Ex 4 Given : S is a square matrix

To find : $\dfrac{\partial\, tr(S)}{\partial S}$

1) Inferring shape of $\dfrac{\partial\, tr(S)}{\partial S}$.

$tr(S)$ would be a scalar value as it is sum of elements across main diagonal

$\dfrac{\partial\, tr(S)}{\partial S}$ would also be $n \times n$ matrix if S is a $n \times n$ matrix.

If $S = \begin{bmatrix} s_{11} & & \cdots & s_{1n} \\ \vdots & s_{22} & & \\ & & s_{33} & \\ \vdots & & & \ddots \\ s_{n1} & & & s_{nn} \end{bmatrix}$

then $\dfrac{\partial\, tr(S)}{\partial S} = \begin{bmatrix} \dfrac{\partial\, tr(S)}{\partial s_{11}} & \cdots & \dfrac{\partial\, tr(S)}{\partial s_{1n}} \\ \vdots & \dfrac{\partial\, tr(S)}{\partial s_{22}} & \vdots \\ \vdots & \cdots & \dfrac{\partial\, tr(S)}{\partial s_{nn}} \end{bmatrix}$

so $tr(S) = s_{11} + s_{22} + \ldots + s_{nn} = \sum_{i=1}^{n} s_{ii}$

2 cases

case 1 : derivative of $tr(S)$ wrt nondg. element s_{kj} where $k \neq j$

$$\frac{\partial\, tr(S)}{\partial s_{lj}} = \frac{\partial \sum_{i=1}^{N} s_{ii}}{\partial s_{lj}} = 0$$ (tr(S) has no non-diagonal term)

---

case 2: derivative of tr(S) wrt diagonal term

$$\frac{\partial\, tr(S)}{\partial s_{ll}} = \frac{\partial \sum_{i=1}^{N} s_{ii}}{\partial s_{ll}} = 1$$

so

$$\frac{\partial\, tr(S)}{\partial S} = \begin{bmatrix} \frac{\partial tr(S)}{\partial s_{11}} & \cdots & \cdots & \frac{\partial tr(S)}{\partial s_{1n}} \\ \vdots & \frac{\partial tr(S)}{\partial s_{22}} & & \vdots \\ \vdots & & \ddots & \\ \vdots & \cdots & \cdots & \frac{\partial tr(S)}{\partial s_{nn}} \end{bmatrix},$$

would look like

$$\frac{\partial\, tr(S)}{\partial S} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & & & \\ 0 & & 1 & & \\ & & & \ddots & \\ & & & & 1 \end{bmatrix}_{N \times N}$$

$$\therefore \boxed{\frac{\partial\, tr(S)}{\partial S} = I}$$

Ex3 $w \in \mathbb{R}^n$, Euclidean norm $\|w\| := \sqrt{w^T w}$

Calculate : $\frac{\partial \|w\|}{\partial w}$

1) Inferring shape of $\frac{\partial \|w\|}{\partial w}$. $\|w\|$ is a scalar value and w is a column vector so,

$$\frac{\partial \|w\|}{\partial w} = \begin{bmatrix} \frac{\partial \|w\|}{\partial w_1} \\ \frac{\partial \|w\|}{\partial w_2} \\ \vdots \\ \frac{\partial \|w\|}{\partial w_n} \end{bmatrix}$$

$$\underset{(1\times N)}{w^T} \underset{(N\times 1)}{w} = [w_1 \cdots w_N] \begin{bmatrix} w_1 \\ \vdots \\ w_N \end{bmatrix} = \sum_{i=1}^{N} w_i \cdot w_i$$

Indexing wrt to a w_l,

$$\frac{\partial \|w\|}{\partial w_l} = \frac{\partial \sqrt{w^T w}}{\partial w_l} = \frac{\partial \sqrt{\sum_{i=1}^{N} w_i w_i}}{\partial w_l}$$

$$= \frac{1}{2\sqrt{\sum_{i=1}^{N} w_i w_i}} \frac{\partial \left(\sum_{i=1}^{N} w_i w_i\right)}{\partial w_l} \quad \text{(chain rule)}$$

Now using product rule $\frac{\partial uv}{\partial x} = u\frac{\partial v}{\partial x} + v\frac{\partial u}{\partial x}$

$$= \frac{1}{2\sqrt{\sum_{i=1}^{N} w_i w_i}} \left[\sum_{i=1}^{N} w_i \cdot \frac{\partial w_i}{\partial w_l} + \sum_{i=1}^{N} \frac{\partial w_i}{\partial w_l} \cdot w_i\right]$$

$$= \frac{1}{2\sqrt{w^T w}} \left[\sum_{i=1}^{N} w_i \cdot \delta_{il} + \sum_{i=1}^{N} \delta_{il} w_i\right]$$

$$= \frac{1}{2\|w\|} \left[2 \sum_{i=1}^{N} w_i \delta_{il}\right] = \frac{w_l}{\|w\|}$$

So $$\frac{\partial \|w\|}{\partial w_l} = \frac{w_l}{\|w\|}$$

Extrapolating it back into vector form,

$$\frac{\partial \|w\|}{\partial w} = \begin{bmatrix} \frac{w_1}{\|w\|} & \cdot \frac{w_2}{\|w\|} & \cdots & \frac{w_l}{\|w\|} & \cdots & \frac{w_n}{\|w\|} \end{bmatrix}_{N\times 1}$$

$$= \frac{1}{\|w\|} \begin{bmatrix} w_1 & \cdot\cdot\, w_2 & \cdot\cdot & w_l & \cdot\cdot & w_n \end{bmatrix}$$

$$\therefore \boxed{\frac{\partial \|w\|}{\partial w} = \frac{w}{\|w\|}}$$

Exercise 2 Given matrices V & W

Find $$\frac{\partial\, tr(VXW)}{\partial X}$$

(1) Inferring shape of (VXW) term, say

X is a general matrix of shape $m \times n$,

$X \subseteq R^{m\times N}$,

Trace only exists for a square matrix,

so matrix mult VXW should be a square matrix so,

$$V_{l \times m} \, X_{m \times n} \, W_{n \times l}$$

(we chose the shapes in such a way that
① matrix mul of VXW is a sq. matrix
② matrix mult happens)

Lets find out how the trace term would look like

$$\begin{bmatrix} v_{11} & \cdots & v_{1m} \\ v_{21} & & v_{2m} \\ \vdots & \ddots & \vdots \\ v_{m1} & \cdots & v_{lm} \end{bmatrix} \times \begin{bmatrix} x_{11} & \cdots & x_{1n} \\ x_{21} & \ddots & \\ \vdots & & \\ x_{m1} & & x_{mn} \end{bmatrix} \times W$$

$$\begin{bmatrix} \boxed{\sum_{i=1}^{m} v_{1i} x_{i1} \quad \sum_{i=1}^{m} v_{1i} x_{i2} \quad \cdots} \\ \boxed{\quad \sum_{i=1}^{m} v_{2i} x_{i2} \quad \cdots} \\ \vdots \quad \ddots \end{bmatrix} \begin{bmatrix} w_{11} & w_{12} & \cdots & w_{1l} \\ \vdots & \vdots & & \\ w_{n1} & w_{n2} & & w_{n} \end{bmatrix}$$

we have now VX multiplied, lets now just focus on diagonal terms after multiplying VX with W,

$$\begin{bmatrix} \boxed{\sum_{j=1}^{n} w_{j1} \left(\sum_{i=1}^{m} v_{1i} x_{ij} \right)} & & \\ & \boxed{\sum_{j=1}^{n} w_{j2} \left[\sum_{i=1}^{m} v_{2i} x_{ij} \right]} & \\ & & \ddots \end{bmatrix}$$

(so & so forth

The tr now would be summed across l terms,

one term $\Rightarrow$ $$\sum_{j=1}^{n} w_{jl} \left(\sum_{i=1}^{m} v_{li} x_{ij} \right)$$

many terms $\Rightarrow$ $$\sum_{k=1}^{l} \left[\sum_{j=1}^{n} w_{jk} \left(\sum_{i=1}^{m} v_{ki} x_{ij} \right) \right]$$

shape of $\frac{\partial \operatorname{tr}(VXW)}{\partial (X)}$ would be $m \times n$ matrix

$$\begin{bmatrix} \frac{\operatorname{tr}(VXW)}{\partial x_{11}} & \cdots & \frac{\partial \operatorname{tr}(VXW)}{\partial x_{1n}} \\ \vdots & \ddots & \\ & \cdots & \frac{\partial \operatorname{tr}(VXW)}{\partial x_{mn}} \end{bmatrix}$$

so deriving wrt one term,

$$\frac{\partial \operatorname{tr}(VXW)}{\partial x_{pq}} = \frac{\partial \sum_{k=1}^{l} \left[\sum_{j=1}^{n} w_{jk} \left(\sum_{i=1}^{m} v_{ki} x_{ij} \right) \right]}{\partial x_{pq}}$$

$$\sum_{k=1}^{l}\left[\sum_{j=1}^{n} w_{jk}\Big(\underline{\sum_{i=1}^{m} v_{ki}\,\delta_{pi}\,\delta_{qj}}\Big)\right]$$

simplifying this term when $p=i$ first,

$$\sum_{k=1}^{l}\left[\underline{\sum_{j=1}^{n} w_{jk}\, v_{kp}\, \delta_{qj}}\right]$$

simplifying this term when $q=j$

$$\sum_{k=1}^{l}\left[w_{qk}\, v_{kp}\right]$$

so

$$\boxed{\frac{\partial\, tr(VXW)}{\partial x_{pq}} = \sum_{k=1}^{l} w_{qk}\, v_{kp}} = \underline{\sum_{k=1}^{l} v_{kp}\, w_{qk}}$$

$$\begin{bmatrix} \sum^{l} v_{k1} w_{1k} & \cdots & \sum^{l} v_{k1} w_{nk} \\ \vdots & & \\ \sum^{l} v_{km} w_{1k} & \cdots & \sum^{l} v_{km} w_{nk} \end{bmatrix} = \begin{bmatrix} \frac{\partial\, tr(VXW)}{\partial x_{11}} & \cdots & \frac{\partial\, tr}{\partial x_{1n}} \\ \vdots & \ddots & \vdots \\ \frac{\partial\, tr}{\partial x_{m1}} & & \frac{\partial\, tr}{\partial x_{mn}} \end{bmatrix}$$

$$\underset{m\times l}{\begin{bmatrix} v_{11} & v_{21} & \cdots & v_{l1} \\ v_{12} & & & \\ \vdots & & & \\ v_{1m} & & & \end{bmatrix}} \times \underset{l\times n}{\begin{bmatrix} w_{11} & w_{21} & \cdots & w_{l1} \\ \vdots & & & \\ w_{1n} & & & \end{bmatrix}}$$

Made with Goodnotes

$= V^T W^T$ (seems a bit hard but try to trace back by inferring shapes)

Exercise 1 $x \in \mathbb{R}^n$ evaluate $\dfrac{\partial x^T B x}{\partial x}$

$B \in \mathbb{R}^{n \times n}$

Let $s = \underset{1\times n}{x^T} \underset{n\times n}{B} \underset{n \times 1}{x} \Rightarrow 1 \times 1$ scalar value

shape of $\dfrac{\partial s}{\partial x} = \left[\dfrac{\partial s}{\partial x_1} \quad \dfrac{\partial s}{\partial x_2} \quad \cdots \quad \dfrac{\partial s}{\partial x_n} \right]$

$\Downarrow$

same shape as x

Lets find value of s,

$$s = \begin{bmatrix} x_1 & \cdots & x_n \end{bmatrix}_{1\times N} \begin{bmatrix} B_{11} & \cdots & B_{1n} \\ \vdots & \ddots & \\ B_{n1} & & B_{nn} \end{bmatrix}_{n\times n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} x_i B_{i1} & \sum_{i=1}^{n} x_i B_{i2} & \cdots & \sum_{i=1}^{n} x_i B_{in} \end{bmatrix}_{1\times n} \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}_{n\times 1}$$

$$= \sum_{j=1}^{n} x_j \cdot \left(\sum_{i=1}^{n} x_i B_{ij} \right)$$

Taking x_j inside,

$$= \sum_{j=1}^{n} \sum_{i=1}^{n} x_i x_j B_{ij} \Leftarrow s.$$

Derivating wrt one term,

$$\frac{\partial s}{\partial x_l} = \frac{\partial \sum_{j=1}^{n} \sum_{i=1}^{n} x_i x_j B_{ij}}{\partial x_l}$$

using product rule,

$$\frac{\partial s}{\partial x_l} = \sum_{j=1}^{n} \sum_{i=1}^{n} \left(\frac{\partial x_i}{\partial x_l}\right) \cdot x_j B_{ij} + \sum_{j=1}^{n} \sum_{i=1}^{n} x_i \frac{\partial x_j}{\partial x_l} B_{ij}$$

$$\therefore \frac{\partial s}{\partial x_l} = \sum_{j=1}^{n} \sum_{i=1}^{n} \delta_{il} \cdot x_j B_{ij} + \sum_{j=1}^{n} \sum_{i=1}^{n} x_i \delta_{jl} B_{ij}$$

$$\therefore \boxed{\frac{\partial s}{\partial x_l} = \sum_{j=1}^{n} x_j B_{lj} + \sum_{i=1}^{n} x_i B_{il}}$$

$$\frac{\partial s}{\partial x} = \begin{bmatrix} \sum_{j=1}^{n} x_j \underline{B_{1j}} + \sum_{i=1}^{n} x_i \underline{B_{i1}} \\ \vdots \\ \sum_{j=1}^{n} x_j \underline{B_{nj}} + \sum_{i=1}^{n} x_i \underline{B_{in}} \end{bmatrix}$$

(observe the underlined terms; a term & its transposed term gets added)

$$= B \cdot x + B^T \cdot x = (B + B^T)x$$