Chain Rule

Recall: 1-variable

$$\begin{cases} w = g(u) = 2u+1 \\ u = f(x) = x^2 \end{cases}$$

w can be regarded as a function x

$$w = g \circ f(x) = g(f(x)) = 2x^2+1$$

(Abuse of notation : $w = w(x)$ or $w = \underset{\substack{\shortparallel \\ 2x^2+1}}{g(x)}$)

Then $\frac{dw}{dx} = \frac{dw}{du}\frac{du}{dx}$ (usual way of writing)

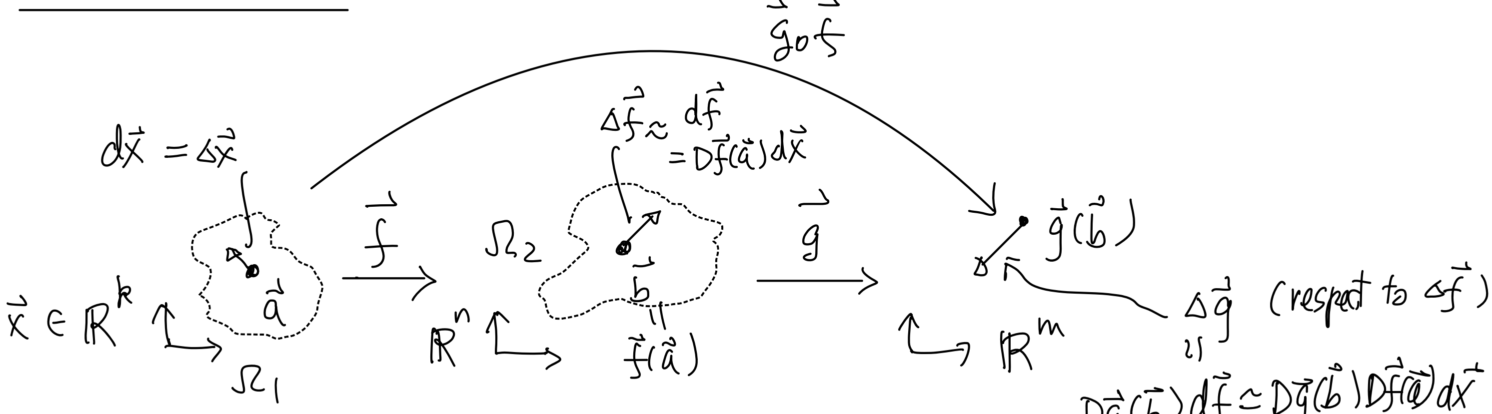
$$\left(\frac{dw}{dx}(x) = \frac{d(g \circ f)}{dx}(x) = \frac{dg}{du}(f(x)) \cdot \frac{df}{dx}(x) \right)$$

$$\frac{dw}{dx} = 2 \cdot 2x = 4x$$

Caution: Abuse of notation :

$\frac{dw}{dx}$ is $\frac{d(g \circ f)}{dx}(x)$ & $\frac{dw}{du}$ is $\frac{df}{dx}(x)$

General dimensions:



Thm (Chain Rule)

Let
- $\vec{f}: \Omega_1 \to \mathbb{R}^n$ ($\Omega_1 \subseteq \mathbb{R}^k$, open)
- $\vec{g}: \Omega_2 \to \mathbb{R}^m$ ($\Omega_2 \subseteq \mathbb{R}^n$, open)
- $\vec{f}(\Omega_1) \subset \Omega_2$,

If
- $\vec{f}$ differentiable at $\vec{a} \in \Omega_1 \subset \mathbb{R}^k$
- $\vec{g}$ differentiable at $\vec{b} = \vec{f}(\vec{a}) \in \Omega_2 \subset \mathbb{R}^n$

Then $\vec{g} \circ \vec{f}$ is differentiable at $\vec{a}$, and

$$D(\vec{g} \circ \vec{f})(\vec{a}) = D\vec{g}(\vec{f}(\vec{a}))\, D\vec{f}(\vec{a})$$

matrix multiplication

eg1: $\vec{f}: \mathbb{R} \to \mathbb{R}^2$,

$$\vec{f}(\theta) = (\cos\theta, \sin\theta) = \begin{bmatrix} \cos\theta \\ \sin\theta \end{bmatrix}$$

$$\vec{g}: \mathbb{R}^2 \to \mathbb{R}^2$$

$$\vec{g}(u,v) = \begin{bmatrix} 2uv \\ u^2 - v^2 \end{bmatrix}$$

Find $D(\vec{g} \circ \vec{f})(\theta)$. $\left(\begin{array}{c} u = \cos\theta \\ v = \sin\theta \end{array} \right)$

Soln: Method 1: Find composition explicitly

$$\vec{g} \circ \vec{f}(\theta) = \vec{g}(\cos\theta, \sin\theta) = \begin{bmatrix} 2\cos\theta\sin\theta \\ \cos^2\theta - \sin^2\theta \end{bmatrix}$$

$$= \begin{bmatrix} \sin 2\theta \\ \cos 2\theta \end{bmatrix}$$

$$D(\vec{g} \circ \vec{f})(\theta) = \begin{bmatrix} \frac{d}{d\theta} \sin 2\theta \\ \frac{d}{d\theta} \cos 2\theta \end{bmatrix} = \begin{bmatrix} 2\cos 2\theta \\ -2\sin 2\theta \end{bmatrix}.$$

Method 2 : Chain Rule

$$D\vec{f}(\theta) = \begin{bmatrix} f_1' \\ f_2' \end{bmatrix} = \begin{bmatrix} \frac{d}{d\theta}\cos\theta \\ \frac{d}{d\theta}\sin\theta \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

$$D\vec{g}(u,v) = \begin{bmatrix} -\vec{\nabla} g_1 - \\ -\vec{\nabla} g_2 - \end{bmatrix} = \begin{bmatrix} -\vec{\nabla}(2uv) - \\ -\vec{\nabla}(u^2-v^2) - \end{bmatrix} = \begin{bmatrix} 2v & 2u \\ 2u & -2v \end{bmatrix}$$

By Chain Rule :

$$D(\vec{g}\circ\vec{f})(\theta) = D\vec{g}(\vec{f}(\theta))\, D\vec{f}(\theta)$$

matrix multiplication

$$= \left. \begin{bmatrix} 2v & 2u \\ 2u & -2v \end{bmatrix} \right|_{\binom{u}{v} = \binom{\cos\theta}{\sin\theta}} \cdot \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} 2\sin\theta & 2\cos\theta \\ 2\cos\theta & -2\sin\theta \end{bmatrix} \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} -2\sin^2\theta + 2\cos^2\theta \\ -2\sin\theta\cos\theta - 2\sin\theta\cos\theta \end{bmatrix}$$

$$= \begin{bmatrix} 2\cos 2\theta \\ -2\sin 2\theta \end{bmatrix} .$$

eg2 (Abuse notations)

$$f(x,y) = (x^2, 3xy, x+y^2) \quad (=\vec{f})$$

$$g(u,v,w) = \frac{uw}{v}$$

Consider $g \circ f$:

$$\begin{matrix} x \\ y \end{matrix} \xmapsto{f} \begin{matrix} (f_1 =)\ u \\ (f_2 =)\ v \\ (f_3 =)\ w \end{matrix} \xmapsto{g} g$$

Find $\frac{\partial g}{\partial x}(1,1)$ (regard g as a function of x,y)

<u>Soln</u>: (g is real-variable, $Dg = \vec{\nabla} g$)

$$Dg = \vec{\nabla} g = \left[\frac{\partial g}{\partial u}, \frac{\partial g}{\partial v}, \frac{\partial g}{\partial w}\right] = \left[\frac{w}{v}, -\frac{uw}{v^2}, \frac{u}{v}\right]$$

$$Dg(\vec{f}(1,1)) = \vec{\nabla} g(1,3,2)$$

$$= \left[\frac{2}{3}, -\frac{2}{9}, \frac{1}{3}\right]$$

$$\left(\begin{matrix} u = x^2 = 1 \\ v = 3xy = 3 \text{ at } (1,1) \\ w = x+y^2 = 2 \end{matrix}\right)$$

$$D\vec{f} = \begin{bmatrix} \vec{\nabla} f_1 \\ \vec{\nabla} f_2 \\ \vec{\nabla} f_3 \end{bmatrix} = \begin{bmatrix} \vec{\nabla} x^2 \\ \vec{\nabla}(3xy) \\ \vec{\nabla}(x+y^2) \end{bmatrix} = \begin{bmatrix} 2x & 0 \\ 3y & 3x \\ 1 & 2y \end{bmatrix}$$

$$D\vec{f}(1,1) = \begin{bmatrix} 2 & 0 \\ 3 & 3 \\ 1 & 2 \end{bmatrix}$$

Hence, chain rule $\Rightarrow$

$$D(g\circ f)(1,1) = \left[\tfrac{2}{3}, -\tfrac{2}{9}, \tfrac{1}{3}\right] \begin{bmatrix} 2 & 0 \\ 3 & 3 \\ 1 & 2 \end{bmatrix}$$

$$= \left[\underset{\frac{\partial g}{\partial x}(1,1)}{1}, \underset{\frac{\partial g}{\partial y}(1,1)}{0}\right] \qquad \text{(check!)}$$

$$\Rightarrow \quad \frac{\partial g}{\partial x}(1,1) = 1$$

<u>Remark</u>: We should just calculate the 1^{st} column

$$\left[\tfrac{2}{3}, -\tfrac{2}{9}, \tfrac{1}{3}\right] \begin{bmatrix} 2 & * \\ 3 & * \\ 1 & * \end{bmatrix}$$

$$\left[\frac{\partial g}{\partial u}, \frac{\partial g}{\partial v}, \frac{\partial g}{\partial w}\right] \qquad \begin{bmatrix} \frac{\partial f_1}{\partial x} & * \\ \frac{\partial f_2}{\partial x} & * \\ \frac{\partial f_3}{\partial x} & * \end{bmatrix}$$

$$\frac{\partial g}{\partial x} = \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial g}{\partial w}\frac{\partial w}{\partial x} \qquad \begin{pmatrix} u = f_1 \\ v = f_2 \\ w = f_3 \end{pmatrix}$$

In general, for 2-variables $\xrightarrow{\vec{f}}$ 3-variables $\xrightarrow{g}$ real-value

ie $\quad k=2, \; n=3, \; m=1$

$$\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \xrightarrow{\vec{f}} \begin{pmatrix} f_1(x_1,x_2) \\ f_2(x_1,x_2) \\ f_3(x_1,x_2) \end{pmatrix} \xrightarrow{g} g(f_1, f_2, f_3)$$

we usually use classical notation,

(x,y) for (x_1, x_2),

$u = f_1(x,y)$, $v = f_2(x,y)$, $w = f_3(x,y)$, and

$g = g(u,v,w)$

$(x,y) \longmapsto (u,v,w) \longmapsto g$ can be regarded as function of (x,y) : $g = g(x,y)$

↑ Abuse of notations

Then the Chain rule is

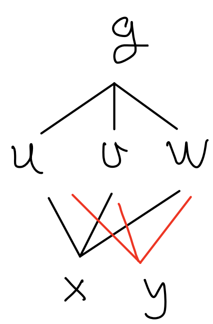
$$\boxed{\begin{aligned} \frac{\partial g}{\partial x} &= \frac{\partial g}{\partial u}\cdot\frac{\partial u}{\partial x} + \frac{\partial g}{\partial v}\cdot\frac{\partial v}{\partial x} + \frac{\partial g}{\partial w}\cdot\frac{\partial w}{\partial x} \\ \frac{\partial g}{\partial y} &= \frac{\partial g}{\partial u}\cdot\frac{\partial u}{\partial y} + \frac{\partial g}{\partial v}\cdot\frac{\partial v}{\partial y} + \frac{\partial g}{\partial w}\cdot\frac{\partial w}{\partial y} \end{aligned}}$$

$$\begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix} = D(g\circ\vec{f}) = Dg(\vec{f})\, D\vec{f} = \begin{bmatrix} \frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} & \frac{\partial g}{\partial w} \end{bmatrix} \begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y} \end{bmatrix}$$

(1×2 matrix) (1×3)(3×2) matrix

(Similarly for other low dimensional situations)

<u>Remark</u>: Branch Diagram (in Textbook)

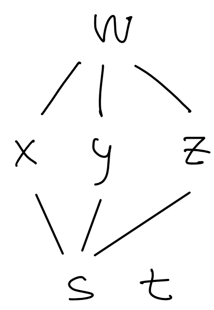


output

intermediate variables

input

eg 3 $\quad w(x,y,z) = \sqrt{x^2+y^2+z^2}$



$$\begin{cases} x = 3e^t \sin s \\ y = 3e^t \cos s \\ z = 4e^t \end{cases}$$

Find $\dfrac{\partial w}{\partial s}$ at $s=t=0$.

Soln: $$\frac{\partial w}{\partial s} = \frac{\partial w}{\partial x}\frac{\partial x}{\partial s} + \frac{\partial w}{\partial y}\frac{\partial y}{\partial s} + \frac{\partial w}{\partial z}\frac{\partial z}{\partial s}$$

$$= \frac{x}{\sqrt{x^2+y^2+z^2}} \cdot 3e^t \cos s + \frac{y}{\sqrt{x^2+y^2+z^2}} \cdot 3e^t(-\sin s)$$

$$+ \frac{z}{\sqrt{x^2+y^2+z^2}} \cdot 0$$

Put $s=t=0$, $(x = 3e^t \sin s = 0,\ y = 3e^t \cos s = 3,\ z = 4)$

$$\frac{\partial w}{\partial s}(0,0) = 0 \cdot 3 + \frac{3}{\sqrt{0^2+3^2+4^2}} \cdot 0 + 0 = 0$$

eg 4. John is walking with position at time t given by

$$\begin{cases} x(t) = t^3+1 \\ y(t) = 2t^2 \end{cases}$$

Altitude is $H(x,y) = x^2 - y^2 + 100$

(1) Is John going up/down at $t=1$?

(2) Which direction should he go instead at $t=1$ to go down most quickly?

Soln: (1) Find $\left.\frac{dH}{dt}\right|_{t=1}$ $\qquad t < {x \atop y} > H$

Chain rule: $\frac{dH}{dt} = \frac{\partial H}{\partial x}\frac{dx}{dt} + \frac{\partial H}{\partial y}\frac{dy}{dt}$

$= 2x \cdot 3t^2 + (-2y)\cdot(4t)$

At $t=1$, $(x,y) = (2,2)$

$\therefore \left.\frac{dH}{dt}\right|_{t=1} = (2\cdot 2)\cdot(3\cdot 1^2) - (2(2))\cdot(4\cdot 1)$

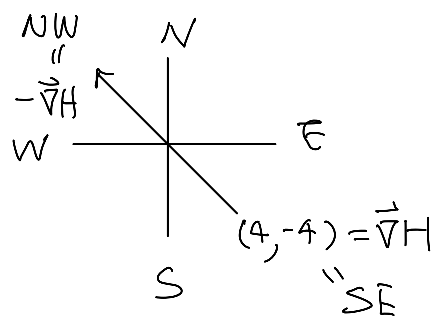
$= 12 - 16 = -4 < 0$

$\therefore$ John is going down at $t=1$.

(2) Go down most quickly in the direction $-\vec{\nabla}H$
(by the geometric interpretation of $\vec{\nabla}H$)

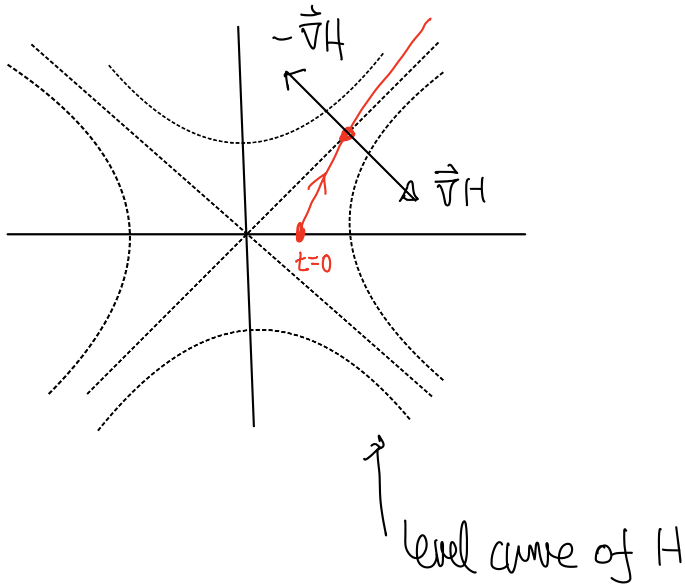
$\vec{\nabla}H = (2x, -2y)$ at $(x,y) = (2,2)$ (when $t=1$)

$= (4, -4)$

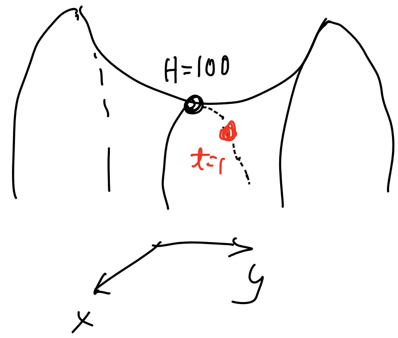


$\therefore$ H decreases most rapidly in the direction of $-\vec{\nabla}H(2,2) = (-4,4)$

i.e. John should go NW (North-West)



$H = x^2 - y^2 + 100$



Idea of Pf of Chain Rule

- $\vec{f} : \Omega_1 \to \mathbb{R}^n$ $\quad (\Omega_1 \subseteq \mathbb{R}^k$, open$)$
- $\vec{g} : \Omega_2 \to \mathbb{R}^m$ $\quad (\Omega_2 \subseteq \mathbb{R}^n$, open$)$
- $\vec{f}(\Omega_1) \subset \Omega_2$,
- $\vec{f}$ differentiable at $\vec{a} \in \Omega_1 \subset \mathbb{R}^k$
- $\vec{g}$ differentiable at $\vec{b} = \vec{f}(\vec{a}) \in \Omega_2 \subset \mathbb{R}^n$

$$\Rightarrow \quad \begin{aligned} \vec{f}(\vec{x}) - \vec{f}(\vec{a}) &= D\vec{f}(\vec{a})(\vec{x} - \vec{a}) + \vec{\varepsilon}_f(\vec{x}) \quad \text{—— (i)} \\ \vec{g}(\vec{y}) - \vec{g}(\vec{b}) &= D\vec{g}(\vec{b})(\vec{y} - \vec{b}) + \vec{\varepsilon}_g(\vec{y}) \quad \text{—— (ii)} \end{aligned}$$

Put $\vec{y} = f(\vec{x})$ (and $\vec{b} = \vec{f}(\vec{a})$), we have

$$\vec{g}(\vec{f}(\vec{x})) - \vec{g}(\vec{b}) = D\vec{g}(\vec{b})\left[D\vec{f}(\vec{a})(\vec{x} - \vec{a}) + \vec{\varepsilon}_f(\vec{x})\right] + \vec{\varepsilon}_g(\vec{f}(\vec{x}))$$

$$= D\vec{g}(\vec{b})\, D\vec{f}(\vec{a})(\vec{x} - \vec{a}) + \vec{\varepsilon}_{\vec{g}\circ\vec{f}}(\vec{x})$$

where $\vec{\varepsilon}_{\vec{g}\circ\vec{f}}(\vec{x}) = D\vec{g}(\vec{b})\vec{\varepsilon}_f(\vec{x}) + \vec{\varepsilon}_g(\vec{f}(\vec{x}))$

Then one can show that $\displaystyle\lim_{\vec{x}\to\vec{a}} \frac{\|\vec{\varepsilon}_{\vec{g}\circ\vec{f}}(\vec{x})\|}{\|\vec{x}-\vec{a}\|} = 0$

$$\therefore \quad D(\vec{g}\circ\vec{f})(\vec{a}) = D\vec{g}(\vec{f}(\vec{a}))D\vec{f}(\vec{a})$$

(Pf omitted, see MATH2050 for the case of 1-variable)

<u>Summary</u>: <u>Jacobian Matrix</u>

(1) 1-variable, real-valued $\quad f: \Omega^{\subseteq \mathbb{R}} \longrightarrow \mathbb{R}$, $\quad x \longmapsto f(x)$

$$Df(x) = \frac{df}{dx} \quad (1\times 1 \text{ matrix, a scalar})$$

(2) multivariable, real-valued $\quad f: \Omega^{\subseteq \mathbb{R}^n} \longrightarrow \mathbb{R}$, $\quad \vec{x} \longmapsto f(\vec{x})$

$$Df(\vec{x}) = \vec{\nabla} f(\vec{x}) = \left(\frac{\partial f}{\partial x_1}(\vec{x}), \cdots, \frac{\partial f}{\partial x_n}(\vec{x})\right)$$

($1\times n$ matrix, (row) vector in $\mathbb{R}^n$)

(3) Multivariable, vector-valued $\vec{f} : \Omega \subset \mathbb{R}^n \to \mathbb{R}^m$

$$\vec{x} \longmapsto \vec{f}(\vec{x})$$

$$D\vec{f}(\vec{x}) = \begin{bmatrix} — \vec{\nabla} f_1 — \\ \vdots \\ — \vec{\nabla} f_m — \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \quad \text{((}m\times n\text{) matrix)}$$

(4) 1-variable, vector-valued $\vec{\gamma} : I \subset \mathbb{R} \to \mathbb{R}^m$
(a curve in $\mathbb{R}^n$)

$$t \longmapsto \vec{\gamma}(t) = (x_1(t), \cdots, x_m(t))$$

$$D\vec{\gamma}(t) = \begin{bmatrix} \frac{dx_1}{dt} \\ \vdots \\ \frac{dx_m}{dt} \end{bmatrix} \quad (m\times 1 \text{ matrix, column vector in } \mathbb{R}^m)$$

<u>Chain Rule</u> in classical notation

$$(x_1, \cdots, x_k) \longrightarrow (y_1, \cdots, y_n) \longrightarrow (g_1, \cdots, g_m)$$

$$\begin{cases} g_i = g_i(y_1, \cdots, y_n) & \text{are functions of } y_1, \cdots, y_n \\ y_j = y_j(x_1, \cdots, x_k) & \text{are functions of } x_1, \cdots, x_k \end{cases}$$

We can regard $g_i = g_i(x_1, \cdots, x_k)$ as functions of $x_1, \cdots, x_k$

abuse of notation

Then the ij-entry of the Chain rule

$$\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_k} \end{bmatrix} = \begin{bmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_n} \\ \vdots & & \vdots \\ \frac{\partial g_m}{\partial y_1} & \cdots & \frac{\partial g_m}{\partial y_n} \end{bmatrix} \begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_k} \\ \vdots & & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_k} \end{bmatrix}$$

is

$$\boxed{\begin{aligned} \frac{\partial g_i}{\partial x_j} &= \frac{\partial g_i}{\partial y_1} \cdot \frac{\partial y_1}{\partial x_j} + \cdots + \frac{\partial g_i}{\partial y_n} \cdot \frac{\partial y_n}{\partial x_j} \\ &\left(= \sum_{l=1}^{n} \frac{\partial g_i}{\partial y_l} \cdot \frac{\partial y_l}{\partial x_j} \right) \end{aligned}}$$

Application of Chain Rule

Level Set

Recall: $f: \Omega \to \mathbb{R}$, $\Omega \subseteq \mathbb{R}^n$

$$\vec{x} \longrightarrow c = f(\vec{x}) \in \mathbb{R}$$

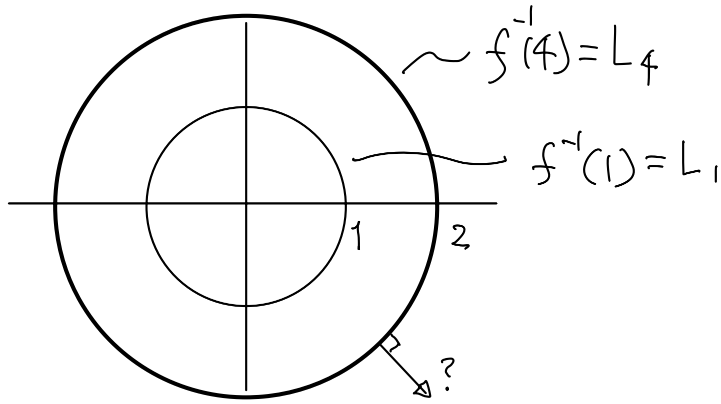
$L_c = f^{-1}(c) = \{\vec{x} \in \Omega : f(\vec{x}) = c\}$.

Level set of f (at level c).

eg: $f(x,y) = x^2+y^2$

$f^{-1}(1) = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 = 1\}$

$f^{-1}(4) = \{(x,y) \in \mathbb{R}^2 : x^2+y^2 = 4\}$



Thm Let
- $f: \Omega \to \mathbb{R}$ ($\Omega \subset \mathbb{R}^n$, open)
- $c \in \mathbb{R}$
- $\vec{a} \in S = f^{-1}(c)$ (S = a level set of f)

Suppose
- f is <u>differentiable</u> at $\vec{a}$,
- $\vec{\nabla} f(\vec{a}) \neq \vec{0}$

Then $\vec{\nabla} f(\vec{a}) \perp S$ at $\vec{a}$