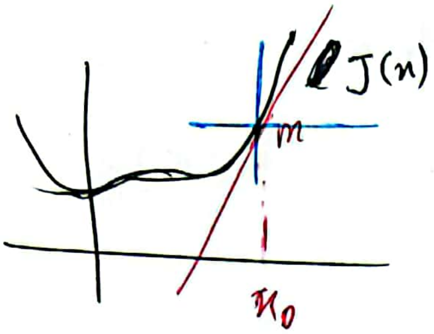
linearization & differentiability



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

$$J(x) - J(x_0) = m(x-x_0)$$

$$f(x) \simeq f(x_0) + m(x-x_0)$$

$\exists$ a linear map $l:\mathbb{R}\to\mathbb{R}$

$$\lim_{x\to x_0} \frac{|f(x)-f(x_0)-l(x-x_0)|}{|x-x_0|} = 0$$

l linear $\Longrightarrow l(x-x_0) = \underset{1\times 1}{m}(x-x_0)$

$f:\mathbb{R}\to\mathbb{R}^n$ $\quad f(x) = \begin{bmatrix} f_1(x) \\ f_2(x) \\ \vdots \\ f_n(x) \end{bmatrix}$ $\quad f' = \begin{bmatrix} f_1'(x) \\ f_2'(x) \\ \vdots \\ f_n'(x) \end{bmatrix} \in \mathbb{R}^n$

$$\lim_{x\to x_0} \frac{|f(x)-f(x_0)-l(x-x_0)|}{|x-x_0|} = 0$$

$$\lim_{x\to x_0} \left| \frac{f(x)-f(x_0)}{x-x_0} - \frac{l(x-x_0)}{x-x_0} \right|$$

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

$l(x) = \vec{m}\,x$

$\vec{m}$ $\quad n\times 1$

Multivariable functions (real valued)

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

$$f(x_1, x_2, \dots, x_n) = f\left(\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right) = f(\vec{x}) \qquad \vec{x} \in \mathbb{R}^n$$

$$f(x,y) = \log(1+|x|)\, e^y$$

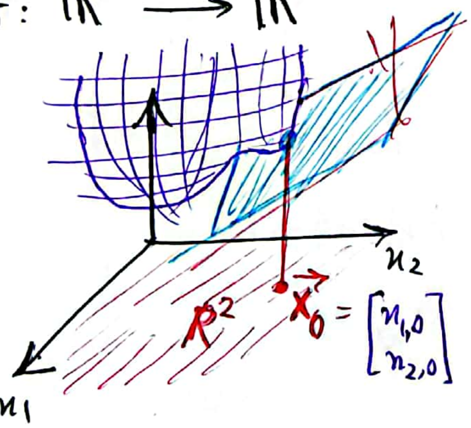
$$f: \mathbb{R} \times \mathbb{R} \longrightarrow \mathbb{R}$$

$$f(\vec{x}) = f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = \log(1+|x_1|)\, e^{x_2}$$

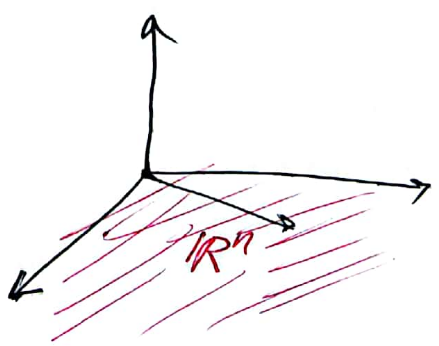
$$f: \mathbb{R}^2 \longrightarrow \mathbb{R}$$

$$f(\vec{x}) = \frac{1}{1+e^{-a^T\vec{x}}}$$

$f: \mathbb{R}^2 \longrightarrow \mathbb{R}$



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

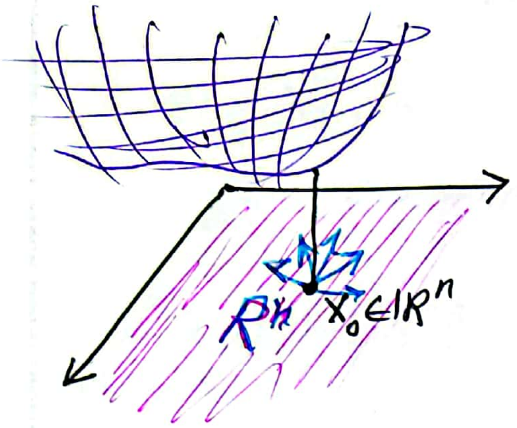


$$\underbrace{J(\vec{x}) - J(\vec{x}_0)}_{\mathbb{R}} = \underbrace{\vec{m}^T}_{1\times 2} \underbrace{(\vec{x} - \vec{x}_0)}_{\mathbb{R}^2}$$

$$f(\vec{x}) - f(\vec{x}_0) \simeq \vec{m}^T(\vec{x} - \vec{x}_0) \quad \text{around } \vec{x}_0$$

$\exists\, \ell: \mathbb{R}^n \longrightarrow \mathbb{R}$ مشتق پذیری

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



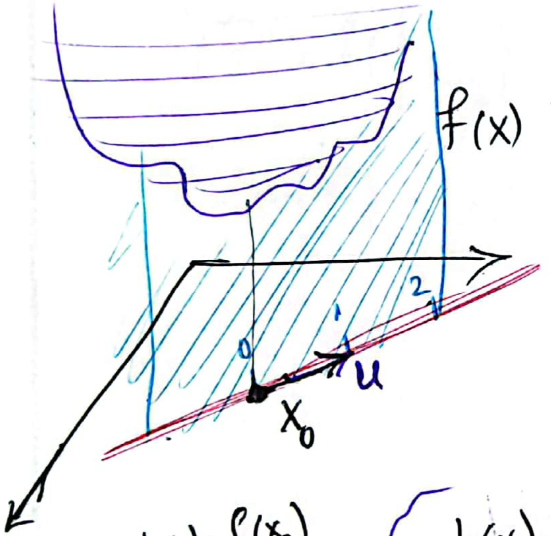
directional derivative

of $f: \mathbb{R}^n \to \mathbb{R}$ at $\vec{x}_0 \in \mathbb{R}^n$

along the direction $\vec{u} \in \mathbb{R}^n$

$$\left.\frac{\partial f}{\partial \vec{u}}\right|_{x_0} = \partial_u f = D_u f\Big|_{x_0} = (D_u f)(x_0)$$

$$= D[u]f\Big|_{x_0} = D[u]f(x_0) = (D[u]f)(x_0)$$



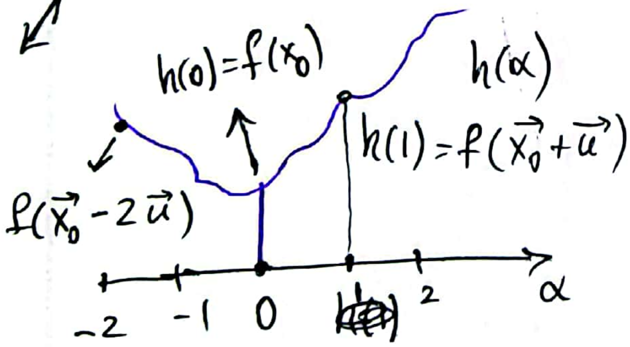
$\|\vec{u}\| = 1$

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

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

$$D[u]f(\vec{x}_0) = \left.\frac{d}{d\alpha} h\right|_{\alpha=0} = h'(0)$$

$$= \underbrace{\left.\frac{d}{d\alpha} f(\vec{x}_0 + \alpha \vec{u})\right|_{\alpha=0}}_{\text{easy to calculate}}$$



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

Do away with $\|\vec{u}\| = 1$. $\vec{u}$ can be anything in $\mathbb{R}^n$!

$$D[\vec{u}]f(\vec{x}_0) = \left.\frac{d}{d\alpha} f(\vec{x}_0 + \alpha \vec{u})\right|_{\alpha=0}$$

$$D[2\vec{u}]f(\vec{x}_0) = \left.\frac{d}{d\alpha} f(\vec{x}_0 + \alpha 2\vec{u})\right|_{\alpha=0} = 2\left.\frac{d}{d\alpha} f(\vec{x}_0 + \alpha \vec{u})\right|$$

$$= \lim_{\alpha \to 0} \frac{f(\vec{x}_0 + 2\alpha \vec{u}) - f(x_0)}{\alpha} = \lim_{2\alpha \to 0} 2\, \frac{f(\vec{x}_0 + 2\alpha u) - f(x_0)}{\underbrace{2\alpha}_{\beta}} \quad \alpha = 0$$

$D[\cdot]f(x_0) : \mathbb{R}^n \to \mathbb{R}$ (as a function of $\vec{u}$) MA17 II

Fixing x_0, is $D[u]f(x_0)$ linear as a function of $\vec{u} \in \mathbb{R}^n$?

$$D[\beta u]f(x_0) = \beta D[u]f(x_0) \checkmark$$

$$D[u+v]f(x_0) \overset{?}{=} D[u]f(x_0) + D[v]f(x_0)$$

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

$$f(\vec{x}) = f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix}\right) = x_1^2 \sin x_2$$

$$f(\vec{x}+\alpha\vec{u}) = f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \alpha \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}\right) = (x_1 + \alpha u_1)^2 \sin(x_2 + \alpha u_2)$$

$$\frac{d}{d\alpha} f(\vec{x}+\alpha\vec{u})\Big|_{\alpha=0} = 2(x_1+\alpha u_1)u_1 \sin(x_2+\alpha u_2) + (x_1+\alpha u_1)^2 \cos(x_2+\alpha u_2)\, u_2 \Big|_{\alpha=0}$$

$$= 2x_1 u_1 \sin(x_2) + u_2 x_1^2 \cos(x_2)$$

$$= \underbrace{\left[2x_1 \sin(x_2),\ x_1^2 \cos(x_2)\right]}_{2\times 1} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$$

$$\frac{d}{d\alpha} f(\vec{x} + \alpha(\vec{u}+\vec{v})) = \frac{d}{d\alpha} f\left(\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} + \alpha \begin{bmatrix} u_1+v_1 \\ u_2+v_2 \end{bmatrix}\right) \quad 2\times 1$$

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

For a <u>differentiable</u> function at $\vec{X}_0$

$D[u]f(x_0)$ is linear as a function of $\vec{u}$. $\vec{u} \in \mathbb{R}^n$

$\exists M \in$

$\Longrightarrow D[u]f(x_0)$

Scanned with CamScanner

$D[u]f(x_0)$ is linear as a function of $\vec{u} \in \mathbb{R}^n$

$\exists M \in \mathbb{R}^{1\times n}$ $\quad D[u]f(x_0) = \underset{1\times n}{M}\ \underset{n\times 1}{\vec{u}} = M(x_0)\,\vec{u}$

$[m_1\ m_2\ -\ m_n]$ $\quad$ M is a function of $\vec{x}_0$

$\exists \vec{m} \in \mathbb{R}^n$ $\quad D[u]f(x_0) = \vec{m}^T\vec{u}$

$= \langle \vec{m}, \vec{u} \rangle$

$\vec{m} = \vec{m}(x_0)$ is call the gradient of f at $\vec{x}_0$.

گرادیان

How to ~~compute~~ calculate the gradient vector at $\vec{x}_0$?

The hard way (usually)

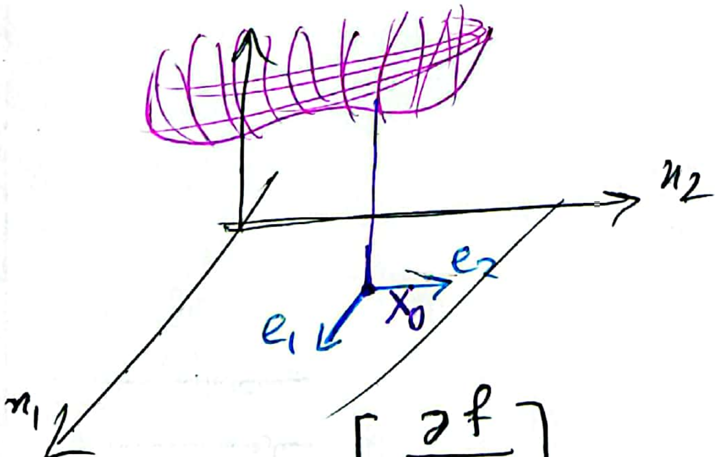
$\vec{m} \in \mathbb{R}^n \Rightarrow \vec{m} = \begin{bmatrix} m_1 \\ m_2 \\ \vdots \\ m_n \end{bmatrix}$ $\quad$ what is m_1?

$m_1 = [m_1\ m_2\ \cdots\ m_n]\underbrace{\begin{bmatrix} 1 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}}_{\vec{e}_1} = m^T\vec{e}_1 = D[\vec{e}_1]f(x_0)$

$= \left.\dfrac{\partial f}{\partial x_1}\right|_{x_0}$

$= \dfrac{d}{d\alpha} f\left(\vec{x} + \alpha\begin{bmatrix} 1 \\ 0 \\ \vdots \end{bmatrix}\right)$

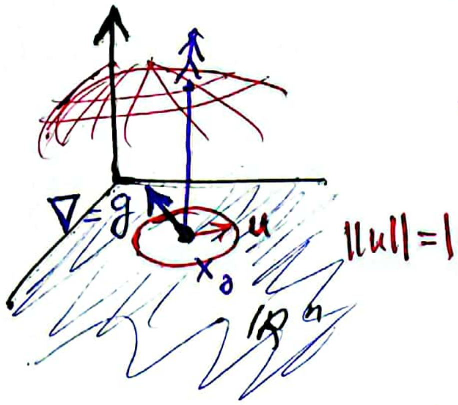
$= \dfrac{d}{d\alpha} f\left(\begin{bmatrix} x_1 + \alpha \\ x_2 \\ \vdots \\ x_n \end{bmatrix}\right)$

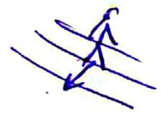


$\vec{m} = \left.\begin{bmatrix} \dfrac{\partial f}{\partial x_1} \\ \dfrac{\partial f}{\partial x_2} \\ \vdots \\ \dfrac{\partial f}{\partial x_n} \end{bmatrix}\right|_{x = x_0}$

gradient is usually denoted by $\vec{\nabla} f|_{x_0}$

$\nabla f(x_0)$ org





$$\nabla = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_n \end{bmatrix} \in \mathbb{R}^n$$

$$\max_{\|u\|=1} D[u]f(x_0) = \max_{\|u\|=1} \langle \vec{u}, \vec{\nabla} \rangle \qquad \vec{u} = \frac{\vec{\nabla}}{\|\nabla\|}$$

$$\min_{\|u\|=1} D[u]f(x_0) = \min_{\|u\|=1} \langle \vec{u}, \vec{\nabla} \rangle \qquad \vec{u} = \frac{-\nabla}{\|\nabla\|}$$

$$\vec{u} = \frac{\vec{\nabla}}{\|\nabla\|} \qquad \langle \vec{u}, \vec{\nabla} \rangle = \left\langle \frac{\nabla}{\|\nabla\|}, \nabla \right\rangle$$

$$= \frac{1}{\|\nabla\|} \langle \vec{\nabla}, \vec{\nabla} \rangle = \frac{\|\nabla\|^2}{\|\nabla\|}$$

$$= \|\nabla\|$$

$$u \perp \nabla \Rightarrow \langle \vec{u}, \vec{\nabla} \rangle = 0$$