Derivatives of matrices, vectors and scalar forms.

(You can find more details and properties in "matrix cookbook").

Notations: $\beta = [\beta_1, \beta_2, \dots, \beta_n]^T$ be an $n\times 1$ column vector of variables,

$a = [a_1, a_2, \dots, a_n]^T$ be an $n\times 1$ constant column vector.

A be an $m\times n$ constant matrix ; Y be an $n\times n$ matrix of variables.

Examples: Please compute the following first-order derivative $\frac{\partial f(\beta)}{\partial \beta}$:

1. Let $f(\beta) = a^T\beta$, so that $f(\beta): \mathbb{R}^n \longmapsto \mathbb{R}^1$.
2. Let $f(\beta) = \beta^T a$, so that $f(\beta): \mathbb{R}^n \longmapsto \mathbb{R}^1$.
3. Let $f(\beta) = A\beta$. so that $f(\beta): \mathbb{R}^n \to \mathbb{R}^m$.
4. Let $f(\beta) = \beta^T A^T$. so that $f(\beta): \mathbb{R}^n \longmapsto \mathbb{R}^{1\times m}$.
5. Let $f(\beta) = \beta^T(A^TA)\beta$, so that $f(\beta): \mathbb{R}^n \longmapsto \mathbb{R}^1$.
6. Please compute $\frac{\partial f(Y)}{\partial Y}$, where $f(Y) = tr(Yaa^T)$, so that $f(Y): \mathbb{R}^{n\times n} \longmapsto \mathbb{R}^1$

Solution 1:

$$f(\beta) = a^T\beta = [a_1, a_2, \dots, a_n]\begin{bmatrix}\beta_1\\ \beta_2\\ \vdots\\ \beta_n\end{bmatrix} = \sum_{i=1}^n a_i\beta_i,$$

By definition, we have

$$\underbrace{\frac{\partial f(\beta)}{\partial \beta}}_{\frac{\partial\, scalar}{\partial\, vector}} = \begin{bmatrix}\frac{\partial f(\beta)}{\partial \beta_1}\\ \frac{\partial f(\beta)}{\partial \beta_2}\\ \vdots\\ \frac{\partial f(\beta)}{\partial \beta_n}\end{bmatrix} = \begin{bmatrix}\frac{\partial \sum_{i=1}^n a_i\beta_i}{\partial \beta_1}\\ \frac{\partial \sum_{i=1}^n a_i\beta_i}{\partial \beta_2}\\ \vdots\\ \frac{\partial \sum_{i=1}^n a_i\beta_i}{\partial \beta_n}\end{bmatrix} = \begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_n\end{bmatrix} = a$$

Solution 2:

$f(\beta) = \beta^T a = a^T\beta = \sum_{i=1}^n a_i\beta_i$, similar we can derive:

$$\underbrace{\frac{\partial f(\beta)}{\partial \beta}}_{\frac{\partial\, scalar}{\partial\, vector}} = \begin{bmatrix}\frac{\partial f(\beta)}{\partial \beta_1}\\ \frac{\partial f(\beta)}{\partial \beta_2}\\ \vdots\\ \frac{\partial f(\beta)}{\partial \beta_n}\end{bmatrix} = \begin{bmatrix}\frac{\partial \sum_{i=1}^n a_i\beta_i}{\partial \beta_1}\\ \frac{\partial \sum_{i=1}^n a_i\beta_i}{\partial \beta_2}\\ \vdots\\ \frac{\partial \sum_{i=1}^n a_i\beta_i}{\partial \beta_n}\end{bmatrix} = \begin{bmatrix}a_1\\ a_2\\ \vdots\\ a_n\end{bmatrix} = a$$

Solution 3:

$$f(\beta) = A\beta = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m\times n} \cdot \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}_{n\times 1} = \Big[\underbrace{\sum_{j=1}^{n} a_{1j}\beta_j}_{f_1(\beta)}, \underbrace{\sum_{j=1}^{n} a_{2j}\beta_j}_{f_2(\beta)}, \cdots, \underbrace{\sum_{j=1}^{n} a_{mj}\beta_j}_{f_m(\beta)} \Big]^T_{m\times 1}$$

By definition, we have

$$\underbrace{\frac{\partial f(\beta)}{\partial \beta}}_{\frac{\partial \text{ vector}}{\partial \text{ vector}}} = \begin{bmatrix} \frac{\partial f_1(\beta)}{\partial \beta_1} & \frac{\partial f_2(\beta)}{\partial \beta_1} & \cdots & \frac{\partial f_m(\beta)}{\partial \beta_1} \\ \frac{\partial f_1(\beta)}{\partial \beta_2} & \frac{\partial f_2(\beta)}{\partial \beta_2} & \cdots & \frac{\partial f_m(\beta)}{\partial \beta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \underbrace{\frac{\partial f_1(\beta)}{\partial \beta_n}}_{\frac{\partial f_1(\beta)}{\partial \cdot\cdot}} & \underbrace{\frac{\partial f_2(\beta)}{\partial \beta_n}}_{\frac{\partial f_2(\beta)}{\partial \cdot\cdot}} & \cdots & \underbrace{\frac{\partial f_m(\beta)}{\partial \beta_n}}_{\frac{\partial f_m(\beta)}{\partial \cdot\cdot}} \end{bmatrix}_{m\times n}$$

$$= \begin{bmatrix} \frac{\partial \sum_{j=1}^{n} a_{1j}\beta_j}{\partial \beta_1} & \frac{\partial \sum_{j=1}^{n} a_{2j}\beta_j}{\partial \beta_1} & \cdots & \frac{\partial \sum_{j=1}^{n} a_{mj}\beta_j}{\partial \beta_1} \\ \frac{\partial \sum_{j=1}^{n} a_{1j}\beta_j}{\partial \beta_2} & \frac{\partial \sum_{j=1}^{n} a_{2j}\beta_j}{\partial \beta_2} & \cdots & \frac{\partial \sum_{j=1}^{n} a_{mj}\beta_j}{\partial \beta_2} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial \sum_{j=1}^{n} a_{1j}\beta_j}{\partial \beta_n} & \frac{\partial \sum_{j=1}^{n} a_{2j}\beta_j}{\partial \beta_n} & \cdots & \frac{\partial \sum_{j=1}^{n} a_{mj}\beta_j}{\partial \beta_n} \end{bmatrix}_{m\times n} = \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & \cdots & a_{m2} \\ \vdots & \vdots & \ddots & \vdots \\ a_{1n} & a_{2n} & \cdots & a_{mn} \end{bmatrix}_{m\times n} = \boxed{A^T}$$

Solution 4:

$$f(\beta) = \beta^T A^T = [\beta_1, \beta_2, \cdots, \beta_n]_{1\times n} \begin{bmatrix} a_{11} & a_{21} & \cdots & a_{m1} \\ a_{12} & a_{22} & & a_{m2} \\ \vdots & \vdots & & \vdots \\ a_{1n} & a_{2n} & & a_{mn} \end{bmatrix}_{n\times m} = \Big[\underbrace{\sum_{j=1}^{n} a_{1j}\beta_j}_{f_1(\beta)}, \underbrace{\sum_{j=1}^{n} a_{2j}\beta_j}_{f_2(\beta)} \cdots \underbrace{\sum_{j=1}^{n} a_{mj}\beta_j}_{f_m(\beta)} \Big]_{1\times m}$$

Since $\dfrac{\partial f^T(\beta)}{\partial \beta} = \dfrac{\partial (\beta^T A^T)^T}{\partial \beta} = \dfrac{\partial A\beta}{\partial \beta} = A^T$ (from solution 3).

So we have $\dfrac{\partial f(\beta)}{\partial \beta} = \left(\dfrac{\partial f^T(\beta)}{\partial \beta}\right)^T = (A^T)^T = \boxed{A}$

Solution 5:

Using the formula $\dfrac{d h^T(x) g(x)}{dx} = \dfrac{d h(x)}{dx} g(x) + \dfrac{d g(x)}{dx} h(x)$,

Let $h^T(x) = h^T(\beta) = \beta^T A^T$; $g(x) = g(\beta) = A\beta$

then we have
$$\frac{d f(\beta)}{d\beta} = \frac{d h^T(\beta) g(\beta)}{d\beta} = \frac{d h(\beta)}{d\beta} g(\beta) + \frac{d g(\beta)}{d\beta} h(\beta)$$
$$= A^T A\beta + A^T A\beta$$
$$= 2A^T A\beta$$

Notice that $h(\beta) = g(\beta) = A\beta$ in this example

Exercise: please calculate $\dfrac{d f(\beta)}{d\beta}$, where $f(\beta) = \beta^T X \beta$, X is an $n \times n$ constant matrix.

Answer: $\dfrac{d(\beta^T X \beta)}{d\beta} = (X^T + X)\beta$

Solution 6:

$$f(Y) = tr(Yaa^T) = tr\begin{bmatrix} \sum_{i=1}^{n} Y_{1i} a_i a_1 & \cdots & & \\ \vdots & \sum_{i=1}^{n} Y_{2i} a_i a_2 & & \\ \vdots & & \ddots & \\ & & & \sum_{i=1}^{n} Y_{ni} a_i a_n \end{bmatrix}_{n\times n} = \sum_{j=1}^{n}\sum_{i=1}^{n} Y_{ji} a_i a_j$$

By definition:

$$\underbrace{\frac{d f(Y)}{dY}}_{\frac{d\,\text{scalar}}{d\,\text{matrix}}} = \begin{bmatrix} \frac{d f(Y)}{d Y_{11}} & \frac{d f(Y)}{d Y_{12}} & \cdots & \frac{d f(Y)}{d Y_{1n}} \\ \frac{d f(Y)}{d Y_{21}} & \frac{d f(Y)}{d Y_{22}} & \cdots & \frac{d f(Y)}{d Y_{2n}} \\ \vdots & & \ddots & \vdots \\ \frac{d f(Y)}{d Y_{n1}} & \frac{d f(Y)}{d Y_{n2}} & \cdots & \frac{d f(Y)}{d Y_{nn}} \end{bmatrix}_{n\times n} = \begin{bmatrix} a_1^2 & a_2 a_1 & \cdots & a_n a_1 \\ a_1 a_2 & a_2^2 & & a_n a_2 \\ \vdots & \vdots & \ddots & \vdots \\ a_1 a_n & a_2 a_n & \cdots & a_n^2 \end{bmatrix}_{n\times n} = aa^T$$

Exercise: please calculate $\dfrac{d f(Y)}{dY}$, where $f(Y) = tr(YX)$, X is an $n \times n$ constant matrix.

Answer: $\dfrac{d\, tr(YX)}{dX} = X^T$