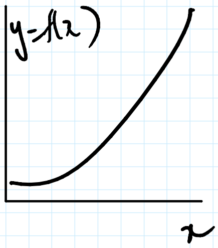
Gradient

12 March 2021 09:05

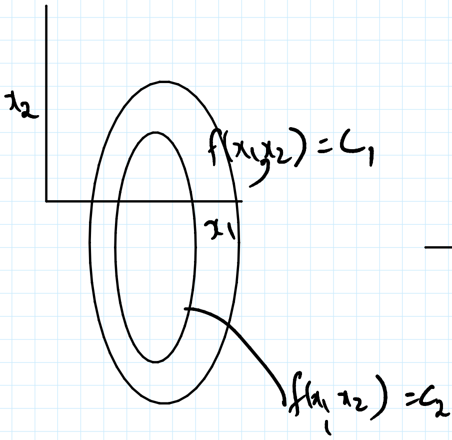
· $f : \underset{\sim}{\mathbb{R}^n} \to \mathbb{R}$

$f(\underbrace{x_1, x_2 \cdots x_n}_{x}) \in \mathbb{R}$

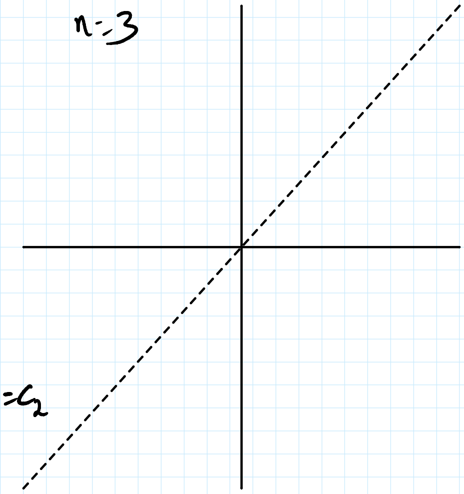
$n=1$



$n=2$



$n=3$



f: continuous, differentiable
no gaps)

$x \to x+h$

$$\lim_{|h|\to 0} \frac{f(x+h)-f(x)}{|h|} \quad \text{exists}$$

$n=1$ $$f(x+h) - f(x) = h\, f'(x) + \underbrace{\frac{h^2}{2} f''(x)}_{\text{neglect these term}} + \cdots$$ (Taylor series)

h is very small

$$f(x+h) \approx f(x) + h\, f'(x)$$

$n \geq 1$ $$\underbrace{f(x+h)}_{\in \mathbb{R}} \approx \underbrace{f(x)}_{\in \mathbb{R}} + \underbrace{\underbrace{\nabla f^T(x)}_{1\times n} \; \underbrace{h}_{n\times 1}}_{\mathbb{R}}.$$

∇f: $n \times 1$

h: $n \times 1$

$\nabla f^T h \in \mathbb{R}$

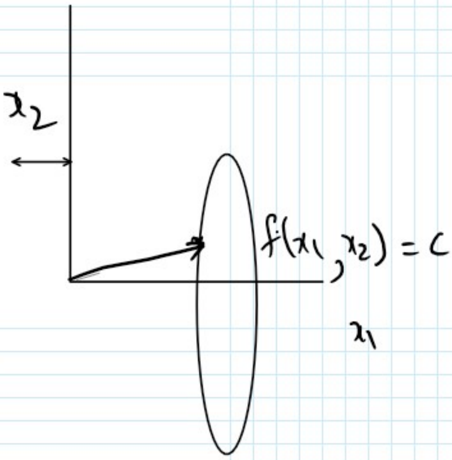
$$\underbrace{\nabla f}_{\text{gradient}} = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix}$$

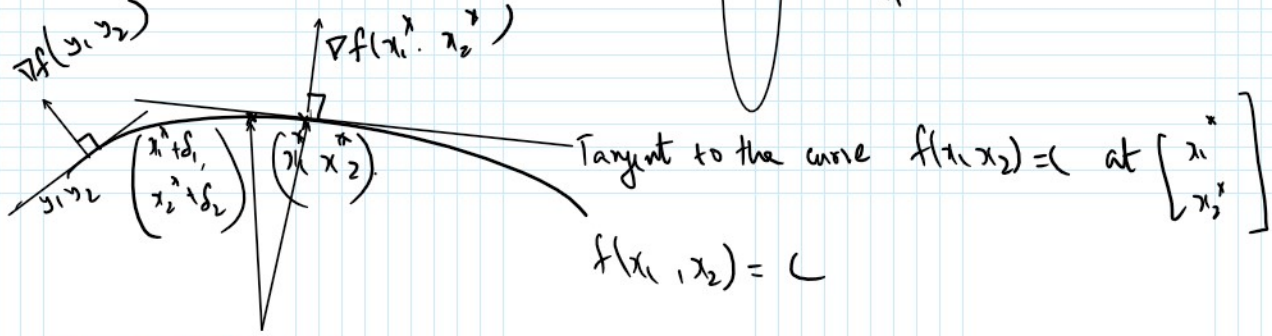
$n=1$ $\nabla f = \frac{df}{}$

$\frac{\partial f}{\partial x_n}$

$n=1 \quad \nabla f = \frac{df}{dx}$

$n=2 \quad \nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix}$ $\quad x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$





∇f is $\perp$ar to the surface

$$\underbrace{f(x_1^* + \delta_1, x_2^* + \delta_2)}_{C} = \underbrace{f(x_1^*, x_2^*)}_{C} + \nabla f^T \begin{bmatrix} \cancel{x_1^*} + \delta_1 - \cancel{x_1^*} \\ \cancel{x_2^*} + \delta_2 - \cancel{x_2^*} \end{bmatrix} + \ldots \text{ higher order terms}$$

when $\delta_1 \to 0$
$\delta_2 \to 0$

$$\cancel{C} = \cancel{C} + \nabla f^T \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} + \underbrace{\text{higher order terms}}_{\to 0}$$

$$\nabla f^T \begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix} = 0$$

∇f $\perp$ar to $\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$

as $\delta_1 \to 0$ $\delta_2 \to 0$ $\begin{bmatrix} \delta_1 \\ \delta_2 \end{bmatrix}$ lies on the tangent

$\therefore$ ∇f $\perp$ar to the tangent

∇f is normal to the curve $f(x_1, x_2) = C$

∇f is (locally) the direction of max increase (steepest ascent)

$-\nabla f$: steepest descent (locally : most decrease)

Formulae for gradient :

$f(x_1 x_2 .. x_n) = c$ (c. constant)

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$f(x_1 x_2 .. x_n) = c_1 x_1 + c_2 x_2 + .. + c_n x_n + c_0$$

$$= \underbrace{[c_1 \ c_2 \ .. \ c_n]}_{c^T} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} + c_0$$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \vdots \\ \frac{\partial f}{\partial x_n} \end{bmatrix} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \underset{\sim}{c}$$

$$f(x) = \underset{1 \times n}{\underset{\sim}{c}^T} x + c_0$$

$\nabla f = \underset{n \times 1}{\underset{\sim}{c}}$ (constant everywhere)

c is a normal vector to the plane $c^T x + c_0 = \alpha$

c is my normal vector to the plane $c^T x + c_0 = \alpha$

$f(x) = \alpha$ is a surface of constant value (α)

$$f(x) = c_0 + c^T x + \frac{1}{2} \underset{1\times n}{x^T} \underset{n\times n}{H} \underset{n\times 1}{x}$$

H is a real symmetric matrix $(n\times n)$

$n = 2$

$$f(x) = c_0 + c_1 x_1 + c_2 x_2 + \frac{1}{2} \begin{bmatrix} x_1 & x_2 \end{bmatrix} \begin{bmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$= c_0 + c_1 x_1 + c_2 x_2 + \frac{1}{2} h_{11} x_1^2 + \frac{1}{2} h_{22} x_2^2 + h_{12} x_1 x_2$$

$$\nabla f = \begin{bmatrix} \frac{\partial f}{\partial x_1} \\ \frac{\partial f}{\partial x_2} \end{bmatrix} = \begin{bmatrix} c_1 + \frac{1}{2} \cdot 2 h_{11} x_1 + h_{12} x_2 \\ c_2 + \frac{1}{2} 2 \cdot h_{22} x_2 + h_{12} x_1 \end{bmatrix}$$

$$= \underset{\sim}{\begin{bmatrix} c_1 \\ c_2 \end{bmatrix}} + \begin{bmatrix} h_{11} x_1 + h_{12} x_2 \\ h_{12} x_1 + h_{22} x_2 \end{bmatrix}$$

$$= c + \underbrace{\begin{bmatrix} h_{11} & h_{12} \\ h_{12} & h_{22} \end{bmatrix}}_{H} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

$$\nabla f = \underset{2\times1}{c} + \underset{2\times2}{H}\underset{2\times1}{x}.$$

$(n=1)$ $f(x) = c_0 + c_1 x + \frac{1}{2} h x^2$

$$\frac{df}{dx} = c_1 + hx$$

Verify that in general

when $f = c_0 + c^T x + \frac{1}{2} x^T H x$ (H: symmetric)

$$\boxed{\nabla f = 0 + c + Hx = Hx + c}$$

If the matrix is not symmetric

$$f = c_0 + c^T x + \frac{1}{2} x^T A x \quad (A \text{ is not symmetric})$$

$$= c_0 + c^T x + \frac{1}{2} x^T \left(\frac{A + A^T}{2}\right) x$$

$\frac{A + A^T}{2}$: symmetric $= H$

$$f = c_0 + c^T x + \frac{1}{2} x^T H x$$

symmetry has been restored!

$$x^T A x = x^T \left(\frac{A + A^T}{2}\right) x$$

$$= \frac{x^T A x}{2} + \frac{x^T A^T x}{2}$$

$$= \frac{1}{2}\left[x^T A x + x^T A^T x\right] = \frac{1}{2}\left[x^T A x + x^T A x\right]$$

$$(x^T A x)^T = x^T A^T x \qquad = x^T A x$$