Lecture 10.5 (Midterm day)

Matrix calculus
(Taking derivative of matrix-valued scalar fns w.r.t. matrix)

$f : \mathbb{R}^{m\times n} \mapsto \mathbb{R}$
(often $m=n$)
e.g. $\det(\cdot)$, $\mathrm{tr}(\cdot)$

Given $f(X)$ compute $\frac{\partial f}{\partial X} \rightarrow$ a matrix

e.g. 1

$$\boxed{f(X) = \mathrm{tr}(X^2), \quad \frac{\partial f}{\partial X} = 2X^T}$$

$$D_Z f(X) = \lim_{h\to 0} \frac{\mathrm{tr}\left((X+hZ)^2\right) - \mathrm{tr}(X^2)}{h}$$

$$= \lim_{h\to 0} \frac{\mathrm{tr}\left[(X+hZ)(X+hZ)\right] - \mathrm{tr}(X^2)}{h}$$

$$= \lim_{h \to 0} \frac{\mathrm{tr}\left[X^2 + h(XZ + ZX) + h^2 Z^2\right] - \mathrm{tr}(X^2)}{h}$$

ignore

$$= \lim_{h \to 0} \frac{\cancel{\mathrm{tr}(X^2)} + 2\cancel{h}\,\mathrm{tr}(XZ) - \cancel{\mathrm{tr}(X^2)}}{\cancel{h}}$$

$$= 2\,\mathrm{tr}(XZ)$$

$$\Rightarrow \underbrace{D_Z f(X)}_{\shortparallel} = 2\,\mathrm{tr}(XZ)$$

$$\Rightarrow \mathrm{tr}\left(\left(\frac{\partial f}{\partial X}\right)^T Z\right) = \mathrm{tr}\left((2X^T)^T Z\right)$$

$$\boxed{\therefore \frac{\partial f}{\partial X} = 2X^T}, \quad f(X) = X^2$$

eg. 2 $f(X) = tr(X^T X)$, $\frac{\partial f}{\partial X} = 2X$

$$D_Z f(X) = \lim_{h \to 0} \frac{tr((X+hZ)^T(X+hZ)) - tr(X^T X)}{h}$$

$$= \lim_{h \to 0} \frac{tr(\cancel{X^T X} + h X^T Z + h Z^T X) - \cancel{tr(X^T X)}}{h}$$

$$= \lim_{h \to 0} \frac{\cancel{h}}{\cancel{h}} \boxed{tr(X^T Z + Z^T X)}$$

$$\Rightarrow tr\left(\left(\frac{\partial f}{\partial X}\right)^T Z\right) = 2\,tr(X^T Z)$$

$$\Rightarrow \boxed{\frac{\partial f}{\partial X} = 2X}$$

e.g. 3

$$f(X) = \operatorname{tr}(X^{-1}), \quad \frac{\partial f}{\partial X} = ? \quad \boxed{-(X^{-2})^T}$$

$$D_Z f(X) = \lim_{h \to 0} \frac{\operatorname{tr}((X+hZ)^{-1}) - \operatorname{tr}(X^{-1})}{h}$$

$$= \lim_{h \to 0} \frac{\operatorname{tr}((I+hX^{-1}Z)^{-1} X^{-1}) - \operatorname{tr}(X^{-1})}{h}$$

$$= \lim_{h \to 0} \frac{\operatorname{tr}[(I-hX^{-1}Z) X^{-1}] - \operatorname{tr}(X^{-1})}{h}$$

$$= \lim_{h \to 0} \frac{\cancel{\operatorname{tr}(X^{-1})} - \cancel{h}\operatorname{tr}(X^{-1}ZX^{-1}) - \cancel{\operatorname{tr}(X^{-1})}}{\cancel{h}}$$

$$\Rightarrow D_Z f(X) = -\operatorname{tr}(X^{-1} Z X^{-1})$$

$$\Rightarrow \operatorname{tr}\left(\left(\frac{\partial f}{\partial X}\right)^T Z\right) = -\operatorname{tr}(X^{-2} Z)$$

$$\Rightarrow \boxed{\frac{\partial f}{\partial X} = -\left(X^{-2}\right)^T}$$

---

eg. 4 $\quad f(X) = \det(X)$, $\quad \frac{\partial f}{\partial X} = ?$

$$D_Z f(X) = \lim_{h \to 0} \frac{\det(X + hZ) - \det(X)}{h}$$

$$= \lim_{h \to 0} \frac{\det(X(I + hX^{-1}Z)) - \det(X)}{h}$$

$$= \lim_{h \to 0} \frac{\det(X)\det(I + hX^{-1}Z) - \det(X)}{h}$$

Now,

$$\det(I + hY) = 1 + h\,\mathrm{tr}(Y) + O(h^2)$$

(Jacobi's formula)

$$\therefore\ D_Z f(X) = \lim_{h \to 0} \frac{\det(X)\left[1 + h\,\mathrm{tr}(X^{-1}Z)\right] - \det(X)}{h}$$

$$\Rightarrow \mathrm{tr}\left(\left(\frac{\partial f}{\partial X}\right)^T Z\right) = \lim_{h \to 0} \frac{\cancel{h}\det(X)}{\cancel{h}}\,\mathrm{tr}(X^{-1}Z)$$

$$= \mathrm{tr}\left[\left(\det(X)\,X^{-T}\right)^T Z\right]$$

$$\Rightarrow \boxed{\frac{\partial f}{\partial X} = \det(X)\,X^{-T}}$$

Eg 5

$$\frac{\partial}{\partial X} \log\det(X^{-1})$$

appears in optimization problems as $\boxed{\text{minimize } \log\det(X^{-1})}$

Why minimize $\boxed{\log\det(X^{-1})}$

$$f(X) = \log\det(X^{-1})$$

$$= \log\left(\det(X)\right)^{-1}$$

$$= -\log\det(X)$$

$$\max \log\det(X)$$

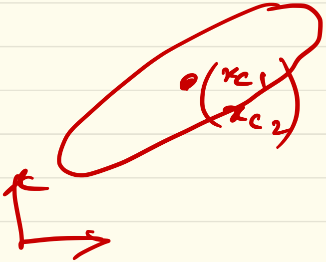
$$\Updownarrow$$

$$\max \det(X)$$

Ellipsoid:

$$\mathcal{E}(\underline{x}_c, Q) = \left\{ \underline{x} \in \mathbb{R}^n \;\middle|\; (\underline{x} - \underline{x}_c)^T Q^{-1} (\underline{x} - \underline{x}_c) \le 1 \right\}$$

$$Q \succ 0$$



$$\mathrm{vol}(\mathcal{E}(\underline{x}_c, Q)) = \frac{\text{constant}}{\sqrt{\det(Q^{-1})}} = \text{constant} \sqrt{\det(Q)}$$

$$\Leftrightarrow \mathrm{vol}(\cdot) \propto \det(Q)$$

Now, $f(X) = -\log\det(X)$

$$D_Z f(X) = \lim_{h\to 0} \frac{-\log\det(X+hZ) + \log\det(X)}{h}$$

$$= \lim_{h\to 0} \frac{-\log\{\det(X) + h\det(X)\,\mathrm{tr}(X^{-1}Z)\} + \log\det(X)}{h}$$

$$= \lim_{h\to 0} \frac{-\log\{\det(X)(1 + h\,\mathrm{tr}(X^{-1}Z))\} + \log\det(X)}{h}$$

$$= \lim_{h\to 0} \frac{\cancel{-\log\det(X)} - \log(1 + h\,\mathrm{tr}(X^{-1}Z)) \cancel{+ \log\det(X)}}{h}$$

Recall: $\underbrace{\log(1+y)}_{\parallel} \approx y + O(y^2)$

$$\int_0^y \underbrace{(1+t)^{-1}}_{\approx 1-t+t^2} dt = y - \frac{y^2}{2} + \frac{y^3}{3} - \frac{y^4}{4} + \ldots$$

$$\therefore \quad D_Z f(X) = \lim_{h \to 0} \frac{-\log(1 + h\,\mathrm{tr}(X^{-1}Z))}{h}$$

$$\Rightarrow \mathrm{tr}\left(\left(\frac{\partial f}{\partial X}\right)^T Z\right) = \lim_{h\to 0} - \frac{\cancel{h}\,\mathrm{tr}(X^{-1}Z)}{\cancel{h}} + O(h^2)$$

$$= -\mathrm{tr}(X^{-1}Z)$$

$$\Rightarrow \boxed{\frac{\partial f}{\partial X} = -X^{-T}}$$