5. Linear Algebra & Matrix Calculus.

① p-Norm of vector

p should be $[0, \infty)$

$$\|x\|_p = \left(\sum_i |x_i|^p\right)^{1/p}$$

E.g. $\|X\|_1 = \sum_i |x_i|$

$$\|x\|_2 = \sqrt{\sum_i x_i^2}$$

$$\|x\|_0 = \sum_i \mathbb{1}(x_i \neq 0)$$

$$\|x\|_\infty = \max(\{x_i\})$$

② p-Norm of matrix

$$\|A\|_p = \sup_{x \neq 0} \frac{\|Ax\|_p}{\|x\|_p}$$

E.g $\|A\|_1 = \max_j \sum_{i=1}^{m} |A_{ij}|$ maximum column sum

$$\|A\|_2 = \sigma_{max}(A)$$

$$\|A\|_\infty = \max_i \sum_j |A_{ij}| = \|A^T\|_1$$

$$\|A\|_F = \sqrt{\sum_{i,j} (A_{ij})^2} = \sqrt{\sum \sigma_i^2} \quad \text{Frobenius Norm}$$

Proof:

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|} = \max_{x \neq 0} \|A \frac{x}{\|x\|}\| = \max_{\|u\|=1} \|Au\|$$

$$= \max_{\|u\|_2=1} \|U\Sigma V^T u\| = \max_{\|u\|_2=1} \|\Sigma \hat{u}\| = \max_{\|u\|_2} \|\Sigma u\| = \sigma_{max}(A)$$

$$\|A\|_F = \left(\sum_{i,j} (A_{ij})^2\right)^{1/2} = \sqrt{Tr(AA^T)} = \sqrt{Tr(V\Sigma U^T U \Sigma V^T)} = \sqrt{Tr(\Sigma\Sigma)}$$
$$= \sqrt{\sum \sigma_i^2}$$

$$\|A\|_\infty = \max_{x \neq 0} \frac{\|Ax\|_\infty}{\|x\|_\infty} = \max_{x \neq 0} \|A \cdot \frac{x}{\|x\|_\infty}\|_\infty$$

Let $x' = x / \|x\|_\infty$

$\max \|Ax'\|_\infty \Rightarrow x' = \mathbb{1}$ and the result $\Rightarrow \max \sum row$.

$$\|A\|_1 = \max \frac{\|Ax\|_1}{\|x\|_1} \Rightarrow x^* = [0, 0, \cdots 1, 0, \cdots 0]^T$$

So it's obvious $\|A\|_1 = \|A^T\|_\infty$

③ High-dimentional Derivative.

* Dimention principles.

1) denorminator's dimension will be arranged first.

E.g. $$\frac{\partial A^{x\times Y\times Z}}{\partial B^{M\times N\times K}} = C^{\underbrace{M\times N\times K}\times\underbrace{x\times Y\times Z}}$$

1) Scalar by vector $\quad Z\in\mathbb{R},\ V\in\mathbb{R}^M$

$$\frac{\partial Z}{\partial V} = \begin{bmatrix} \frac{\partial Z}{\partial V_1} \\ \vdots \\ \frac{\partial Z}{\partial V_M} \end{bmatrix}_{M\times 1}$$

2) Vector by scalar $\quad Z\in\mathbb{R},\ V\in\mathbb{R}^M$

$$\frac{\partial V}{\partial Z} = \left[\frac{\partial V_1}{\partial Z}, \frac{\partial V_2}{\partial Z}, \cdots, \frac{\partial V_M}{\partial Z}\right]$$

3) Vector by Vcctor $\quad X\in\mathbb{R}^M,\ Y\in\mathbb{R}^N$

$$\frac{\partial Y}{\partial X} = \left[\frac{\partial Y_1}{\partial X}, \cdots, \frac{\partial Y_N}{\partial X}\right]$$

$$= \begin{bmatrix} \frac{\partial Y}{\partial X_1} \\ \vdots \\ \frac{\partial Y}{\partial X_M} \end{bmatrix} = \begin{bmatrix} \frac{\partial Y_1}{\partial X_1} & \cdots & \frac{\partial Y_N}{\partial X_1} \\ \vdots & \ddots & \vdots \\ \frac{\partial Y_1}{\partial X_M} & \cdots & \frac{\partial Y_N}{\partial X_M} \end{bmatrix}_{M\times N}$$

Corallary 1: $A \in \mathbb{R}^{k \times M}$, $V \in \mathbb{R}^{M \times 1}$ $X \in \mathbb{R}^{N \times 1}$

A is not a function of x

$$\left[\frac{\partial AV}{\partial x}\right]^{N \times k} = \frac{\partial V}{\partial x} \cdot A^T$$

Corallary 2: $U \in \mathbb{R}^{M \times 1}$ $V \in \mathbb{R}^{M \times 1}$ $x \in \mathbb{R}^{N \times 1}$

$$\left[\frac{\partial U^T V}{\partial x}\right] = \frac{\partial U}{\partial x} \cdot V + \frac{\partial V}{\partial x} \cdot U$$

Corallary 3: Chain Rule.

$$\frac{\partial}{\partial x} f(g(x)) = \frac{\partial g(x)}{\partial x} \cdot \frac{\partial f(g(x))}{\partial g(x)}$$

4) Scalar by Matrix $z \in \mathbb{R}$, $A \in \mathbb{R}^{k \times N}$

$$\frac{\partial z}{\partial A} = \left[\frac{\partial z}{\partial A_{ij}}\right]_{k \times N}$$

Corallary 4. $A \in \mathbb{R}^{M \times N}$, $U \in \mathbb{R}^{M \times 1}$, $V \in \mathbb{R}^{N \times 1}$,

U, V are not function of A

$$\frac{\partial U^T A V}{\partial A} = V U^T$$

5) Vector by Matrix $x \in \mathbb{R}^{k\times 1}$ $A \in \mathbb{R}^{M\times N}$

$$\left[\frac{\partial x}{\partial A}\right] = \left[\frac{\partial x_1}{\partial A}, \frac{\partial x_2}{\partial A}, \cdots, \frac{\partial x_k}{\partial A}\right]_{M\times N\times k} \qquad \left[\frac{\partial x_i}{\partial A}\right]_{M\times N\times 1}$$

$$= \begin{bmatrix} \frac{\partial x}{\partial A(1,:)} \\ \frac{\partial x}{\partial A(2,:)} \\ \vdots \\ \frac{\partial x}{\partial A(M,:)} \end{bmatrix}_{M\times N\times k} \qquad \left[\frac{\partial x}{\partial A(1,:)}\right]_{1\times N\times k}$$

④ Tensor Multiplication

1) $A \in \mathbb{R}^{M\times N}$ $X \in \mathbb{R}^{k\times 1}$ $Y \in \mathbb{R}^{k\times 1}$

$$\left[\frac{\partial x}{\partial A}\cdot Y\right]_{M\times N} = \left[\frac{\partial x_1}{\partial A}, \frac{\partial x_2}{\partial A}, \cdots, \frac{\partial x_k}{\partial A}\right] \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_k \end{bmatrix}$$

$$= \sum_{i=1}^{k} \frac{\partial x_i}{\partial A}\cdot y_i$$

2) $A \in \mathbb{R}^{M\times N}$, $x \in \mathbb{R}^{k\times 1}$ $B \in \mathbb{R}^{k\times D}$

$$\left[\frac{\partial x}{\partial A} B\right] = \left[\frac{\partial x_1}{\partial A}, \cdots, \frac{\partial x_k}{\partial A}\right] \cdot \begin{bmatrix} B(1,:) \\ B(2,:) \\ \vdots \\ B(k,:) \end{bmatrix} = \sum_{i=1}^{k} \frac{\partial x_i}{\partial A}\cdot B(i,:)$$

$$= \left[\frac{\partial x_1}{\partial A}, \cdots \frac{\partial x_k}{\partial A}\right]\cdot\left[B_1, \cdots, B_D\right]$$

Note that $\frac{\partial x_i}{\partial A} B_i$ will construct i^{th} channel of final tensor.