Notation & Matrix Differentiation.

- $S(\theta) =$ 1×1 scaler
- $\theta = \begin{bmatrix} \theta_1 \\ \theta_2 \\ \vdots \\ \theta_k \end{bmatrix}$ $k\times 1$ vector , $\theta' = [\theta_1 \ \ \theta_2 \cdots \theta_k]$

$$\Rightarrow \frac{\partial S(\theta)}{\partial \theta} = \begin{bmatrix} \frac{\partial S(\theta)}{\partial \theta_1} \\ \vdots \\ \frac{\partial S(\theta)}{\partial \theta_k} \end{bmatrix}$$

$k\times 1$ vector
Gradient

$$\Rightarrow \frac{\partial S(\theta)}{\partial \theta'} = \left[\frac{\partial S(\theta)}{\partial \theta_1} \cdots \frac{\partial S(\theta)}{\partial \theta_k} \right]$$

$1\times k$ vector
Jaccobian

For scaler

$$\frac{\partial S(\theta)}{\partial \theta}' = \frac{\partial S(\theta)}{\partial \theta'}$$

※分母表示形状.

$$\Rightarrow \frac{\partial S(\theta)}{\partial \theta \partial \theta'} = \frac{\partial}{\partial \theta}\left(\frac{\partial S(\theta)}{\partial \theta'}\right) = \frac{\partial}{\partial \theta}\left[\frac{S(\theta)}{\partial \theta_1} \cdots \frac{\partial S(\theta)}{\partial \theta_k}\right] = \begin{bmatrix} \frac{\partial S(\theta)}{\partial \theta_1 \partial \theta_1} & \cdots & \frac{\partial S(\theta)}{\partial \theta_1 \partial \theta_k} \\ \vdots & & \vdots \\ \frac{\partial S(\theta)}{\partial \theta_k \partial \theta_1} & \cdots & \frac{\partial S(\theta)}{\partial \theta_k \partial \theta_k} \end{bmatrix}$$

$k\times k$ vector
Hessian

- $A(\theta) = \begin{bmatrix} A_1(\theta) \\ A_2(\theta) \\ \vdots \\ A_r(\theta) \end{bmatrix}$ $r\times 1$ matrix

$$\Rightarrow \frac{\partial A(\theta)}{\partial \theta'} = \left[\frac{\partial A(\theta)}{\partial \theta_1} \cdots \frac{\partial A(\theta)}{\partial \theta_k} \right] = \begin{bmatrix} \frac{\partial A_1(\theta)}{\partial \theta_1} & \cdots & \frac{\partial A_1(\theta)}{\partial \theta_k} \\ \vdots & & \vdots \\ \frac{\partial A_r(\theta)}{\partial \theta_1} & \cdots & \frac{\partial A_r(\theta)}{\partial \theta_k} \end{bmatrix}$$

$r\times k$ matrix
Jaccobian

※ : $\frac{\partial A(\theta)}{\partial \theta} \equiv \left(\frac{\partial A(\theta)}{\partial \theta'}\right)'$ $k\times r$ matrix

Jaccobian^T

内積

$$\frac{\partial}{\partial\theta}\Big(\underset{1\times k}{a^T}\underset{k\times 1}{\theta}\Big) = \frac{\partial}{\partial\theta}(a\cdot\theta) = \frac{\partial}{\partial\theta}(\theta\cdot a) = \frac{\partial}{\partial\theta}(\theta^T a) \qquad \Rightarrow \frac{\partial}{\partial\theta}(a^T\theta) = \frac{\partial}{\partial\theta}(\theta^T a) = a$$

$$= \frac{\partial}{\partial\theta}(a_1\theta_1 + \cdots a_k\theta_k) = a_1\frac{\partial\theta_1}{\partial\theta} + \cdots a_k\frac{\partial\theta_k}{\partial\theta}$$

$$= a_1\begin{bmatrix}1\\0\\0\\0\\ \vdots\end{bmatrix} + \cdots a_k\begin{bmatrix}0\\ \vdots\\ 1\end{bmatrix} = \begin{bmatrix}a_1\\ \vdots\\ a_k\end{bmatrix} = a$$

$$\underset{k\times 1}{\frac{\partial}{\partial\theta}}\Big(\overset{1\times r}{a^T}\overset{r\times 1}{F(\theta)}\Big) = \frac{\partial}{\partial\theta}\Big(\underset{1\times 1}{a_1}\underset{1\times 1}{F_1(\theta)} + \cdots \underset{1\times 1}{a_r}\underset{1\times 1}{F_r}\Big)$$

$$= \frac{\partial F_1(\theta)}{\partial\theta}a_1 + \cdots + \frac{\partial F_r}{\partial\theta}a_r$$

$$= k\begin{bmatrix}\frac{\partial F_1(\theta)}{\partial\theta_1}a_1\\ \vdots\\ \frac{\partial F_1(\theta)}{\partial\theta_k}a_1\end{bmatrix} + \cdots \begin{bmatrix}\frac{\partial F_r}{\partial\theta_1}a_r\\ \vdots\\ \frac{\partial F_r(\theta)}{\partial\theta_k}a_r\end{bmatrix}$$

$$= k\begin{bmatrix}\frac{\partial F_1(\theta)}{\partial\theta_1}a_1 + \cdots + \frac{\partial F_r(\theta)}{\partial\theta_1}a_r\\ \vdots \qquad\qquad \vdots\\ \frac{\partial F_1(\theta)}{\partial\theta_k}a_1 + \cdots + \frac{\partial F_r}{\partial\theta_k}a_r\end{bmatrix} = k\overset{\qquad 1 \qquad\qquad r}{\begin{bmatrix}\frac{\partial F_1(\theta)}{\partial\theta_1} & \cdots & \frac{\partial F_r(\theta)}{\partial\theta_1}\\ \vdots & & \vdots\\ \frac{\partial F_1(\theta)}{\partial\theta_k} & \cdots & \frac{\partial F_r}{\partial\theta_k}\end{bmatrix}}\begin{bmatrix}a_1\\ \vdots\\ a_r\end{bmatrix}$$

But $\frac{\partial F(\theta)}{\partial\theta'} = \text{Jacobian}(F) = \begin{bmatrix}\frac{\partial F_1}{\partial\theta_1} & \cdots & \frac{\partial F_1}{\partial\theta_k}\\ \vdots & & \\ \frac{\partial F_r}{\partial\theta_1} & \cdots & \frac{\partial F_r}{\partial\theta_k}\end{bmatrix}$ transpose

$$= \underset{k\times r}{\underset{r\times k}{\left(\frac{\partial F(\theta)}{\partial\theta'}\right)'}}\underset{r\times 1}{a} \overset{\text{denote}}{=} \underset{k\times r}{\frac{\partial F(\theta)}{\partial\theta}}\underset{r\times 1}{a}$$

$$\underset{k\times 1}{\frac{\partial}{\partial\theta}}\Big[\underset{1\times r}{F(\theta)'}\,\underset{r\times 1}{G(\theta)}\Big] = \frac{\partial}{\partial\theta}\Big[F(\theta)\cdot G(\theta)\Big]$$

$$= \frac{\partial}{\partial\theta}\Big[F_1(\theta)G_1(\theta) + \cdots F_r(\theta)G_r(\theta)\Big] = \frac{\partial}{\partial\theta}\sum_{i=1}^{r} F_i(\theta)G_i(\theta)$$

$$= \begin{bmatrix} \frac{\partial}{\partial\theta_1}\sum F_i(\theta)G_i(\theta) \\ \vdots \\ \frac{\partial}{\partial\theta_k}\sum F_i(\theta)G_i(\theta) \end{bmatrix}$$

$$= \begin{bmatrix} \sum F_i(\theta)\frac{\partial G_i(\theta)}{\partial\theta_1} + G_i(\theta)\frac{\partial F_i(\theta)}{\partial\theta_1} \\ \vdots \\ \sum F_i(\theta)\frac{\partial G_i(\theta)}{\partial\theta_k} + G_i(\theta)\frac{\partial F_i(\theta)}{\partial\theta_k} \end{bmatrix} = \begin{bmatrix} \sum F_i(\theta)\frac{\partial G_i(\theta)}{\partial\theta_1} \\ \vdots \\ \sum F_i(\theta)\frac{\partial G_i(\theta)}{\partial\theta_k} \end{bmatrix} + \begin{bmatrix} \sum G_i(\theta)\frac{\partial F_i(\theta)}{\partial\theta_1} \\ \vdots \\ \sum G_i(\theta)\frac{\partial F_i(\theta)}{\partial\theta_k} \end{bmatrix}$$

$$= \begin{bmatrix} F_1(\theta)\frac{\partial G_1(\theta)}{\partial\theta_1} + F_2(\theta)\frac{\partial G_2(\theta)}{\partial\theta_1} + \cdots & F_r(\theta)\frac{\partial G_r(\theta)}{\partial\theta_1} \\ \vdots & \\ F_1(\theta)\frac{\partial G_1(\theta)}{\partial\theta_k} + F_2(\theta)\frac{\partial G_2(\theta)}{\partial\theta_k} + \cdots & F_r(\theta)\frac{\partial G_r(\theta)}{\partial\theta_k} \end{bmatrix} + F \leftrightarrow G$$

$$= \underbrace{\begin{bmatrix} \frac{\partial G_1(\theta)}{\partial\theta_1} & \frac{\partial G_2(\theta)}{\partial\theta_1} & \cdots & \frac{\partial G_r(\theta)}{\partial\theta_1} \\ \vdots & & & \\ \frac{\partial G_1(\theta)}{\partial\theta_k} & \frac{\partial G_2(\theta)}{\partial\theta_k} & \cdots & \frac{\partial G_r(\theta)}{\partial\theta_k} \end{bmatrix}}_{\text{Jacobian transposed}} \begin{pmatrix} F_1(\theta) \\ F_2(\theta) \\ \vdots \\ F_r(\theta) \end{pmatrix} + F \leftrightarrow G$$

The Jacobian

$$\frac{\partial G\downarrow}{\partial\theta'\rightarrow} = \begin{bmatrix} \frac{\partial G_1}{\partial\theta_1} & \frac{\partial G_1}{\partial\theta_2} & \cdots & \frac{\partial G_1}{\partial\theta_k} \\ \vdots & & & \\ \frac{\partial G_r}{\partial\theta_1} & \cdots & & \frac{\partial G_r}{\partial\theta_k} \end{bmatrix}$$

$$= \left(\frac{\partial G(\theta)}{\partial\theta'}\right)' F(\theta) + \left(\frac{\partial F(\theta)}{\partial\theta'}\right)' G(\theta)$$

$$= \underset{k\times r}{\frac{\overset{1\times r}{\partial G(\theta)'}}{\underset{k\times 1}{\partial\theta}}}\ \underset{r\times 1}{F(\theta)} + \underset{k\times r}{\frac{\partial F(\theta)'}{\partial\theta}}\ \underset{r\times 1}{G(\theta)}$$