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 $w = \underset{\overset{\shortparallel}{2x^2+1}}{g(x)}$)

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

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

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

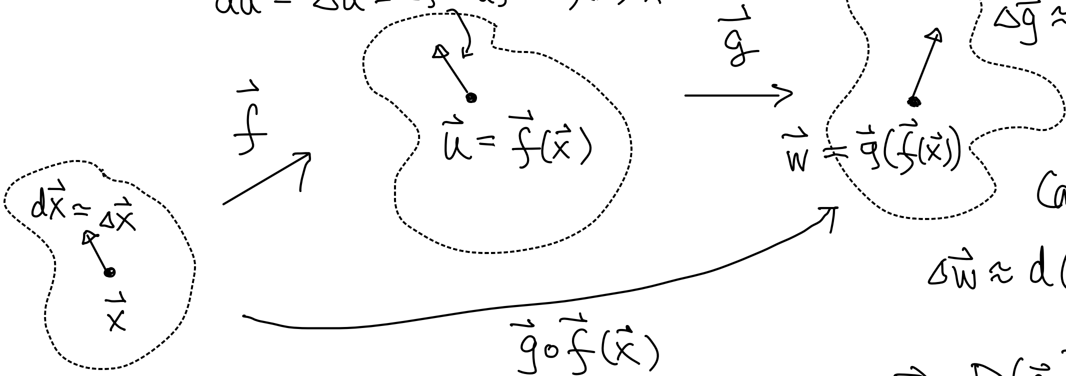
<u>Caution</u>: Abuse of notation :

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

<u>General dimensions</u>:

$d\vec{u} = \Delta\vec{u} = \Delta\vec{f} \simeq d\vec{f} = D\vec{f}(\vec{x})d\vec{x}$

$\Delta\vec{w}$
$\shortparallel$
$\Delta\vec{g} \approx d\vec{g} = D\vec{g}(\vec{u})\,d\vec{u}$
$= D\vec{g}(\vec{f}(\vec{x}))D\vec{f}(\vec{x})d\vec{x}$



Compare to

$\Delta\vec{w} \approx d(\vec{g}\circ\vec{f}) = D(\vec{g}\circ\vec{f})(\vec{x})d\vec{x}$

$\Rightarrow D(\vec{g}\circ\vec{f})(\vec{x}) = D\vec{g}(\vec{f}(\vec{x}))D\vec{f}(\vec{x})$

Thm (**Chain Rule**)

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

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 $D(\vec{g} \circ \vec{f})(\theta)$. $\quad \begin{pmatrix} u = \cos\theta \\ v = \sin\theta \end{pmatrix}$

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

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

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

By Chain Rule

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

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

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

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

eg2 (Abuse of notations)

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

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

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

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

Solu: (g is real-valued, $Dg = \vec{\nabla} g$)

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

At $(1,1)$, $\begin{cases} u = x^2 = 1 \\ v = 3xy = 3 \\ w = x + y^2 = 2 \end{cases} \Rightarrow f(1,1) = (1,3,2)$

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

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

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

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

$$= [1, 0]$$

$\left[\frac{\partial g}{\partial x}, \frac{\partial g}{\partial y} \right]$

$$\therefore \frac{\partial g}{\partial x}(1,1) = 1.$$

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

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

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

$$\begin{pmatrix} u = f_1 \\ v = f_2 \\ w = f_3 \end{pmatrix}$$

$$\frac{\partial g}{\partial x} = \frac{\partial g}{\partial u}\frac{\partial u}{\partial x} + \frac{\partial g}{\partial v}\frac{\partial v}{\partial x} + \frac{\partial g}{\partial w}\frac{\partial w}{\partial x}$$

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

ie $k=2$, $n=3$, $m=1$

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

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

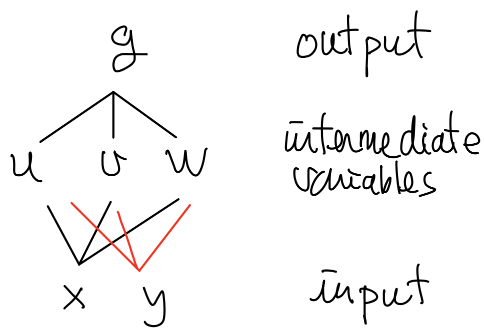
Then the Chain rule is

$$\boxed{\begin{aligned}\frac{\partial g}{\partial x} &= \frac{\partial g}{\partial u}\cdot\frac{\partial u}{\partial x} + \frac{\partial g}{\partial v}\cdot\frac{\partial v}{\partial x} + \frac{\partial g}{\partial w}\cdot\frac{\partial w}{\partial x}\\[2ex] \frac{\partial g}{\partial y} &= \frac{\partial g}{\partial u}\cdot\frac{\partial u}{\partial y} + \frac{\partial g}{\partial v}\cdot\frac{\partial v}{\partial y} + \frac{\partial g}{\partial w}\cdot\frac{\partial w}{\partial y}\end{aligned}}$$

$$\underbrace{\begin{bmatrix}\frac{\partial g}{\partial x} & \frac{\partial g}{\partial y}\end{bmatrix}}_{(1\times 2 \text{ matric})} = D(g\circ\vec{f}) = Dg(\vec{f})\, D\vec{f} = \underbrace{\begin{bmatrix}\frac{\partial g}{\partial u} & \frac{\partial g}{\partial v} & \frac{\partial g}{\partial w}\end{bmatrix}\begin{bmatrix}\frac{\partial u}{\partial x} & \frac{\partial u}{\partial y}\\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y}\\ \frac{\partial w}{\partial x} & \frac{\partial w}{\partial y}\end{bmatrix}}_{(1\times 3)(3\times 2) \text{ matric}}$$

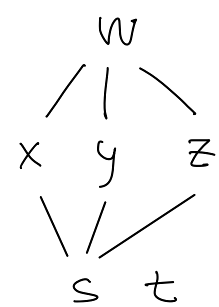
(Similarly for other low dimensional situations)

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



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

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



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

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

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

$$+\left(\frac{\partial}{\partial z}\sqrt{x^2+y^2+z^2}\right)\frac{\partial}{\partial s}(4e^t)$$

$$= \frac{x}{\sqrt{x^2+y^2+z^2}}\, 3e^t \cos s + \frac{y}{\sqrt{x^2+y^2+z^2}}(-3e^t \sin s) + 0$$

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

$$\frac{\partial W}{\partial s}(0,0) = 0+0 = 0 \quad ※$$

eg4. 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 $\left.\frac{dH}{dt}\right|_{t=1}$

$$t < {x \atop y} > H$$

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

$$\therefore \quad \left.\frac{dH}{dt}\right|_{t=1} = (2\cdot 2)\cdot(3\cdot 1^2) - 2(2)\cdot(4\cdot 1) = -4 < 0$$

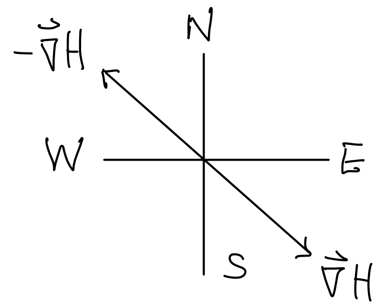
$\therefore$ John is <u>going down</u> at $t=1$.

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

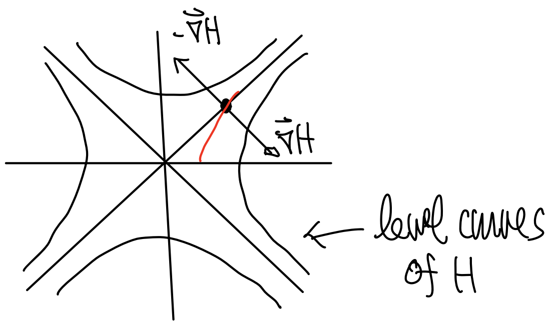
$$\vec{\nabla} H = (2x, -2y) \quad \text{at } (x,y)=(2,2) \text{ (when } t=1)$$
$$= (4,-4)$$

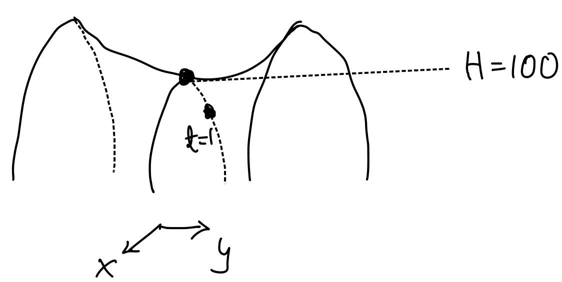
$\therefore$ H decreases most rapidly in the <u>direction of</u>

$$-\vec{\nabla} H(2,2) = (-4,4)$$



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





Idea of Pf of Chain Rule

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

$$\Rightarrow \begin{cases} \vec{f}(\vec{x}) - \vec{f}(\vec{a}) = D\vec{f}(\vec{a})(\vec{x}-\vec{a}) + \vec{\varepsilon}_{\vec{f}}(\vec{x}) & \text{———— (1)} \\ \vec{g}(\vec{y}) - \vec{g}(\vec{b}) = D\vec{g}(\vec{b})(\vec{y}-\vec{b}) + \vec{\varepsilon}_{\vec{g}}(\vec{y}) & \text{———— (2)} \end{cases}$$

Put $\vec{y} = f(\vec{x})$ (and $\vec{b} = \vec{f}(\vec{a})$) in (2), we have

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

$$(\text{by } (1)) \qquad = D\vec{g}(\vec{b})\left(D\vec{f}(\vec{a})(\vec{x}-\vec{a}) + \vec{\varepsilon}_{\vec{f}}(\vec{x})\right) + \vec{\varepsilon}_{\vec{g}}(\vec{f}(\vec{x}))$$

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

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

So one need to show that

$$\lim_{\vec{x}\to\vec{a}} \frac{\| D\vec{g}(\vec{b})\vec{\varepsilon}_{\vec{f}}(\vec{x}) + \vec{\varepsilon}_{\vec{g}}(\vec{f}(\vec{x})) \|}{\|\vec{x}-\vec{a}\|} = 0$$

(Proof: Omitted
see MATH 2050
for 1-variable case)

in order to prove that $D(\vec{g}\circ\vec{f})(\vec{x}) = D\vec{g}(\vec{b})D\vec{f}(\vec{a})$. #

Summary: Jacobian Matrix

(1) 1-variable, real-valued :

$$\begin{array}{ccc} f : \overset{\subseteq \mathbb{R}}{\Omega} & \longrightarrow & \mathbb{R} \\ \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\ x & \longmapsto & f(x) \end{array}$$

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

(2) multivariable, real-valued

$$\begin{array}{ccc} f : \overset{\subseteq \mathbb{R}^n}{\Omega} & \longrightarrow & \mathbb{R} \\ \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\ \vec{x} & \longmapsto & f(\vec{x}) \end{array}$$

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

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

(3) Multivariable, vector-valued

$$\begin{array}{ccc} \vec{f} : \overset{\subseteq \mathbb{R}^n}{\Omega} & \longrightarrow & \mathbb{R}^m \\ \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\ \vec{x} & \longmapsto & \vec{f}(\vec{x}) \end{array}$$

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

(4) 1-variable, vector-valued
(a curve in $\mathbb{R}^n$)

$$\begin{array}{ccl} \vec{\gamma} : \overset{\subset \mathbb{R}}{I} & \longrightarrow & \mathbb{R}^m \\ \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\ t & \longmapsto & \vec{\gamma}(t) = (x_1(t), \cdots, x_m(t)) \end{array}$$

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

Chain Rule 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

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