(1)

Lecture 6 EE 127/227AT, Feb 6, 2020

Vector Calculus

- Taylor's theorem (functions of scalars)
- Gradient & Hessian (functions of vectors)
- Taylor's theorem (Vector)
- Least Squares revisited
- Matrix Inner Product (Not covered)
- Gradient (functions of matrices) (Not covered).

Admin

HW 2 due Friday

HW 1 self grade due Friday.

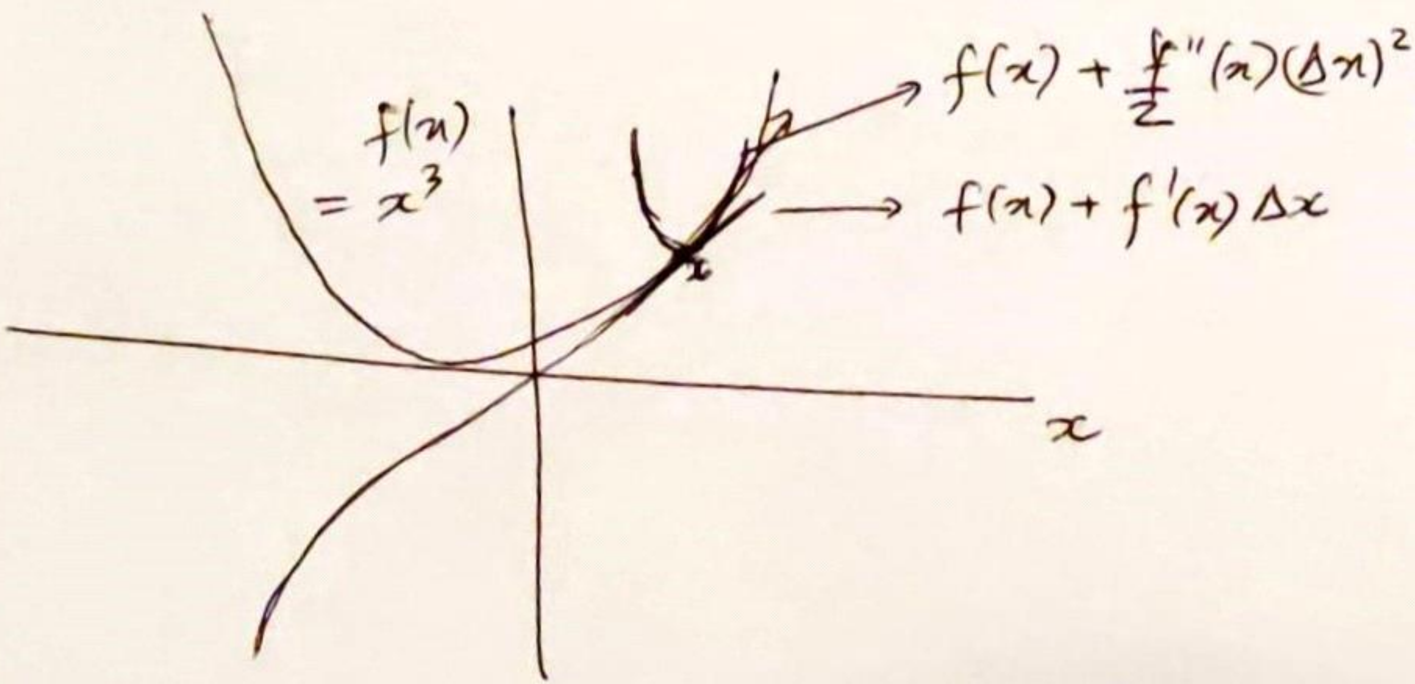
② Vector calculus local behavior of functions.

$f: \mathbb{R} \to \mathbb{R}$ $f(x) = x^3$.

Taylor's theorem (scalar version).

$$f(x+\Delta x) = f(x) + f'(x)\Delta x + \frac{1}{2!} f''(x)(\Delta x)^2 +$$ higher order terms.

$$f'(x) = \frac{\partial f}{\partial x}(x) \qquad f''(x) = \frac{\partial}{\partial x} f'(x) = \frac{\partial^2 f}{\partial x^2}(x).$$



③ Functions of vectors

$f : \mathbb{R}^n \rightarrow \mathbb{R}$ $\qquad \vec{x} \in \mathbb{R}^n \qquad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$

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

Derivative

$$\frac{\partial f(\vec{x})}{\partial \vec{x}} = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\vec{x}) , & \frac{\partial f}{\partial x_2}(\vec{x}) & \cdots & \frac{\partial f}{\partial x_n}(\vec{x}) \end{bmatrix}$$

Row vector.

Gradient

$$\nabla_{\vec{x}} f(\vec{x}) = \begin{bmatrix} \frac{\partial f}{\partial x_1}(\vec{x}) \\ \frac{\partial f}{\partial x_2}(\vec{x}) \\ \vdots \\ \frac{\partial f}{\partial x_n}(\vec{x}) \end{bmatrix}.$$

Hessian

$$\nabla^2_{\vec{x}} f(\vec{x}) = \begin{array}{c c} & n \\ n & \begin{bmatrix} \frac{\partial^2 f}{\partial x_1^2}(\vec{x}) & \frac{\partial^2 f}{\partial x_1 \partial x_2} & \\ \frac{\partial^2 f}{\partial x_2 \partial x_1} f(\vec{x}) & & \\ & & \end{bmatrix} \end{array}$$

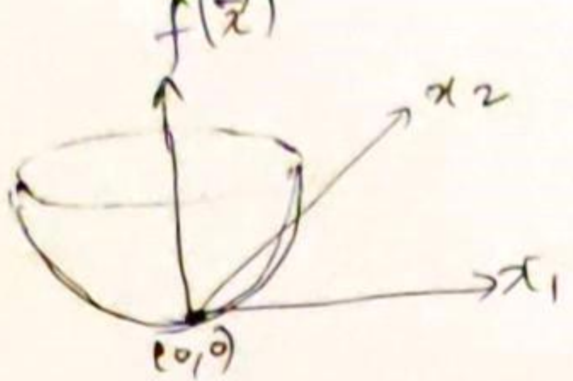
$$\left(\nabla^2_{\vec{x}} f(\vec{x})\right)_{11} = \frac{\partial}{\partial x_1} \left(\nabla f(\vec{x})\right)_1 = \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_1} f(\vec{x}) = \frac{\partial^2 f(\vec{x})}{\partial x_1^2}$$

$$\left(\nabla^2_{\vec{x}} f(\vec{x})\right)_{21} = \frac{\partial}{\partial x_2} \left(\nabla f(\vec{x})\right)_1 = \frac{\partial}{\partial x_2} \frac{\partial}{\partial x_1} f(x) = \frac{\partial^2 f(\vec{x})}{\partial x_2 \partial x_1}$$

$$\left(\nabla^2_{\vec{x}} f(\vec{x})\right)_{12} = \frac{\partial}{\partial x_1} \left(\nabla f(\vec{x})\right)_2 = \frac{\partial}{\partial x_1} \frac{\partial}{\partial x_2} f(x) = \frac{\partial^2 f(\vec{x})}{\partial x_1 \partial x_2}$$

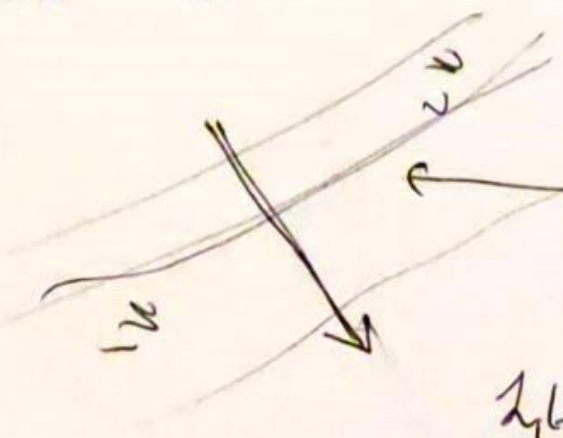
④ Example: $\vec{x} \in \mathbb{R}^2$ $\quad \vec{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$

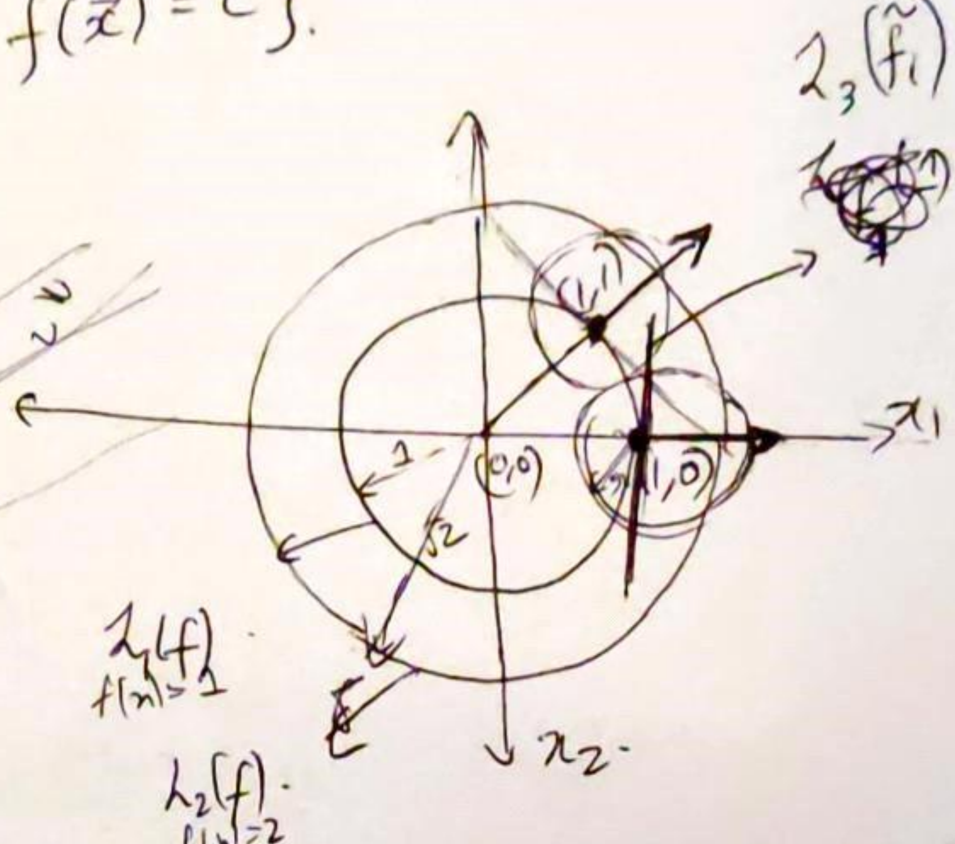
$f(\vec{x}) = x_1^2 + x_2^2$



Level set: $L_c(f) = \{ \vec{x} \mid f(\vec{x}) = c \}$.

$L_1(f) = \{ \vec{x} \mid x_1^2 + x_2^2 = 1 \}$.





$$\nabla f(\vec{x}) = \begin{bmatrix} 2x_1 \\ 2x_2 \end{bmatrix} \qquad \frac{\partial}{\partial x_1}\left(x_1^2 + x_2^2\right) = 2x_1$$

Taylor's theorem (first order).

$$\tilde{f}_1(x_1 + \Delta x_1, x_2 + \Delta x_2) = f(x_1, x_2) + \nabla f(x_1, x_2)^T \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix}$$

$$= x_1^2 + x_2^2 + [2x_1 \; 2x_2] \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix}$$

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

$x_1 = 1, \; x_2 = 0$

$$\tilde{f}_1(1 + \Delta x_1, \Delta x_2) = 1^2 + 0^2 + 2(1)\Delta x_1 + 2(0)\Delta x_2$$

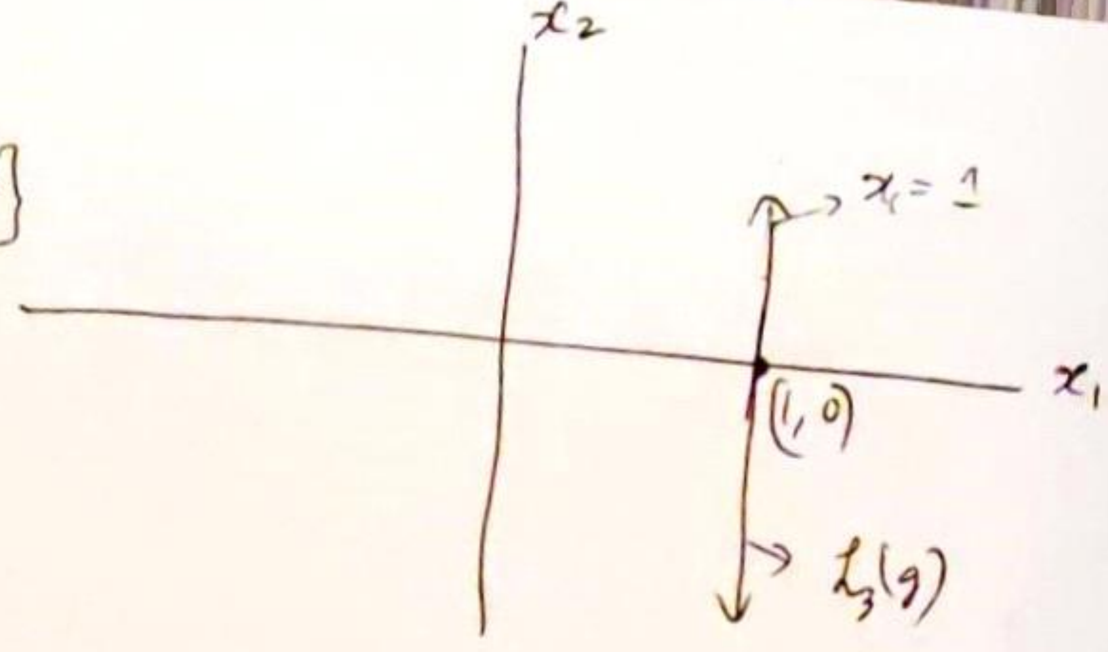
$$= 1 + 2\Delta x_1$$

Level sets of first order approximation are perpendicular to the gradient.

$\cancel{g(x_1) = 1 + 2}$

$$g(x_1, x_2) = 1 + 2x_1$$

$$L_3(g) = \{(x_1, x_2) \mid \underbrace{1 + 2x_1 = 3}_{\Rightarrow x_1 = 1}\}$$



(7) Hessian of $f(\vec{x}) = x_1^2 + x_2^2$

$$\frac{\partial^2}{\partial x_1^2}(x_1^2 + x_2^2) = \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_1}(x_1^2 + x_2^2)$$

$$= \frac{\partial}{\partial x_1}(2x_1)$$

$$= 2$$

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

$$\tilde{f}_2(x_1 + \Delta x_1, x_2 + \Delta x_2) = x_1^2 + x_2^2 + 2x_1 \Delta x_1 + 2x_2 \Delta x_2 + \frac{1}{2}[\Delta x_1 \ \Delta x_2]\begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix}\begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix}$$

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

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

$$= f(x_1 + \Delta x_1, x_2 + \Delta x_2).$$

8. More examples

① $f(\vec{x}) = \vec{x}^T \vec{a}$, $\vec{x} \in \mathbb{R}^n$ $\vec{a} \in \mathbb{R}^n$.

$= \sum_i x_i a_i$

$$\left(\nabla f(\vec{x})\right)_k = \frac{\partial}{\partial x_k}\left(\sum_i x_i a_i\right) = a_k$$

$$\boxed{\nabla f(\vec{x}) = \vec{a}}$$

② $f(\vec{x}) = \vec{x}^T A \vec{x}$ $\vec{x} \in \mathbb{R}^n$ $A \in \mathbb{R}^{n\times n}$.

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

$$= \sum_i x_i^2 A_{ii} + \sum_i \sum_{j \neq i} x_i A_{ij} x_j$$

$=$ terms that don't contain x_k

$$+ x_k^2 A_{kk} + \sum_{j \neq k} x_k A_{kj} x_j + \sum_{i \neq k} x_i A_{ik} x_k.$$

$\frac{\partial f(\vec{x})}{\partial x_k}$.

$=$ other terms

$$+ x_k^2 A_{kk} + x_k \sum_{j\neq k} \left(A_{kj} x_j + x_j A_{jk}\right))$$

matrix A unit vector $\vec{e}_j = \begin{bmatrix} 0 \\ \vdots \\ 1 \\ 0 \end{bmatrix} \leftarrow$ j^{th} entry.

$A\vec{e}_j = j^{th}$ column of A

$\vec{e}_i^T A = i^{th}$ row of A

$$\vec{x} = \sum_i x_i \vec{e}_i$$

$$\vec{x}^T A \vec{x} = \left(\sum_i x_i \vec{e}_i\right)^T A \left(\sum_j x_j \vec{e}_j\right)$$

$$\vec{e}_i^T A \vec{e}_j = A_{ij}$$

$$f(\vec{x}) = x_k^2 A_{kk} + x_k \left(\sum_{j \neq k} (A_{kj} + A_{jk}) x_j \right) + \text{other terms}$$

$$(\nabla f(\vec{x}))_k = 2 A_{kk} x_k + \sum_{j \neq k} (A_{kj} + A_{jk}) x_j$$

$$= \sum_j (A_{kj} + A_{jk}) x_j \Rightarrow \boxed{\nabla f(\vec{x}) = (A + A^T) \vec{x}}$$

$$= \left((A + A^T) \vec{x} \right)_k$$

Vector least squares.

$$x^* = \underset{\vec{x}}{\operatorname{argmin}} \| A \vec{x} - \vec{b} \|_2^2$$

$$f(\vec{x}) = \| A \vec{x} - \vec{b} \|_2^2 = (A \vec{x} - \vec{b})^T (A \vec{x} - \vec{b})$$

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

$$\nabla f(\vec{x}) = (A^T A + (A^T A)^T) \vec{x} + 0 - 2 A^T \vec{b}$$

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

(10) For convex functions find minimizer $\vec{x}^*$, by finding $\vec{x}$ such that $\nabla f(\vec{x}) = 0$.

$$\nabla f(\vec{x}^*) = 0 \Rightarrow 2(A^T A \vec{x}^* - A^T \vec{b}) = 0$$

$$\Rightarrow A^T A \vec{x}^* = A^T \vec{b}.$$

$$\vec{x}^* = (A^T A)^{-1} A^T \vec{b}$$