Math 2010 Week 8

Recall: Matrix multiplication

Let A, B be matrices,

A is $\underbrace{m\times n}_{\text{m rows, n columns}}$, B is $n\times k$ $\Rightarrow$ AB is $m\times k$

Let $A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} — \vec{a}_1 — \\ \vdots \\ — \vec{a}_m — \end{bmatrix}$

If $b = \begin{bmatrix} | \\ \vec{b} \\ | \end{bmatrix} \in \mathbb{R}^n$ (written as a column vector)

then

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

Similarly,

$$\underset{\substack{1\times n \text{ matrix} \\ \text{(row vector)}}}{\begin{bmatrix} — \vec{a} — \end{bmatrix}} \underset{n\times k \; = B}{\begin{bmatrix} | & & | \\ \vec{b}_1 & \cdots & \vec{b}_k \\ | & & | \end{bmatrix}} = \underset{1\times k}{\begin{bmatrix} \vec{a}\cdot\vec{b}_1 & \cdots & \vec{a}\cdot\vec{b}_k \end{bmatrix}}$$

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

$$= \begin{bmatrix} — \vec{a}_1 — \\ \vdots \\ — \vec{a}_m — \end{bmatrix} B = \begin{bmatrix} — \vec{a}_1 B — \\ \vdots \\ — \vec{a}_m B — \end{bmatrix}$$

e.g.

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

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

$$\begin{bmatrix} 1 & 2 \end{bmatrix} B = \begin{bmatrix} 21 & 24 & 27 \end{bmatrix}$$

$$\begin{bmatrix} 3 & 4 \end{bmatrix} B = \begin{bmatrix} 47 & 54 & 61 \end{bmatrix}$$

Vector-valued Functions

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

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

Suppose $\frac{\partial f_i}{\partial x_j}(\vec{a})$ exists for each i, j

For each $1 \le i \le m$,

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

Here, regard $\vec{\nabla} f_i(\vec{a})$ as a row vector

$\vec{x} - \vec{a}$ as a column vector

in order to use matrix multiplication

Writing $(*)$ for $1 \le i \le m$ in a matrix:

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

<u>Defn</u> Define the <u>Jacobian matrix</u> of $\vec{f}$ at $\vec{a}$ to be

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

and the <u>linearization</u> of $\vec{f}$ at $\vec{a}$ to be

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

$\vec{f}$ is said to be <u>differentiable</u> at $\vec{a}$ if

<u>error term</u> $\vec{\varepsilon}(\vec{x}) = \vec{f}(\vec{x}) - \vec{L}(\vec{x})$

of linearization satisfies $\displaystyle\lim_{\vec{x}\to\vec{a}} \frac{\|\vec{\varepsilon}(\vec{x})\|}{\|\vec{x}-\vec{a}\|} = 0$

Rmk

① $\left[D\vec{f}(\vec{a})\right]_{ij} = \dfrac{\partial f_i}{\partial x_j}(\vec{a})$

② $\vec{f}(\vec{x}) = \vec{f}(\vec{a}) + D\vec{f}(\vec{a})(\vec{x}-\vec{a}) + \vec{\varepsilon}(\vec{x})$

$m\times 1 \quad m\times 1 \quad m\times n \quad n\times 1 \quad m\times 1$

③ If f is real-valued ($m=1$), then

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

④ $\|\vec{\varepsilon}(\vec{x})\|$, $\|\vec{x}-\vec{a}\|$ is length in $\mathbb{R}^m$, $\mathbb{R}^n$ resp.

⑤ $\displaystyle\lim_{\vec{x}\to\vec{a}} \frac{\|\vec{\varepsilon}(\vec{x})\|}{\|\vec{x}-\vec{a}\|} = 0 \Leftrightarrow \lim_{\vec{x}\to\vec{a}} \frac{\varepsilon_i(\vec{x})}{\|\vec{x}-\vec{a}\|} = 0 \quad \forall i = 1,\cdots,m$

Hence

$$\boxed{\vec{f} \text{ is differentiable at } \vec{a} \Leftrightarrow f_i \text{ is differentiable at } \vec{a},\ \forall i=1,\cdots,m}$$

Approximation:

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

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

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

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

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

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

<u>Rmk</u> Compare with $f: \mathbb{R} \to \mathbb{R}$

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

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

$$\vec{f}(x,y) = [\underbrace{(y+1)\ln x}_{= f_1}, \underbrace{x^2 - \sin y + 1}_{= f_2}]$$

$$= \begin{bmatrix} (y+1)\ln x \\ x^2 - \sin y + 1 \end{bmatrix}$$

(Rewrite as column vector)

① Find $D\vec{f}(1,0)$

② Approximate $\vec{f}(0.9, 0.1)$

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

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

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

$$\vec{\nabla} f_2 = \begin{bmatrix} 2x & -\cos y \end{bmatrix}$$

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

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

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

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

$$= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}\begin{bmatrix} x-1 \\ y \end{bmatrix}$$

$$\vec{f}(0.9, 0.1) \approx \vec{L}(0.9, 0.1)$$

$$= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 2 & -1 \end{bmatrix}\underbrace{\begin{bmatrix} 0.9-1 \\ 0.1 \end{bmatrix}}_{\Delta\vec{x} = d\vec{x} = \text{change in } \vec{x}}$$

$$= \begin{bmatrix} 0 \\ 2 \end{bmatrix} + \underbrace{\begin{bmatrix} -0.1 \\ -0.3 \end{bmatrix}}_{d\vec{f} = \text{approximated change of } \vec{f}}$$

$$= \begin{bmatrix} -0.1 \\ 1.7 \end{bmatrix}$$

Rmk Actual change in $\vec{f}$:

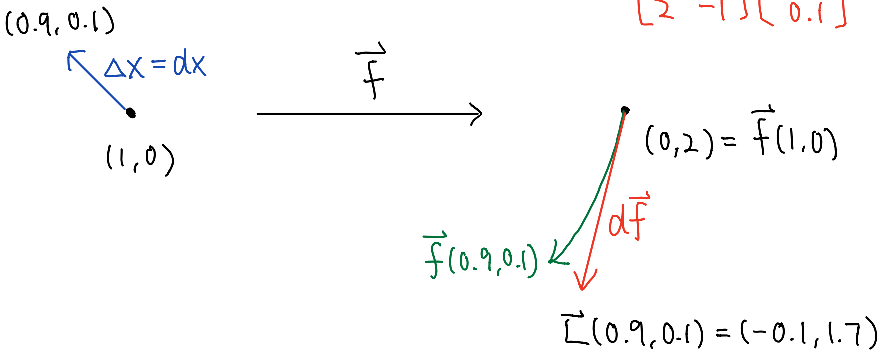
$$\Delta\vec{f} = \vec{f}(0.9, 0.1) - \vec{f}(1,0) = \begin{bmatrix} -0.1159\ldots \\ -0.2898\ldots \end{bmatrix}$$

Picture $\vec{f}(1,0)=\begin{bmatrix}0\\2\end{bmatrix}$ $D\vec{f}(1,0)=\begin{bmatrix}1&0\\2&-1\end{bmatrix}$. Approximate $\vec{f}(0.9,0.1)$.

$$\Delta\vec{x}=\begin{bmatrix}0.9\\0.1\end{bmatrix}-\begin{bmatrix}0\\2\end{bmatrix}=\begin{bmatrix}-0.1\\0.1\end{bmatrix}$$

$$\Delta\vec{f}\approx d\vec{f}=D\vec{f}(1,0)\,\Delta\vec{x}$$

$$=\begin{bmatrix}1&0\\2&-1\end{bmatrix}\begin{bmatrix}-0.1\\0.1\end{bmatrix}=\begin{bmatrix}-0.1\\-0.3\end{bmatrix}$$



Rmk Total differential can also be written in matrix form

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

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

Chain Rule

(Au 4.1, 4.2
Thomas 14.4)

Recall One variable:

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

$$(g\circ f)'(x) = g'(f(x))\,f'(x) \quad \text{or}$$

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

Rmk We may write

$$g\circ f(x) = g(f(x)) = g(x) \quad \left(\text{Abuse of notation}\right)$$

$$\frac{dg}{dx} = \frac{dg}{du}\cdot\frac{du}{dx}$$

For multivariable functions:

Thm (Chain Rule)

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

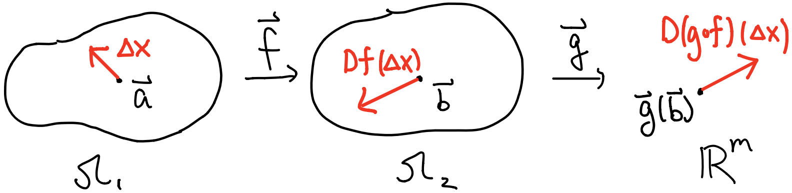
$\vec{g} : \Omega_2 \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$

Suppose that $\vec{f}$ is differentiable at $\vec{a}$

$\vec{g}$ is differentiable at $\vec{b} = \vec{f}(\vec{a})$

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

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



Rmk For simplicity, we may omit → for vectors from now on: $\vec{f} = f$, $\vec{x} = x$

eg $f : \mathbb{R} \to \mathbb{R}^2$, $g : \mathbb{R}^2 \to \mathbb{R}^2$

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

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

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

Sol **Method 1 : Find composition explicitly**

$$(g \circ f)(\theta) = g(\cos\theta, \sin\theta)$$

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

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

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

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,

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

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

$$= \begin{bmatrix} -2\sin^2\theta + 2\cos^2\theta \\ -4\cos\theta\sin\theta \end{bmatrix}$$

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

(same answer)

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

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

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

$\uparrow$

Rmk Regard g as a function of x, y

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

$Dg(f(1,1)) = Dg(1,3,2)$

$= \begin{bmatrix} \frac{2}{3} & -\frac{2}{9} & \frac{1}{3} \end{bmatrix}$

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

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

$\therefore D(g \circ f)(1,1)$

$= Dg(f(1,1))\, Df(1,1)$

$= \begin{bmatrix} \frac{2}{3} & -\frac{2}{9} & \frac{1}{3} \end{bmatrix} \begin{bmatrix} 2 & 0 \\ 3 & 3 \\ 1 & 2 \end{bmatrix}$

$= \begin{bmatrix} 1 & 0 \end{bmatrix}$

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

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

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

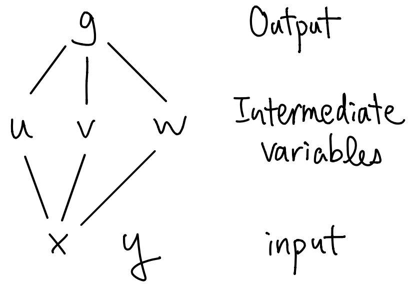
$$\begin{array}{c} \frac{\partial g}{\partial x} \\ \downarrow \end{array} \qquad\qquad\qquad\qquad \frac{\partial f_i}{\partial x}$$

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

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

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

Branch Diagram for $\frac{\partial g}{\partial x}$



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

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

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

Sol

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

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

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

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

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

eg John is walking with position at time t

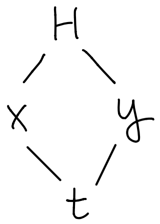
given by $\begin{cases} x(t) = t^3 + 1 \\ y(t) = 2t^2 \end{cases}$

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?

Sol ① Find $\left.\dfrac{dH}{dt}\right|_{t=1}$



$$\frac{dH}{dt} = \frac{\partial H}{\partial x}\cdot\frac{dx}{dt} + \frac{\partial H}{\partial y}\cdot\frac{dy}{dt}$$

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

$$= 2(t^3+1)(3t^2) - 2(2t^2)(4t)$$

$$= 6t^5 - 16t^3 + 6t^2$$

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

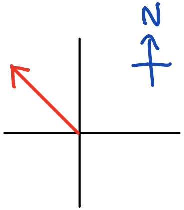
$\therefore$ John is going down at $t=1$

② At $t=1$, $(x,y) = (2,2)$

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

$$\nabla H(2,2) = (4, -4)$$

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



$\therefore$ John should go NW

Rmk slope in x- and y- direction gradient

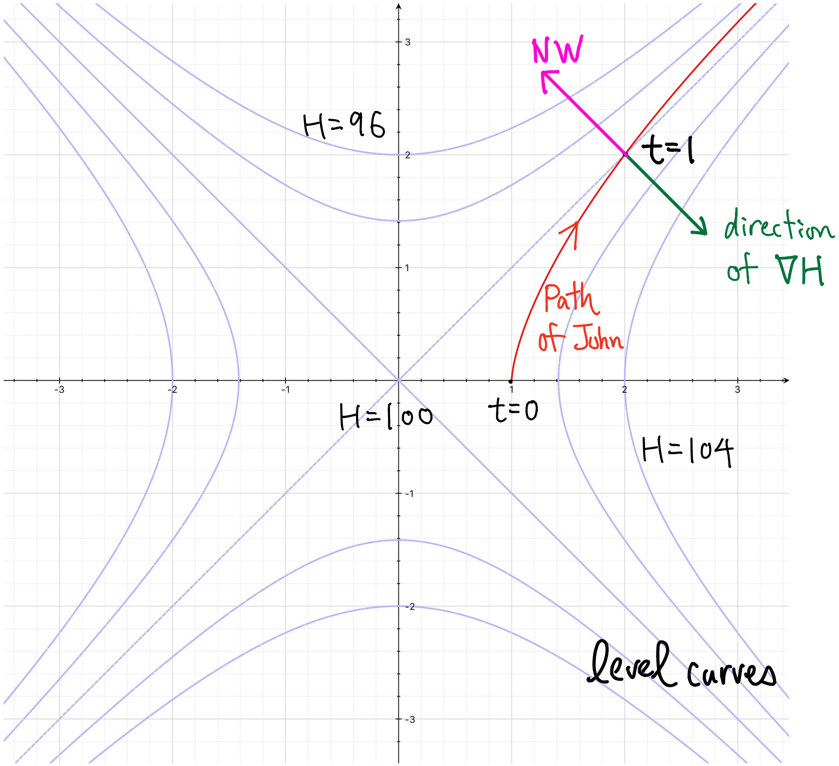
$$\frac{dH}{dt} = \frac{\partial H}{\partial x}\cdot\frac{dx}{dt} + \frac{\partial H}{\partial y}\cdot\frac{dy}{dt} = \nabla H \cdot \left[\frac{dx}{dt} \;\; \frac{dy}{dt}\right]$$

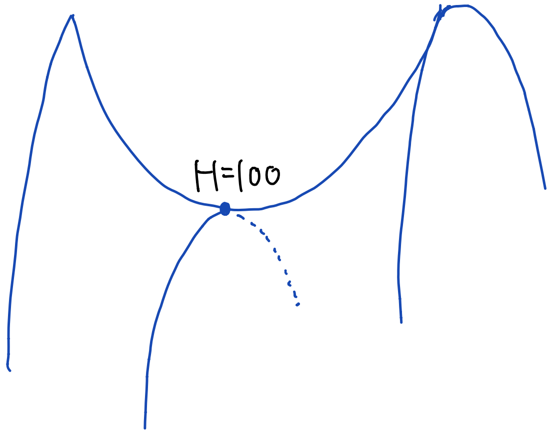
velocity in x- and y-direction velocity

Picture

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

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





Graph of $z = H(x,y)$

Idea of Pf of Chain Rule

Suppose $f : \Omega_1 \subseteq \mathbb{R}^k \longrightarrow \mathbb{R}^n$, differentiable at a

$g : \Omega_2 \subseteq \mathbb{R}^n \longrightarrow \mathbb{R}^m$, differentiable at $b = f(a) \in \Omega_2$

For $x \in \Omega_1$, $f(x) - f(a) = Df(a)(x-a) + \varepsilon_f(x)$ ---- (i)

$y \in \Omega_2$, $g(y) - g(b) = Dg(b)(y-b) + \varepsilon_g(y)$ --- (ii)

Put $y = f(x)$, $b = f(a)$ and (i) into (ii):

$$g(f(x)) - g(f(a)) = Dg(f(a))\left[Df(a)(x-a) + \varepsilon_f(x)\right] + \varepsilon_g(f(x))$$

$$= \underbrace{Dg(f(a))Df(a)(x-a)}_{\text{linear in } x-a} + \underbrace{Dg(f(a))\,\varepsilon_f(x) + \varepsilon_g(f(x))}_{\text{Let it be } \varepsilon_{g\circ f}(x)}$$

Math 2050 for details when $m=n=k=1$

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

(f continuous at $a \Rightarrow \|f(x)-f(a)\|$ small $\Rightarrow \varepsilon_g(f(x))$ small

$\varepsilon_f(x)$ small $\Rightarrow Dgf((a))\,\varepsilon_f(x)$ small)

$\Rightarrow$ $g\circ f$ is differentiable at a with $D(g\circ f)(a) = Dg(f(a))Df(a)$

Summary: Jacobian Matrix

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

$$x \longmapsto f(x)$$

$$Df(x) = \frac{df}{dx}$$ (scalar, 1x1 matrix)

② $f: \Omega \subseteq \mathbb{R}^n \to \mathbb{R}$ (multivariable, real-valued)

$$x = (x_1, \cdots, x_n) \longmapsto f(x) = f(x_1, \cdots, x_n)$$

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

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

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

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

Chain Rule (in classical notations)

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

$g_i = g_i(y_1, \cdots, y_n)$ 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 form:

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

i-th row, j-th column $\Rightarrow$

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