A^T$$

Ex: $f(x) = x^T A x$

$\nabla f = x^T (A + A^T)$ via product rule

$$\nabla^2 f_{ij} = \frac{\partial}{\partial x_j}\left(x^T(A + A^T)\right)_i = \frac{\partial}{\partial x_j}\left(\sum_k x_k (A_{ki} + A_{ik})\right)$$

$$\nabla^2 f = A + A^T \qquad = A_{ij} + A_{ji}$$

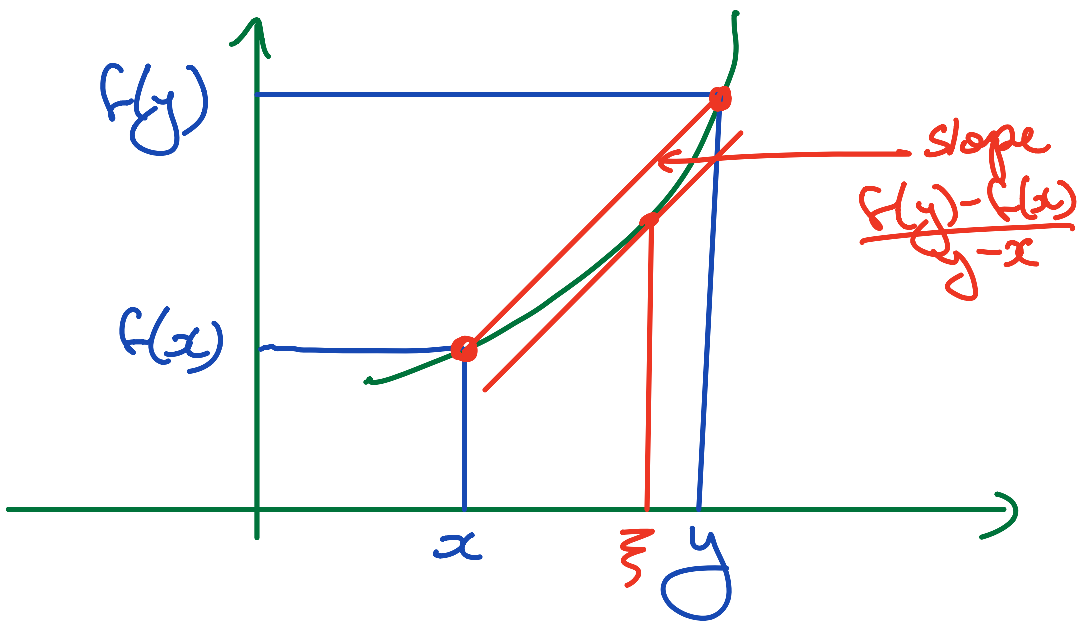
Taylor's theorem:

First, recall Taylor's thm for functions of a single variable

$$f(y) = f(x) + f'(\xi)(y-x)$$

$\xi \in (x,y)$

$$\left(\Leftrightarrow \frac{f(y)-f(x)}{y-x} = f'(\xi) \right)$$



More generally, in one variable

$$f(y) = f(x) + f'(x)(y-x) + \frac{f''(x)(y-x)^2}{2!} + \cdots + f^{(n)}(\xi)\frac{(y-x)^n}{n!}$$

$\xi \in (x,y)$

What if $f: \mathbb{R}^n \to \mathbb{R}$?

For now, it suffices for our purposes to state the following version.

1st order Taylor series approx of f around the point x.

If $f: \mathbb{R}^n \to \mathbb{R}$ is twice continuosly differentiable

$$f(y) = f(x) + \nabla f(x)^T (y-x) + \frac{1}{2}(y-x)^T \nabla^2 f\big|_{\xi} (y-x)$$

here ξ is a point on the line segment connecting x & y

so $\xi = ty + (1-t)x$

where $t \in (0,1)$

For 1 variable

$$f(y) = f(x) + f'(x)(y-x) + f''(\xi)(y-x)^2$$