Lecture 7

- Gradients
- Hessians
- The main theorem (Varaiya)

- we have a fcn $f(x): \mathbb{R} \to \mathbb{R}$
  - derivative: how f changes wrt changes in x (describes changes to the fcn as the variable changes)

Taylor's Thm:

- Let $x_0 \in \mathbb{R}$ be a fixed point & Δx is some variation on x, then we can write the fcn at a point close/local to x_0 as a fcn of x_0 & a correction term

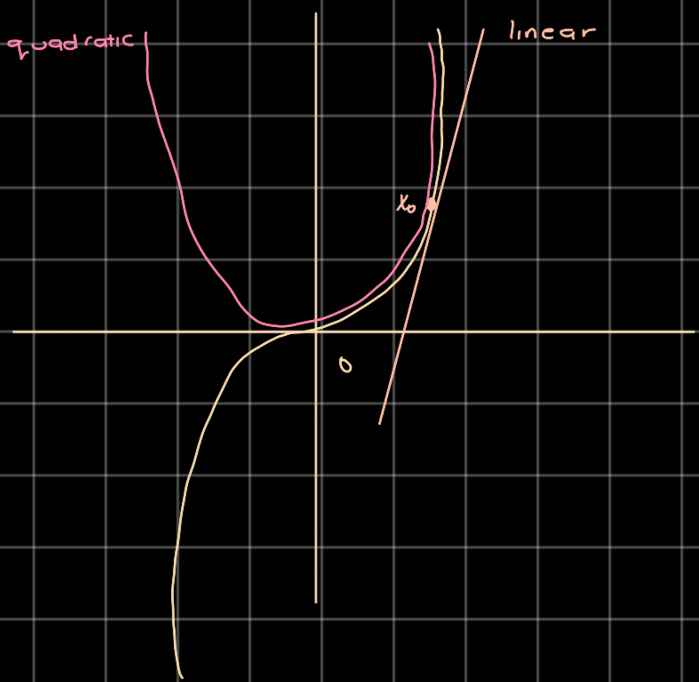
fixed

$$F(x_0 + \underline{\Delta x}) = F(x_0) + \frac{dF}{dx}\bigg|_{x=x_0} \Delta x + \frac{1}{2!}\frac{d^2F}{dx^2}(\Delta x)^2 + \dots + (\text{higher order terms})$$

variable

↳ way of approximating $F(x)$ as a polynomials

eg: $f(x) = x^3$



Taylor's Thm for Vectors

$f(\vec{x}) : \underbrace{\mathbb{R}^n \to \mathbb{R}}$

eg: $f(\vec{x}) = \|\vec{x}\|_2^2$

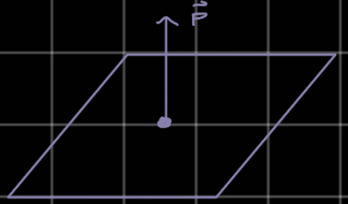
$$f(\vec{x}_0 + \Delta\vec{x}) = \boxed{f(\vec{x}_0) + \underbrace{\overset{\text{gradient}}{\nabla f^T}\Big|_{\vec{x}=\vec{x}_0}}_{\substack{\text{must yield} \\ \text{a scalar} \\ \text{(HAS to be} \\ \text{a row vector)}}} \Delta\vec{x}} + \frac{1}{2!}(\Delta\vec{x})^T \underbrace{\nabla^2 f}_{\substack{\text{must be a matrix} \\ \text{HESSIAN}}}\Big|_{\vec{x}=\vec{x}_0} \Delta\vec{x}$$

First order approximation

↳ if we want something in the generic quadratic form (for vectors), it's in the form $\vec{x}^T A \vec{x}$

in the form $\vec{p}^T \vec{x} = \vec{b}$, which gives us a hyperplane

↳ generalization of a line & the quadratic 2nd order term gives you the generalization of a parabola



$\vec{x}_0 + \Delta\vec{x} = \vec{a}$

$\Delta\vec{x} = \vec{a} - \vec{x}_0$

Gradient

$\nabla f(\vec{x})$ captures change according to all components of $\vec{x}$.

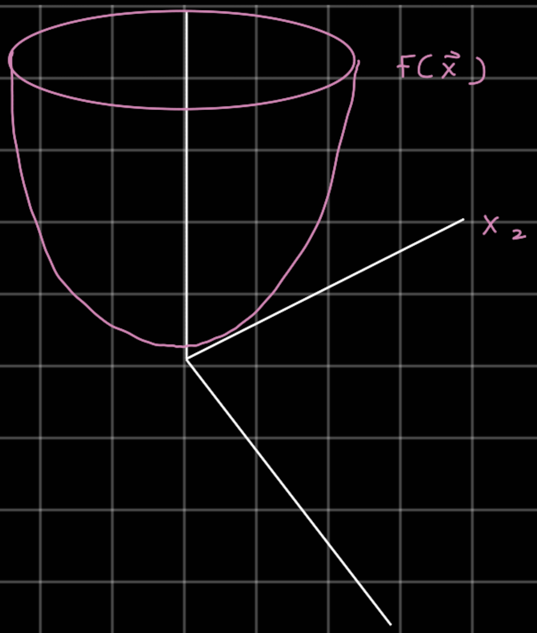
$$\nabla f(\vec{x}) = \left[\frac{\partial f}{\partial x_1} \quad \frac{\partial f}{\partial x_2} \quad \cdots \quad \frac{\partial f}{\partial x_n} \right]^T$$

Hessian

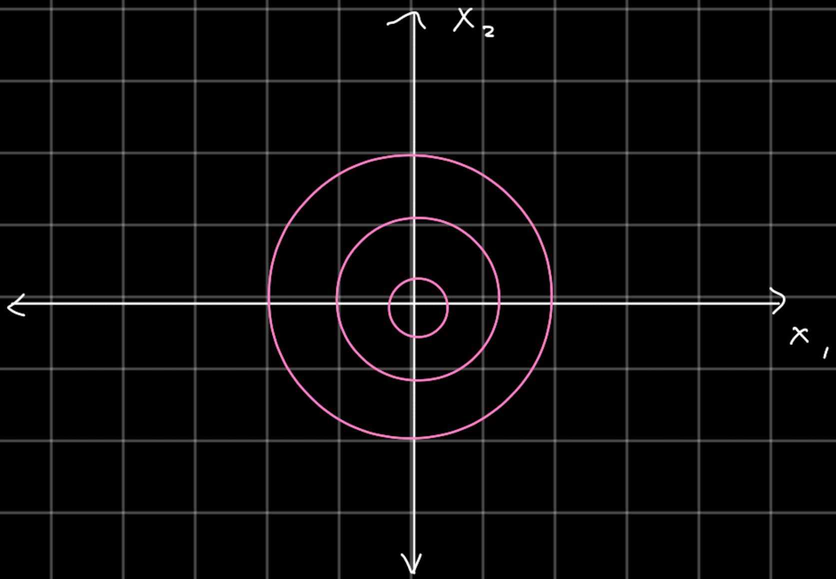
$$\nabla f^2(\vec{x})_{ij} = \frac{\partial f}{\partial x_i \partial x_j}$$

↳ often symmetric (bc sometimes the $\partial x_i \partial x_j$ can be interchanged)

Example: $f(\vec{x}) = \|\vec{x}\|_2^2 \qquad f: \mathbb{R}^2 \to \mathbb{R}$



Levels sets: if $f(\vec{x})$ is a constant, what are the values that can be taken



$$\nabla f(\vec{x}) = \nabla \|x\|_2^2$$
$$= \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} = 2\vec{x}$$

$$\nabla^2 f(\vec{x}) = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}$$

$$f(\vec{x} + \Delta\vec{x}) = f(x_1 + \Delta x_1,\ x_2 + \Delta x_2) = f(\vec{x}) + \nabla f\Big|_{x=x}^T \Delta x + \frac{1}{2!}\Delta x^T \nabla^2 f\Big|_{x=x} \Delta x$$

$\vec{x} = (x_1, x_2)$

$\Delta\vec{x} = (\Delta x_1, \Delta x_2)$

$$= \|x\|_2^2 + (2\vec{x})^T \Delta\vec{x} + \frac{1}{2!}\Delta\vec{x}^T \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} \Delta\vec{x}$$

$$= x_1^2 + x_2^2 + 2\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}[\Delta x_1 \ \ \Delta x_2] + \frac{1}{2!}[\Delta x_1\ \Delta x_2]\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix}$$

$$= x_1^2 + x_2^2 + 2x_1\Delta x_1 + 2x_2\Delta x_2 + \Delta x_1^2 + \Delta x_2^2$$

$$= (x_1 + \Delta x_1)^2 + (x_2 + \Delta x_2)^2$$

↳ no error in our approximation bc it's a quadratic fcn (best approximation for a quadratic is a quadratic) examples

(1) $f(\vec{x}^T\vec{a}) = \sum_{i=1}^{n} x_i a_i$

$$\nabla f(\vec{x}^T\vec{a}) = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = \vec{a}$$

(2) $f(\vec{x}) = \vec{x}^T A \vec{x}$

$$f(\vec{x}) = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix} \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ a & \ddots & \vdots \\ \vdots & & \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

$$\cancel{= x_1(a_{11} + a_{12} \cdots + a + a_{11})}$$

$$= \sum_i \sum_j x_i a_{ij} x_j$$

→ terms that have x_i :

$$\sum x_i a_{ij} \cancel{x_j} + \sum x_j a_{ji} x_i + x_i^2 a_{ii}$$

$$\frac{\partial F}{\partial x_i} = \sum_{j \neq i} a_{ij} x_j + \sum_{j \neq i} x_j a_{ij} + 2 x_i a_{ii}$$

$$\frac{\partial F}{\partial x_i} = \sum_j (a_{ij} + a_{ji}) x_j$$

$$\nabla f(\vec{x}) = (A + A^T) \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$\nabla^2 F = \frac{\partial F}{\partial x_j}\left(\sum_j (a_{ij} + a_{ji}) x_j \right)$$

$$= \sum_j a_{ij} + a_{ji}$$

$$= A + A^T$$

The main theorem: $f : \mathbb{R}^n \to \mathbb{R}$, differentiable everywhere

Consider the optimization problem :

$$\underset{\vec{x} \in \Omega}{\text{minimize}} \quad f(\vec{x}) \quad , \quad \Omega : \text{open set} \subseteq \mathbb{R}^n$$

↳ doesn't incl. its boundary

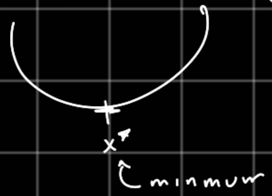
Then if $\vec{x}^*$ is an optimal solution then

$$\frac{\partial F}{\partial x}(x^*) = 0$$

↳ the optimum value is found when the derivative is 0.

of main thm using
Proof: Taylor's Thm

① $f(x^* + \Delta x) = f(x^*) + \frac{dF(x^*)}{dx}\Delta x + \underbrace{O(\Delta x)}_{\text{higher order terms}}$



② $f(x^*) \leq f(x^* + \Delta x)$

$\varepsilon > 0$ s.t. $|x^* - x| < \varepsilon$

$x \in \Omega$

Choose: $0 < \Delta x < \varepsilon$

sub ① into ②

$$f(x^*) \leq f(x^*) + \frac{dF(x^*)}{dx}\Delta x + \underbrace{O(\Delta x)}_{h.o.t}$$

$$0 \leq \frac{dF}{dx}(x^*)\Delta x + h.o.t$$

Divide by Δx

$$0 \leq \frac{dF}{dx}(x^*) + (\text{const})\frac{(\Delta x)^2}{\Delta x} + (\text{cost})\frac{(\Delta x)^2}{\Delta x}$$

$$\lim_{\Delta x \to 0}$$

$$0 \leq \frac{dF(x^*)}{dx}$$

↳ redo proof w/ $(x^* - \Delta x)$ instead and you get $\lim_{\Delta x \to 0}$ $0 \geq \frac{dF(x^*)}{dx}$ → so $dF(x^*)/dx$ must be equal to 0.