Mathematics for AI

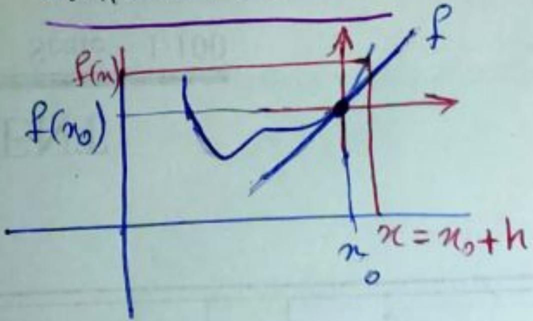
Lecture 15

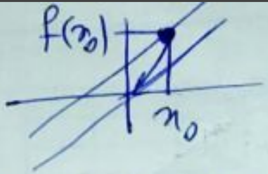
Remember: derivative and linearization

K. N. Toosi
University of Technology

MA 15 I

linearization





$$f(x) - f(x_0) = A(x - x_0)$$

$$\frac{f(x_0+h) - f(x_0)}{h}$$

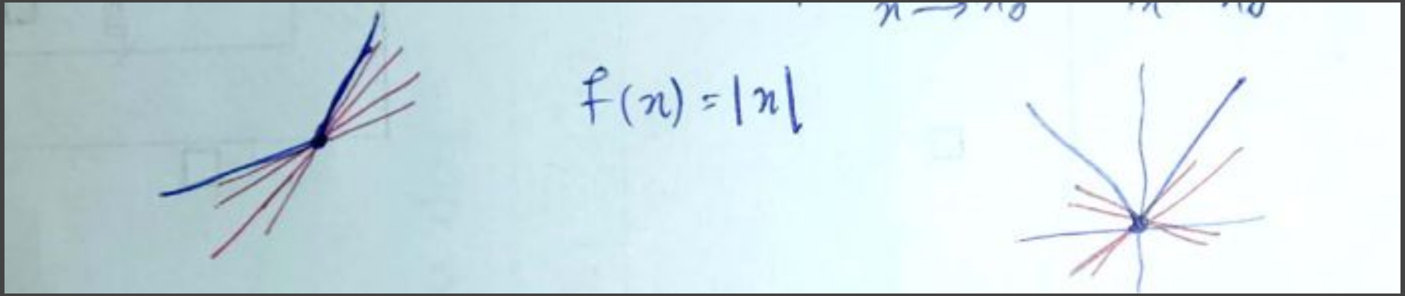
$\mathbb{R} \to \mathbb{R}$

$f(x) = m x$

$$\lim_{x \to x_0} \frac{[f(x) - f(x_0)] - m(x - x_0)}{x - x_0} = 0$$

$$m = f'(x_0) = \lim_{x \to x_0} \frac{f(x) - f(x_0)}{x - x_0}$$

Non-differentiable functions / subgradients



Differentiability classes

differentiable $f: \mathbb{R} \to \mathbb{R}$

$f(x)$ $f'(x_0)$ exist and is contiuous at x_0

f is continuously differentiable at x_0

$f(x)$ is contionuously differentiable everwhere

$f \in C^1$

f: contionues $f \in C^0$

f: differentable f' exists & is continuous $f \in C^1$

f'' exists and is continuous $f \in C^2$

$f^{(n)}$ " " " " $f \in C^n$

$C^0 \supset C^1 \supset C^2 \supset C^3 \ldots \supset C^\infty$

C^∞ smooth $\longrightarrow$ differentiability class

https://en.wikipedia.org/wiki/Smoothness#Differentiability_classes

Vector-valued functions (univariate)

K. N. Toosi
University of Technology

MA1

$f : \mathbb{R} \to \mathbb{R}$

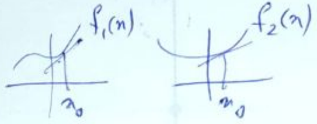
$f : \mathbb{R} \to \mathbb{R}^n$

$$f(x) = \begin{bmatrix} \sin x \\ \cos x \end{bmatrix} \qquad f : \mathbb{R} \to \mathbb{R}^2$$

$\boxed{f(x) \in \mathbb{R}^n} \to$ vector-valued function

$$f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix} \qquad f_i : \mathbb{R} \to \mathbb{R}$$

$$f'(x) = \begin{bmatrix} f_1'(x) \\ f_2'(x) \\ \vdots \\ f_n'(x) \end{bmatrix}$$



$f(x)$ around x_0

$$f(x) \simeq f(x_0) + \underset{n\times 1}{\vec{m}}\,(x - x_0)$$

$$\left.\frac{df}{dx}\right|_{x_0} = \vec{m} = \begin{bmatrix} f_1'(x_0) \\ f_2'(x_0) \\ \vdots \\ f_n'(x_0) \end{bmatrix} \in \mathbb{R}^{n\times 1}$$

Derivative of dot product

$$\frac{d}{dx}\left(f(x)\,g(x)\right) = \left(\frac{d}{dx} f\right) g + f \frac{d}{dx} g$$

$$= f'(x)\,g(x) + f(x)\,g'(x)$$

$f, g : \mathbb{R} \to \mathbb{R}^n$

$$h(x) = f(x)^T g(x)$$

$$h : \mathbb{R} \to \mathbb{R}$$

$$h'(x) = f'(x)^T g(x) + f(x)^T g'(x)$$

Derivative of scalar product

f, g $\quad f: \mathbb{R}^n \to \mathbb{R}$

$\quad\quad g: \mathbb{R} \to \mathbb{R}$

$$h(x) = f(x)\, g(x)$$

$$h'(x) = f'(x)\, g(x) + f(x)\, g'(x)$$

Matrix-valued functions (univariate)

K. N. Toosi
University of Technology

matrix valued functions $f : \mathbb{R} \to \mathbb{R}^{m \times n}$

$$F(\theta) = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \in \mathbb{R}^{2 \times 2}$$

$$F(x) = G(x)\,H(x)$$

derivative of matrix multiplication

$F : \mathbb{R} \to \mathbb{R}^{m \times n}$

$G : \mathbb{R} \to \mathbb{R}^{m \times p}$

$H : \mathbb{R} \to \mathbb{R}^{p \times n}$

MA 15

$$F(x) = G(x)\, H(x)$$

$$F'(x) = G'(x)\, H(x) + G(x)\, H'(x)$$

Functions of multiple variables

- Multivariable functions / Multivariate functions

$f: \mathbb{R}^n \longrightarrow \mathbb{R} \longrightarrow$ (real valued)

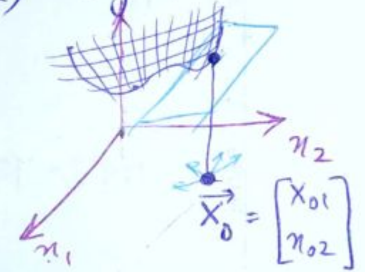
$f(x,y) = \log x \sin y + y^2 + xy$

$\mathbb{R}\times\mathbb{R} \to \mathbb{R} \qquad \mathbb{R}^2 \longrightarrow \mathbb{R}$

$f\left(\begin{bmatrix} x \\ y \end{bmatrix}\right) = \log x \sin y + y^2 + xy$

$f(X) = f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = f(x_1, x_2) = \log x_1 \sin x_2 + x_2^2 + x_1 x_2 \in \mathbb{R}$

$X \in \mathbb{R}^2$

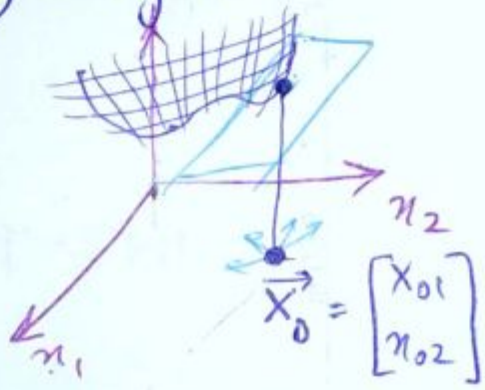


$f: \mathbb{R}^n \to \mathbb{R}$

Linearization of multivariate functions

$$f(\vec{x}) = f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = f(x_1, x_2)$$

$\vec{x} \in \mathbb{R}^2$



$$f: \mathbb{R}^n \to \mathbb{R}$$

$$\underbrace{f(\vec{x}) - f(\vec{x}_0)}_{1\times 1} = \underbrace{m^T}_{1\times n} \underbrace{(x - x_0)}_{n\times 1}$$

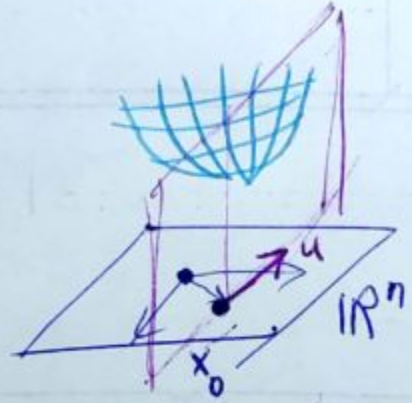
Directional Derivative

directional derivative

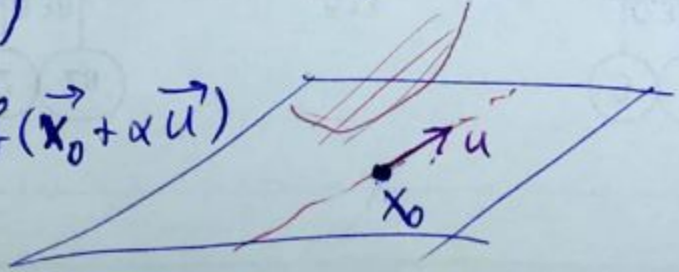
$f: \mathbb{R}^n \longrightarrow \mathbb{R}$

$h(\alpha) \qquad h: \mathbb{R} \longrightarrow \mathbb{R}$

$h(\alpha) = f(\vec{x}_0 + \alpha \vec{u})$



$$D[u]f\Big|_{x=x_0} = \frac{d}{d\alpha} h(\alpha)\Big|_{\alpha=0} = \frac{d}{d\alpha} f(\vec{x}_0 + \alpha \vec{u})$$



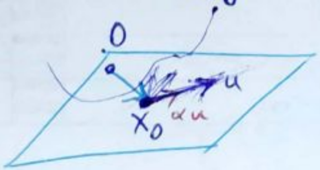
Directional Derivative

K. N. Toosi
University of Technology

MA 15 (IV)

$$\partial_u f(x_0) = D_u f(x_0) = D[u] f(\vec{x}_0) = D[u] f \Big|_{\vec{x}_0}$$

$$D[u] f(x_0) = \frac{d}{d\alpha} f(x_0 + \alpha u)$$



$$= \cancel{f} \quad \cancel{f(x_0+\alpha}$$

$$= \lim_{\alpha \to 0} \frac{f(\vec{x}_0 + \alpha \vec{u}) - f(\vec{x}_0)}{\alpha} \qquad \|u\| = 1$$

let u be of any length (not just 1).

$$D[2u] f(x_0) = f(x_0 + \alpha(2u)) = \lim_{\alpha \to 0} \frac{f(\vec{x}_0 + 2\alpha \vec{u}) - f(x_0)}{\alpha}$$

$$= \lim_{\beta \to 0} \frac{f(x_0 + \beta u) - f(x_0)}{\beta/2}$$

$$= 2 \, \cancel{f} \, \frac{d}{d\alpha} f(x_0 + \beta u)$$

Linearity of directional derivative

$$D[2u]f(x_0) = f(x_0+\alpha(2u)) = \lim_{\alpha\to 0}\frac{f(\vec{x}_0+2\alpha\vec{u}) - f(x_0)}{\alpha}$$

$$= \lim_{\beta\to 0}\frac{f(x_0+\beta u)-f(x_0)}{\beta/2}$$

$$= 2\,\frac{d}{d\beta} f(x_0+\beta u)$$

$$\Rightarrow D[\gamma\vec{u}]f = \gamma D[\vec{u}]f$$

if $f:\mathbb{R}^n\to\mathbb{R}$ is differentiable

$$D[\vec{u}+\vec{v}]f = D[\vec{u}]f + D[\vec{v}]f$$

$\Rightarrow D[u]f$ is linear in u.

$d(u) = D[u]f\big|_{x_0} \Rightarrow d:\mathbb{R}^n\to\mathbb{R}$ is linear

$$\Rightarrow \exists\, m\in\mathbb{R}^m \quad d(u) = \underbrace{m^T}_{1\times m} u$$

Linearity of directional derivative

$\Rightarrow D[u]f$ is linear in u.

$d(u) = D[u]f\Big|_{x_0} \Rightarrow d : \mathbb{R}^n \longrightarrow \mathbb{R}$ is linear

$d \Rightarrow \exists\ m \in \mathbb{R}^m \quad d(u) = \underbrace{m^T}_{1\times m} u$

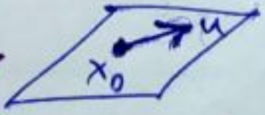
The gradient vector

$d(u) = D[u]f\big|_{x_0} \Rightarrow d: \mathbb{R}^n \longrightarrow \mathbb{R}$ is linear

$d \Rightarrow \exists\ m \in \mathbb{R}^m \quad d(u) = \underbrace{m^T}_{1\times m} u$

$$D[u]f\big|_{x=x_0} = \vec{m}^T u = \vec{m}(x_0)^T u = \langle \vec{m}, \vec{u} \rangle$$

$\vec{m}$ is called the gradient of f at $\vec{x}_0$. $\vec{m} \in \mathbb{R}^n$ and is denoted by ∇_{x_0}



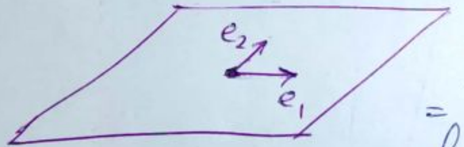
Gradient and partial derivatives

$f:\mathbb{R}^n \to \mathbb{R}$ differentiable at x_0 $\exists \nabla \in \mathbb{R}^n$ MA15 ∇

$$D[u]f(x_0) = \nabla^T u = \langle \nabla, u \rangle \qquad \nabla_{x_0} \in \mathbb{R}^n$$

$$\nabla = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix} \qquad m_1 = \langle \nabla, e_1 \rangle = \left\langle \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}, \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \right\rangle = D[e_1]f$$

$$= \frac{\partial f}{\partial x_1}\Big|_{\vec{X}_0}$$



$$= \lim_{\alpha \to 0} \frac{f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + \alpha \begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right)}{\alpha}$$

$$= \lim \; f\left(\begin{bmatrix} x_1 + \alpha \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right)$$

$$\nabla f\Big|_{x_0} = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \frac{\partial f}{\partial x_2}(x_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_0) \end{bmatrix}$$

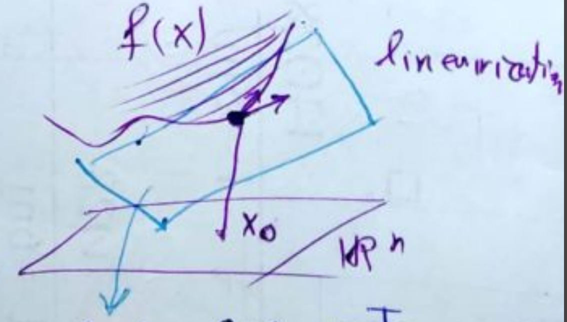
Linearization of multivariate functions

K. N. Toosi
University of Technology

$$\lim_{\alpha \to 0}$$

$$= \lim_{\alpha \to 0} \frac{f\left(\begin{bmatrix} x_1+\alpha \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) - f\left(\begin{bmatrix} x_1 \\ x_2 \\ | \\ x_n \end{bmatrix}\right)}{\alpha}$$

$$\nabla f \Big|_{x_0} = \begin{bmatrix} \frac{\partial f}{\partial x_1}(x_0) \\ \frac{\partial f}{\partial x_2}(x_0) \\ \vdots \\ \frac{\partial f}{\partial x_n}(x_0) \end{bmatrix}$$



$$f(x) - f(x_0) \simeq \nabla^T (x - x_0)$$

$$l(x) = f(x_0) + \nabla^T (x - x_0)$$

$$l(x) = \nabla^T x + (f(x_0) - \nabla^T x_0)$$