MATH 2010 Chapter 8

8.1 Matrix Multiplication

A be an $m \times n$ (m rows, n columns) matrix. Let $b = \begin{bmatrix} | \\ \vec{b} \\ | \end{bmatrix}$ be a (column) vector in $\mathbb{R}^n$.

If we view A as a collection of row vectors:

$$A = \begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \dots & a_{mn} \end{bmatrix} = \begin{bmatrix} -\vec{a}_1- \\ \vdots \\ -\vec{a}_m- \end{bmatrix},$$

then by definition of matrix multiplication we have:

$$A\vec{b} = \begin{bmatrix} -\vec{a}_1- \\ \vdots \\ -\vec{a}_m- \end{bmatrix} \begin{bmatrix} | \\ \vec{b} \\ | \end{bmatrix} = \begin{bmatrix} \vec{a}_1 \cdot \vec{b} \\ \vdots \\ \vec{a}_m \cdot \vec{b} \end{bmatrix} \in \mathbb{R}^m$$

Now let, B be an $n \times k$ matrix. Then, view B as a collection of column vectors:

$$B = \begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_k \\ | & & | \end{bmatrix},$$

we have:

$$AB = A \begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_k \\ | & & | \end{bmatrix} = \begin{bmatrix} | & & | \\ A\vec{b}_1 & \cdots & A\vec{b}_k \\ | & & | \end{bmatrix}$$

Alternatively, we also have:

$$AB = \begin{bmatrix} -\vec{a}_1- \\ \vdots \\ -\vec{a}_m- \end{bmatrix} B = \begin{bmatrix} -\vec{a}_1 B- \\ \vdots \\ -\vec{a}_m B- \end{bmatrix},$$

where:

$$\vec{a}_i B = (\vec{a}_i \cdot \vec{b}_1,\ \vec{a}_i \cdot \vec{b}_2,\ \ldots,\ \vec{a}_i \cdot \vec{b}_k)$$

Example 8.1.

$$\overset{\displaystyle A}{\overset{\shortparallel}{\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}}} \overset{\displaystyle B}{\overset{\shortparallel}{\begin{bmatrix} 5 & 6 & 7 \\ 8 & 9 & 10 \end{bmatrix}}} = \begin{bmatrix} 21 & 24 & 27 \\ 47 & 54 & 61 \end{bmatrix}$$

$$A \begin{bmatrix} 5 \\ 8 \end{bmatrix} = \begin{bmatrix} 21 \\ 47 \end{bmatrix} \quad A \begin{bmatrix} 6 \\ 9 \end{bmatrix} = \begin{bmatrix} 24 \\ 54 \end{bmatrix} \quad A \begin{bmatrix} 7 \\ 10 \end{bmatrix} = \begin{bmatrix} 27 \\ 61 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 2 \end{bmatrix} B = \begin{bmatrix} 21 & 24 & 27 \end{bmatrix} \quad \begin{bmatrix} 3 & 4 \end{bmatrix} B = \begin{bmatrix} 47 & 54 & 61 \end{bmatrix}$$

8.2 Vector-valued Functions

Let $\vec{f} : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$

$$\vec{f}(\vec{x}) = \underbrace{\big(f_1(\vec{x}), \cdots, f_m(\vec{x})\big)}_{\text{vector-valued}} = \begin{bmatrix} f_1(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix}$$

Suppose $\frac{\partial f_i}{\partial x_j}(\vec{a})$ exists for each i, j . For each $1 \leqslant i \leqslant m$,

$$\underset{1\times 1}{f_i(\vec{x})} = \underset{1\times 1}{f_i(\vec{a})} + \underset{1\times n}{\nabla f_i(\vec{a})} \cdot \underset{n\times 1}{(\vec{x} - \vec{a})} + \underset{1\times 1}{\varepsilon_i(\vec{x})} \quad \circledast$$

Here, regard $\nabla f_i(\vec{a})$ as a row vector and $\vec{x} - \vec{a}$ as a column vector,
in order to use multiplication Writing $\circledast$ for $1 \leqslant i \leqslant m$ in a matrix:

$$\begin{bmatrix} f_1(\vec{x}) \\ \vdots \\ f_m(\vec{x}) \end{bmatrix} = \begin{bmatrix} f_1(\vec{a}) \\ \vdots \\ f_m(\vec{a}) \end{bmatrix} + \underbrace{\begin{bmatrix} -\nabla f_1(\vec{a})- \\ \vdots \\ -\nabla f_m(\vec{a})- \end{bmatrix}}_{m\times n \text{ matrix of } \frac{\partial f_i}{\partial x_j}} \begin{bmatrix} x_1 - a_1 \\ \vdots \\ x_n - a_n \end{bmatrix} + \underbrace{\begin{bmatrix} \varepsilon_1(\vec{x}) \\ \vdots \\ \varepsilon_m(\vec{x}) \end{bmatrix}}_{\text{Errors}}$$

Definition 8.2. The **Jacobian matrix** of $\vec{f}$ at $\vec{a}$ is:

$$D\vec{f}(\vec{a}) = \begin{bmatrix} -\nabla f_1(\vec{a})- \\ \vdots \\ -\nabla f_m(\vec{a})- \end{bmatrix} \quad (m \times n \text{ matrix })$$

The **linearization** of $\vec{f}$ at $\vec{a}$ is:

$$\vec{L}(\vec{x}) = \vec{f}(\vec{a}) + D\vec{f}(\vec{a})(\vec{x} - \vec{a})$$

The function $\vec{f}$ is said to be **differentiable** at $\vec{a}$ if the **error term**:

$$\vec{\varepsilon}(\vec{x}) := \vec{f}(\vec{x}) - \vec{L}(\vec{x})$$

of the linearization $\vec{L}$ of $\vec{f}$ satisfies:

$$\lim_{\vec{x}\to\vec{a}} \frac{\|\vec{\varepsilon}(\vec{x})\|}{\|\vec{x} - \vec{a}\|} = 0.$$

Remark. 1.

$$[D\vec{f}(\vec{a})]_{ij} = \frac{\partial f_i}{\partial x_j}(\vec{a})$$

2.

$$\underset{m\times 1}{\vec{f}(\vec{x})} = \underset{m\times 1}{\vec{f}(\vec{a})} + \underset{m\times n}{D\vec{f}(\vec{a})}\underset{n\times 1}{(\vec{x} - \vec{a})} + \underset{m\times 1}{\vec{\varepsilon}(\vec{x})}$$

3. If f is real-valued $(m = 1)$, then

$$Df(\vec{a}) = \nabla f(\vec{a})$$

4. $\|\vec{\varepsilon}(\vec{x})\|$, $\|\vec{x} - \vec{a}\|$ are lengths in $\mathbb{R}^m, \mathbb{R}^n$, respectively.

5.

$$\lim_{\vec{x}\to\vec{a}} \frac{\|\vec{\varepsilon}(\vec{x})\|}{\|\vec{x} - \vec{a}\|} = 0 \Leftrightarrow \lim_{\vec{x}-\vec{a}} \frac{\varepsilon_i(\vec{x})}{\|\vec{x} - \vec{a}\|} = 0,$$

for all $i = 1, \cdots, m$.

Hence,

$\vec{f}$ is differentiable at $\vec{a} \Leftrightarrow f_i$ is differentiable at $\vec{a}$,

for all $i = 1, \cdots, m$.

8.2.1 Approximation:

$$\vec{f}(\vec{x}) \approx L(\vec{x}) = \vec{f}(\vec{a}) + D\vec{f}(\vec{a})(\vec{x} - \vec{a})$$

$$\Rightarrow \underbrace{\vec{f}(\vec{x}) - \vec{f}(\vec{a})}_{\Delta\vec{f}=\text{ change in } f} \approx \underbrace{D\vec{f}(\vec{a})}_{\text{Jacobian Matrix}} \times \underbrace{(\vec{x} - \vec{a})}_{\Delta\vec{x}=\text{ change in } \vec{x}}$$

Can consider $D\vec{f}(\vec{a})$ as a linear map:

$$D\vec{f}(\vec{a}) : \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

$$\Delta\vec{x} \longmapsto \underset{\text{approximated change in } f}{D\vec{f}(\vec{a})\Delta\vec{x} = d\vec{f}}$$

$$\underset{\text{(vector)}}{\Delta\vec{f}} \approx d\vec{f} = \underset{\text{(matrix)}}{D\vec{f}(\vec{a})} \times \underset{\text{(vector)}}{d\vec{x}}$$

Remark. Compare with $f : \mathbb{R} \to \mathbb{R}$

$$\underset{\text{(number)}}{\Delta y} \approx df = \underset{\text{(number)}}{f'(a)} \times \underset{\text{(number)}}{\Delta x}$$

Example 8.3.

$$\begin{aligned} \vec{f}(x, y) &= [\overset{\overset{f_1}{\shortparallel}}{(y + 1)\ln x}, \overset{\overset{f_2}{\shortparallel}}{x^2 - \sin y + 1}] \\ &= \begin{bmatrix} (y+1)\ln x \\ x^2 - \sin y + 1 \end{bmatrix} \text{ (rewrite as column vector)} \end{aligned}$$

1. Find $D\vec{f}(1, 0)$
2. Approximate $\vec{f}(0.9, 0.1)$

Solution.

$$f_1(x, y) = (y + 1)\ln x$$

$$f_2(x, y) = x^2 - \sin y + 1$$

$$\nabla f_1 = \begin{bmatrix} \frac{y+1}{x} & \ln x \end{bmatrix}$$

$$\nabla f_2 = [2x \quad -\cos y]$$

$$\Rightarrow D\vec{f}(x,y) = \begin{bmatrix} \frac{y+1}{x} & \ln x \\ 2x & -\cos y \end{bmatrix}$$

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

Linearization of $\vec{f}$ at (1,0):

$$\begin{aligned} \vec{L}(x,y) &= \vec{f}(1,0) + D\vec{f}(1,0)(\begin{bmatrix} x \\ y \end{bmatrix} - \begin{bmatrix} 1 \\ 0 \end{bmatrix}) \\ &= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} x-1 \\ y \end{bmatrix} \end{aligned}$$

$$\begin{aligned} \vec{f}(0.9, 0.1) &\approx \vec{L}(0.9, 0.1) \\ &= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix} \underbrace{\begin{bmatrix} 0.9-1 \\ 0.1 \end{bmatrix}}_{\Delta\vec{x}=d\vec{x}=\text{change in } \vec{x}} \\ &= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \underbrace{\begin{bmatrix} -0.1 \\ -0.3 \end{bmatrix}} d\vec{f}\text{=approximated change of } \vec{f} \\ &= \begin{bmatrix} -0.1 \\ 1.7 \end{bmatrix} \end{aligned}$$

Remark. Actual change in $\vec{f}$:

$$\Delta\vec{f} = \vec{f}(0.9, 0.1) - \vec{f}(1,0) = \begin{bmatrix} -0.1159\cdots \\ -0.2898\cdots \end{bmatrix}$$

Remark. Total differential can also be written in matrix form:

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

$$\vec{f} = \begin{bmatrix} f_1 \\ \vdots \\ f_m \end{bmatrix}$$

$$d\vec{f} = \begin{bmatrix} df_1 \\ \vdots \\ df_m \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} \begin{bmatrix} dx_1 \\ \vdots \\ dx_n \end{bmatrix} = Df(\vec{a})d\vec{x}$$

8.3 Chain Rule

Recall the **chain rule** for functions in one variable:

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

$$\begin{aligned} (g \circ f)'(x) &= g'(f(x))f'(x) \text{ or} \\ \frac{dw}{dx} &= \frac{dw}{du} \cdot \frac{du}{dx} \\ &= 2 \cdot 2x = 4x \end{aligned}$$

For multivariable functions,

Theorem 8.4 (Chain Rule)**.** *Let:*

$$\vec{f} : \Omega_1 \subseteq \mathbb{R}^k \longrightarrow \mathbb{R}^n$$

$$\vec{g} : \Omega_2 \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$$

Suppose that $\vec{f}$ is differentiable at $\vec{a}$, and $\vec{g}$ is differentiable at $\vec{b} = \vec{f}(\vec{a})$.

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

$$\underbrace{D(\vec{g} \circ \vec{f})(\vec{a})}_{m \times k} = \underbrace{(D\vec{g})(f(\vec{a}))}_{m \times n} \underbrace{(D\vec{f})(\vec{a})}_{n \times k}$$

Remark. For simplicity, we might omit $\rightarrow$ for vectors

From now on: $\vec{f} = f$, $\vec{x} = x$

Example 8.5. Let:

$$\begin{aligned} f &: \mathbb{R} \to \mathbb{R}^2, \\ g &: \mathbb{R}^2 \to \mathbb{R}^2, \end{aligned}$$

where:

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

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

Find $D(g \circ f)(\theta)$.

Solution. **Method 1** Find composition explicitly.

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

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

Method 2 Chain Rule

$$Df(\theta) = \begin{bmatrix} f_1{}' \\ f_2{}' \end{bmatrix} = \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix}$$

$$Dg(u, v) = \begin{bmatrix} \nabla g_1 \\ \nabla g_2 \end{bmatrix} = \begin{bmatrix} 2v & 2u \\ 2u & -2v \end{bmatrix}$$

$$Dg(f(\theta)) = Dg(\cos\theta, \sin\theta) = \begin{bmatrix} 2\sin\theta & 2\cos\theta \\ 2\cos\theta & -2\sin\theta \end{bmatrix}$$

By Chain Rule,

$$\begin{aligned}D(g \circ f)(\theta) &= Dg(f(\theta))Df(\theta) \\ &= \begin{bmatrix} 2\sin\theta & 2\cos\theta \\ 2\cos\theta & -2\sin\theta \end{bmatrix} \begin{bmatrix} -\sin\theta \\ \cos\theta \end{bmatrix} \\ &= \begin{bmatrix} -2\sin^2\theta + 2\cos^2\theta \\ -4\cos\theta\sin\theta \end{bmatrix} \\ &= \begin{bmatrix} 2\cos 2\theta \\ -2\sin 2\theta \end{bmatrix} \text{(same answer)}\end{aligned}$$

Example 8.6.

$$f(x, y) = (x^2, 3xy, x + y^2)$$

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

Consider $g \circ f$:

$$\begin{matrix} x \\ y \end{matrix} \overset{f}{\longmapsto} \begin{matrix} f_1 = u \\ f_2 = v \\ f_3 = w \end{matrix} \overset{g}{\longmapsto} g$$

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

Solution.

$$Dg = \nabla g = \begin{bmatrix} \frac{w}{v} & -\frac{uw}{v^2} & \frac{u}{v} \end{bmatrix}$$

$$\begin{aligned} Dg(f(1,1)) &= Dg(1,3,2) \\ &= \begin{bmatrix} \frac{2}{3} & -\frac{2}{9} & \frac{1}{3} \end{bmatrix} \end{aligned}$$

$$Df = \begin{bmatrix} \nabla f_1 \\ \nabla f_2 \\ \nabla f_3 \end{bmatrix} = \begin{bmatrix} 2x & 0 \\ 3y & 3x \\ 1 & 2y \end{bmatrix}$$

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

Hence,

$$\begin{aligned} D(g \circ f)(1,1) &= Dg(f(1,1))Df(1,1) \\ &= \begin{bmatrix} \frac{2}{3} - \frac{2}{9} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 3 & 3 \\ 1 & 2 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 \end{bmatrix} \end{aligned}$$

Note $D(g \circ f) = \begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \end{bmatrix}$

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

In the previous example, we have:

$$D(g \circ f) = Dg\,Df$$

$$\begin{bmatrix} 1 & 0 \end{bmatrix} = \begin{bmatrix} \frac{2}{3} & -\frac{2}{9} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 3 & 3 \\ 1 & 2 \end{bmatrix} \quad \begin{matrix} f_1 = u \\ f_2 = v \\ f_3 = w \end{matrix}$$

From matrix multiplication, we get another form of chain rule (in classical notation)

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

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

Example 8.7.

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

where:

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

Find $\frac{\partial w}{\partial s}$ at $(s, t) = (0, 0)$.

Solution.

$$\begin{aligned} \frac{\partial w}{\partial s} &= \frac{\partial w}{\partial x} \cdot \frac{\partial x}{\partial s} + \frac{\partial w}{\partial y} \cdot \frac{\partial y}{\partial s} + \frac{\partial w}{\partial z} \cdot \frac{\partial z}{\partial s} \\ &= \frac{x}{\sqrt{x^2 + y^2 + z^2}} \cdot 3e^t \cos s - \frac{y}{\sqrt{x^2 + y^2 + z^2}} \cdot 3e^t \sin s + \frac{z}{\sqrt{x^2 + y^2 + z^2}}(0) \end{aligned}$$

$s = t = 0 \Rightarrow (x, y, z) = (0, 3, 4)$. Hence,

$$\left.\frac{\partial w}{\partial s}\right|_{(s,t)=(0,0)} = 0 - \frac{3}{5}(0) + 0 = 0.$$

Example 8.8. John is hiking with position at time t given by:

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

His altitude is given by: $H(x, y) = x^2 - y^2 + 100$

1. Is John going up/down at $t = 1$?
2. Which direction should he go instead at $t = 1$ to go down most quickly?

Solution. 1. Find $\left.\frac{dH}{dt}\right|_{t=1}$:

$$\begin{aligned}\frac{dH}{dt} &= \frac{\partial H}{\partial x}\cdot\frac{dx}{dt}+\frac{\partial H}{\partial y}\cdot\frac{dy}{dt}\\ &= (2x)(3t^2)+(-2y)(4t)\\ &= 2(t^3+1)(3t^2)-2(2t^2)(4t)\\ &= 6t^5-16t^3+6t^2\end{aligned}$$

$\therefore \left.\frac{dH}{dt}\right|_{t=1} = 6-16+6=-4<0$

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

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

$$\nabla H = (2x,-2y)$$

$$\nabla H = (4,-4)$$

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

$\therefore$ John should go NW .

Remark.

$$\frac{dH}{dt} = \overset{\text{slope in } x\text{- and } y\text{- direction}}{\overset{\downarrow}{\frac{\partial H}{\partial x}}\cdot\underset{\uparrow}{\frac{dx}{dt}}+\overset{\downarrow}{\frac{\partial H}{\partial y}}\cdot\underset{\uparrow}{\frac{dy}{dt}}}_{\text{velocity in } x\text{- and } y\text{- direction}} = \overset{\text{gradient}}{\overset{\downarrow}{\nabla H}} \cdot \underset{\text{velocity}}{\underset{\uparrow}{\begin{bmatrix}\frac{dx}{dt}\,\frac{dy}{dt}\end{bmatrix}}}$$

8.3.1 Idea of Proof of Chain Rule

Suppose

$$\begin{aligned}&f:\Omega_1\subseteq\mathbb{R}^k\longmapsto\mathbb{R}^n,\ \text{differentiable at } a\\ &g:\Omega_2\subseteq\mathbb{R}^n\longmapsto\mathbb{R}^m,\ \text{differentiable at } b=f(a)\in\Omega_2\end{aligned}$$

For

$$x\in\Omega_1,\qquad f(x)-f(a)=Df(a)(x-a)+\varepsilon_f(x) \tag{8.1}$$
$$y\in\Omega_2,\qquad g(y)-g(b)=Dg(b)(y-b)+\varepsilon_g(y) \tag{8.2}$$

Put $y=f(x), b=f(a)$ and (1) into (2):

$$\begin{aligned}g(f(x))-g(f(a)) &= Dg(f(a))[Df(a)(x-a)+\varepsilon_f(x)]+\varepsilon_g(f(x))\\ &= \underbrace{Dg(f(a))Df(a)(x-a)}_{\text{linear in } x-a}+\underbrace{Dg(f(a))\varepsilon_f(x)+\varepsilon_g(f(x))}_{\text{Denote this by } \varepsilon_{g\circ f}(x)}\end{aligned}$$

Then, show that:

$$\lim_{x \to a} \frac{\|\varepsilon_{g\circ f}(x)\|}{\|x-a\|} = 0.$$

Sketch of the argument: For x close to a, the continuity of f at a implies that $\|f(x) - f(a)\|$ is small. The differentiability of g at $f(a)$ then implies that $\varepsilon_g(f(x)))$ is small.

Similarly, the differentiability of f at a implies that $\varepsilon_f(x)$ is small. Hence, $Dg(f(a))\varepsilon_f(x)$ is small.

Hence, $g \circ f$ is differentiable at a, with:

$$D(g \circ f)(a) = Dg(f(a))\, Df(a).$$

8.3.2 Summary

Jacobian Matrix

1. $f : \Omega \subseteq \underset{x \longmapsto f(x)}{\mathbb{R} \to \mathbb{R}}$ (1-variable, real-valued)

   $Df(x) = \frac{df}{dx}$ (scalar, 1×1 matrix)

2. $\underset{x=(x_1,\cdots,x_n) \longmapsto f(x)=f(x_1,\cdots,x_n)}{f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}}$ (multivariable, real-valued)

$$\begin{aligned} Df(x) &= \nabla f(x) \\ &= \left(\frac{\partial f}{\partial x_1}(x), \cdots, \frac{\partial f}{\partial x_n}(x) \right) \quad \begin{pmatrix} \text{vectors in } \mathbb{R}^n \\ 1 \times n \text{ matrix} \end{pmatrix} \end{aligned}$$

3. $f : \Omega \subseteq \mathbb{R}^n \to \mathbb{R}^m$(multivariable, vector-valued)

$$x=(x_1,\cdots,x_n) \longmapsto \begin{bmatrix} f_1(x) \\ \vdots \\ f_m(x) \end{bmatrix} \qquad f_i(x)=f_i(x_1,\cdots,x_n)$$

$$Df(x) = \begin{bmatrix} -\nabla f_1- \\ \vdots \\ -\nabla f_m- \end{bmatrix} = \begin{bmatrix} \frac{\partial f_1}{\partial x_1} & \cdots & \frac{\partial f_1}{\partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_m}{\partial x_1} & \cdots & \frac{\partial f_m}{\partial x_n} \end{bmatrix} (m \times n \text{ matrix})$$

Chain Rule

$$(x_1, \cdots, x_k) \xmapsto{f} (y_1, \cdots, y_n) \xmapsto{g} (g_1, \cdots, g_m)$$

$g_i = g_i(y_1, \cdots, y_n)$ are functions of $y_1, \cdots, y_n$
$y_j = f_j = f_j(x_1, \cdots, x_k)$ are functions of $x_1, \cdots, x_k$
$\therefore$ We can regard $g_i = g_i(x_1, \cdots, x_k)$ as functions of $x_1, \cdots, x_k$

Chain Rule in Matrix Notation

$$\underbrace{\begin{bmatrix} \frac{\partial g_1}{\partial x_1} & \cdots & \frac{\partial g_1}{\partial x_k} \\ \vdots & \cdots & \vdots \\ \frac{\partial g_m}{\partial x_1} & \cdots & \frac{\partial g_m}{\partial x_k} \end{bmatrix}}_{m \times k} = \underbrace{\begin{bmatrix} \frac{\partial g_1}{\partial y_1} & \cdots & \frac{\partial g_1}{\partial y_n} \\ \vdots & \cdots & \vdots \\ \frac{\partial g_m}{\partial y_1} & \cdots & \frac{\partial g_m}{\partial y_n} \end{bmatrix}}_{m \times n} \underbrace{\begin{bmatrix} \frac{\partial y_1}{\partial x_1} & \cdots & \frac{\partial y_1}{\partial x_k} \\ \vdots & \cdots & \vdots \\ \frac{\partial y_n}{\partial x_1} & \cdots & \frac{\partial y_n}{\partial x_k} \end{bmatrix}}_{n \times k}$$

By definition of matrix multiplication:

$$\frac{\partial g_i}{\partial x_j} = \frac{\partial g_i}{\partial y_1}\frac{\partial y_1}{\partial x_j} + \cdots + \frac{\partial g_i}{\partial y_n}\frac{\partial y_n}{\partial x_j}$$