Vector and Matrix Differentiation Rules

I) Scalar numerator

We start by considering a scalar function or scalar field that take vectors $x \in \mathbb{R}^n$ as input. Then, define

partial derivatives along a column!

$$\frac{\partial \, f(x)_{1\times 1}}{\partial \, x_{n\times 1}} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} \\ \vdots \\ \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

As a special case we have $f(x) = a'x$. Therefore

$$\frac{\partial \, a'x}{\partial x} = \begin{bmatrix} \frac{\partial a'x}{\partial x_1} \\ \vdots \\ \frac{\partial a'x}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a.$$

And recall that $a'x = x'a$ because it's a scalar, so that

$$\frac{\partial \, x'a}{\partial x} = \begin{bmatrix} \frac{\partial x'a}{\partial x_1} \\ \vdots \\ \frac{\partial x'a}{\partial x_n} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = a.$$

Similarly,

partial derivatives along a row!

$$\frac{\partial \, f(x)_{1\times 1}}{\partial \, x'_{1\times n}} = \begin{bmatrix} \frac{\partial f(x)}{\partial x_1} & \cdots & \frac{\partial f(x)}{\partial x_n} \end{bmatrix}$$

$$= \left[\frac{\partial f(x)}{\partial x}\right]'$$

RULE 1: $$\frac{\partial f(x)_{1\times 1}}{\partial x'_{1\times k}} = \left[\frac{\partial f(x)_{1\times 1}}{\partial x_{k\times 1}}\right]'$$

II) Vector numerator

Let A be a $m \times n$ matrix, $A = \begin{pmatrix} a_1'{}_{1\times n} \\ \vdots \\ a_m'{}_{1\times n} \end{pmatrix}$ where $a_j \in \mathbb{R}^n$ for $j = 1, \dots, m$.

$$\frac{\partial A x_{m\times 1}}{\partial x'_{1\times n}} = \begin{pmatrix} \frac{\partial a_1' x}{\partial x'}_{1\times 1} \\ \vdots \\ \frac{\partial a_m' x}{\partial x'}_{1\times 1} \end{pmatrix} = \begin{pmatrix} a_1' \\ \vdots \\ a_m' \end{pmatrix} = A_{m\times n}$$

we know how to compute each one of these

$$\frac{\partial x' A'_{1\times m}}{\partial x_{n\times 1}} = \begin{pmatrix} \frac{\partial x' a_1}{\partial x} & \cdots & \frac{\partial x' a_m}{\partial x} \end{pmatrix}$$

$$= \begin{pmatrix} a_1 & \cdots & a_m \end{pmatrix} = A'_{n\times m}.$$

III) Chain Rules on I and II

Multivariate chain rule : let $f(x)$ and $x(\alpha)$, $\alpha \in \mathbb{R}$

$$\frac{\partial f(x)_{1\times 1}}{\partial \alpha_{1\times 1}} = \frac{\partial f(x)}{\partial x_1}\frac{\partial x_1}{\partial \alpha} + \dots + \frac{\partial f(x_n)}{\partial x_n}\frac{\partial x_n}{\alpha}$$

$$= \sum_{i=1}^{n} \frac{\partial f_{1\times 1}}{\partial x_{i\,1\times 1}} \frac{\partial x_{i\,1\times 1}}{\partial \alpha_{1\times 1}} = \left(\frac{\partial f_{1\times 1}}{\partial x_{n\times 1}}\right)' \frac{\partial x_{n\times 1}}{\partial \alpha_{1\times 1}}$$

(Reminder)

Let $\alpha \in \mathbb{R}^r$ and $x = x(\alpha)$. Then

$$\frac{\partial x_{n\times 1}}{\partial \alpha'_{1\times r}} = \begin{pmatrix} \frac{\partial x_{1\,1\times 1}}{\partial \alpha'_{1\times r}} \\ \vdots \\ \frac{\partial x_n}{\partial \alpha'} \end{pmatrix}_{n\times r} = \begin{pmatrix} \frac{\partial x}{\partial \alpha_1}_{n\times 1} & \cdots & \frac{\partial x}{\partial \alpha_r}_{n\times 1} \end{pmatrix}_{n\times r}$$

$$\underset{r\times 1}{\frac{\partial f(x)_{1\times 1}}{\partial \alpha}} = \begin{pmatrix} \frac{\partial f(x)}{\partial \alpha_1} \\ \vdots \\ \frac{\partial f(x)}{\partial \alpha_r} \end{pmatrix}$$

Use Multivariate chain rule

$$= \begin{pmatrix} \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_1} \\ \vdots \\ \sum_{i=1}^{n} \frac{\partial f}{\partial x_i} \frac{\partial x_i}{\partial \alpha_r} \end{pmatrix}$$

Recall

$$\begin{bmatrix} a_1' \\ \vdots \\ a_n' \end{bmatrix} \begin{bmatrix} b_1 \cdots b_k \end{bmatrix} = \begin{bmatrix} a_1' b_1 & \cdots & a_1' b_k \\ \vdots & & \\ a_n' b_1 & \cdots & a_n' b_k \end{bmatrix}$$

$$= \begin{pmatrix} \left(\frac{\partial f}{\partial x}\right)' \frac{\partial x}{\partial \alpha_1} \\ \vdots \\ \left(\frac{\partial f}{\partial x}\right)' \frac{\partial x}{\partial \alpha_r} \end{pmatrix}$$

We can transpose scalars !

$$= \begin{pmatrix} \frac{\partial x'_{1\times n}}{\partial \alpha_1} \frac{\partial f_{1\times 1}}{\partial x_{n\times 1}} \\ \vdots \\ \frac{\partial x'}{\partial \alpha_r} \frac{\partial f}{\partial x} \end{pmatrix} = \begin{pmatrix} \frac{\partial x'}{\partial \alpha_1} \\ \vdots \\ \frac{\partial x'}{\partial \alpha_r} \end{pmatrix} \frac{\partial f}{\partial x}$$

$$= \underbrace{\frac{\partial x'_{1\times n}}{\partial \alpha_{r\times 1}}}_{r\times n} \underbrace{\frac{\partial f_{1\times 1}}{\partial x_{n\times 1}}}_{n\times 1} .$$

$$\boxed{\underline{\text{Rule 2}} : \quad \underset{r\times 1}{\frac{\partial f(x)_{1\times 1}}{\partial \alpha}} = \underbrace{\frac{\partial x'}{\partial \alpha} \frac{\partial f}{\partial x}}_{r\times 1}}$$

Def :- Let $\underset{n\times K}{A} = \Big(\underset{n\times 1}{a_1} \cdots \underset{n\times 1}{a_K} \Big)_{n\times K}$. Then

$$\operatorname{vec}(A) := \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{pmatrix}.$$

Notice that

$$\begin{pmatrix} \frac{\partial x}{\partial \alpha_1}{}_{n\times 1} \\ \vdots \\ \frac{\partial x}{\partial \alpha_r}{}_{n\times 1} \end{pmatrix}_{nr\times 1} = \operatorname{vec}\left(\left[\frac{\partial x}{\partial \alpha_1} \;\cdots\; \frac{\partial x}{\partial \alpha_r} \right] \right) = \operatorname{vec}\left(\frac{\partial x}{\partial \alpha'} \right)$$

Hence

$$\begin{pmatrix} \left(\frac{\partial f}{\partial x}\right)' \frac{\partial x}{\partial \alpha_1} \\ \vdots \\ \left(\frac{\partial f}{\partial x}\right)' \frac{\partial x}{\partial \alpha_r} \end{pmatrix} = \begin{pmatrix} \frac{\partial f}{\partial x'}{}_{1\times n} & \cdots & 0_{1\times n} \\ & \ddots & \\ 0_{1\times n} & \cdots & \frac{\partial f}{\partial x'}{}_{1\times n} \end{pmatrix} \operatorname{vec}\left(\frac{\partial x}{\partial \alpha'}{}_{n\times r} \right)_{nr\times 1}$$

$$= \underbrace{\left(\underset{r\times r}{I_r} \otimes \frac{\partial f}{\partial x'}{}_{1\times n} \right)}_{r\times rn} \operatorname{vec}\left(\frac{\partial x}{\partial \alpha'} \right)_{rn\times 1}$$

$$\boxed{\textbf{Rule 3}: \quad \underset{r\times n}{M}\, \underset{n\times 1}{v} = \left(I_r \otimes v' \right) \operatorname{vec}(M')}$$

Why is this useful?

- $\underbrace{\frac{\partial}{\partial \alpha'}}_{1\times r} \left(\underbrace{\frac{\partial x}{\partial \alpha'}}_{n\times r} \right)$ is not well-defined.

- $\underbrace{\frac{\partial}{\partial \alpha'}}_{1\times r} \underbrace{vec\left(\frac{\partial x}{\partial \alpha'} \right)}_{nr\times 1}$ is well-defined.

This allows us to generalize to the derivative of a vector

- $\frac{\partial \; Ax_{m\times 1}}{\partial \alpha'_{1\times r}} = \begin{pmatrix} \frac{\partial (a_1' x)_{1\times 1}}{\partial \alpha'_{1\times r}} \\ \vdots \\ \frac{\partial (a_m' x)}{\partial \alpha'} \end{pmatrix} = \begin{pmatrix} a_1' \frac{\partial x}{\partial \alpha'} \\ \vdots \\ a_m' \frac{\partial x}{\partial \alpha'} \end{pmatrix}$

  we Transposed Rule 2

  $= \begin{pmatrix} a_1' \\ \vdots \\ a_m' \end{pmatrix} \frac{\partial x}{\partial \alpha'} = A_{m\times n} \frac{\partial x}{\partial \alpha'}_{n\times r}$

- $\frac{\partial \; x' A'_{1\times m}}{\partial \alpha_{r\times 1}} = \begin{pmatrix} \frac{\partial x' a_1}{\partial \alpha} & \cdots & \frac{\partial x' a_m}{\partial \alpha} \end{pmatrix}$

  $= \begin{pmatrix} \frac{\partial x'}{\partial \alpha} a_1 & \cdots & \frac{\partial x'}{\partial \alpha} a_m \end{pmatrix}$

  $= \frac{\partial x'}{\partial \alpha} \begin{pmatrix} a_1 & \cdots & a_m \end{pmatrix} = \frac{\partial x'}{\partial \alpha} A'$

Notice that

$$\frac{\partial x}{\partial x'} = \begin{pmatrix} \frac{\partial x_1}{\partial x'} \\ \vdots \\ \frac{\partial x_n}{\partial x'} \end{pmatrix} = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \vdots & & \\ 0 & 0 & \cdots & 1 \end{pmatrix} = I.$$

Rule 4: $$\frac{\partial \underset{m\times 1}{Ax}}{\partial \underset{1\times r}{\alpha'}} = A \frac{\partial x}{\partial \alpha'}$$

IV) Chain Rule on Quadratic Forms

Let $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$ and A is $m\times n$. Also let $x := x(\alpha)$, $z := z(\alpha)$. Then

$$\frac{\partial (z'Ax)}{\partial \underset{r\times 1}{\alpha}} = \frac{\partial}{\partial \alpha}(z' \underbrace{Ax}_{\substack{\text{fixed} \\ c}}) + \frac{\partial}{\partial \alpha}(\underbrace{z'A}_{\substack{\text{fixed} \\ d'}} x)$$

$$\underset{\text{Rule 2}}{=} \frac{\partial z'}{\partial \alpha} \frac{\partial z'c}{\partial z} + \frac{\partial x'}{\partial \alpha} \frac{\partial d'x}{\partial x}$$

$$= \frac{\partial z'}{\partial \alpha} Ax + \frac{\partial x'}{\partial \alpha} A'z .$$

As a special case we have

$$\frac{\partial x'Ax}{\partial x} = \frac{\partial x' \overbrace{Ax}^{c}}{\partial x} + \frac{\partial \overbrace{x'A}^{d'}x}{\partial x}$$

$$= \underbrace{Ax}_{c} + \underbrace{A'x}_{d} = (A + A')x$$

$= 2Ax$ iff A is symmetric.

Rule 5: $$\frac{\partial z'Ax}{\partial \alpha} = \frac{\partial x'}{\partial \alpha} A'z + \frac{\partial z'}{\partial \alpha} Ax$$

v) Second Derivative

$$\underbrace{\frac{\partial}{\partial \alpha}}_{k\times 1}\left[\underbrace{z'}_{1\times l}\underbrace{A}_{l\times l}\underbrace{\frac{\partial x_{l\times 1}}{\partial \alpha'_{1\times k}}}_{l\times k}\right] = \frac{\partial}{\partial \alpha}\left[z'\underbrace{A\frac{\partial x}{\partial \alpha'}}_{\text{fixed}}\right] + \frac{\partial}{\partial \alpha}\left[\underbrace{z'A}_{\text{fixed}}\frac{\partial x}{\partial \alpha'}\right]$$

from Rule 5
$$= \frac{\partial z'}{\partial \alpha}\left[A\frac{\partial x}{\partial \alpha_1} \;\cdots\; A\frac{\partial x}{\partial \alpha_r}\right] + \frac{\partial}{\partial \alpha}\left[z'A\frac{\partial x}{\partial \alpha'}\right]$$

$$= \underbrace{\frac{\partial z'}{\partial \alpha}}_{k\times l}\underbrace{A}_{l\times l}\underbrace{\frac{\partial x}{\partial \alpha'}}_{l\times k} + \frac{\partial}{\partial \alpha}\left[\underbrace{z'A}_{v'}\underbrace{\frac{\partial x}{\partial \alpha'}}_{M'}\right]$$

from Rule 3
$$= \frac{\partial z'}{\partial \alpha}A\frac{\partial x}{\partial \alpha'} + \underbrace{\frac{\partial}{\partial \alpha}}_{k\times 1}\left\{\underbrace{\mathrm{vec}\left(\frac{\partial x}{\partial \alpha'}\right)'}_{1\times kl}\underbrace{(\underbrace{I_k}_{k\times k}\otimes \underbrace{A'z}_{l\times 1})}_{kl\times k}\right\}$$

$$= \frac{\partial z'}{\partial \alpha}A\frac{\partial x}{\partial \alpha'} + \frac{\partial}{\partial \alpha}\left\{\mathrm{vec}\left(\frac{\partial x}{\partial \alpha'}\right)'\right\}(I_k\otimes A'z)$$

Rule 6: $$\frac{\partial}{\partial \alpha}\left[z'A\frac{\partial x}{\partial \alpha'}\right] = \frac{\partial z'}{\partial \alpha}A\frac{\partial x}{\partial \alpha'} + \frac{\partial}{\partial \alpha}\left\{\mathrm{vec}\left(\frac{\partial x}{\partial \alpha'}\right)'\right\}(I_k\otimes A'z)$$

EXAMPLE 1: Linear GMM

Let X be $n \times k$, Z be $n \times l$, $\theta \in \mathbb{R}^k$, $y \in \mathbb{R}^n$, and W be $l \times l$ and symmetric.
The sample criterion function is

$$Q_n(\theta) = \frac{1}{2}[Z'(Y - X\theta)]' \, W \, [Z'(Y - X\theta)] \quad := a' W a$$

- $$\frac{\partial Q_n(\theta)}{\partial \theta} = \frac{1}{2} \frac{\partial a'}{\partial \theta} W a + \frac{1}{2} \frac{\partial a'}{\partial \theta} W' a$$

$$= \frac{\partial a'}{\partial \theta} W a$$

$$= \frac{\partial}{\partial \theta} \Big[Y'Z - \theta' \underbrace{X'Z}_{k \times n \;\; n \times l} \Big] W \, [Z'(Y - X\theta)]$$

(*) $c_1 \dots c_l$ are the columns of $X'Z$

$$= \frac{\partial}{\partial \theta} [\theta' c_1 \;\; \dots \;\; \theta' c_l] \; W \; [Z'(Y - X\theta)]$$

$$= [c_1 \;\; \dots \;\; c_l] \; W \; [Z'(Y - X\theta)]$$

$$= X'Z \; W \; Z'(Y - X\theta)]$$

- $$\frac{\partial^2 Q_n(\theta)}{\partial \theta \, \partial \theta'} = \frac{\partial}{\partial \theta} \Big[a' W \underbrace{\frac{\partial a}{\partial \theta'}}_{\text{doesn't depend on } \theta \text{ anymore!}} \Big]$$

$$= \frac{\partial}{\partial \theta} \Big[(Y - X\theta)' Z \, W \, Z' X \Big]$$

$$= \frac{\partial}{\partial \theta} \Big[\theta' X'Z \, W Z' X \Big]$$

$$= \frac{\partial}{\partial \theta} \begin{bmatrix} \theta' c_1 & \cdots & \theta' c_K \end{bmatrix}$$

$$= X' Z W Z' X .$$

In summation notation it's not that different.

$$Q_n(\theta) = \frac{1}{2} \Big[\frac{1}{n} \sum_{i=1}^{n} (y_i - x_i'\theta) \underset{1\times \ell}{z_i'} \Big] \underset{\ell \times \ell}{W} \Big[\frac{1}{n} \sum_{i=1}^{n} \underset{\ell \times 1}{z_i} (y_i - x_i'\theta) \Big]$$

- $$\frac{\partial Q_n(\theta)}{\partial \theta} = \frac{\partial a'}{\partial \theta} W a$$

$$= \frac{1}{n} \sum_{i=1}^{n} \frac{\partial}{\partial \theta} \Big[\theta' \underset{k\times 1}{x_i} \underset{1\times \ell}{z_i'} \Big] W \Big[\frac{1}{n} \sum_{i=1}^{n} z_i (y_i - x_i'\theta) \Big]$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i z_i' \quad W \Big[\frac{1}{n} \sum_{i=1}^{n} z_i (y_i - x_i'\theta) \Big]$$

- $$\frac{\partial^2 Q_n(\theta)}{\partial \theta \, \partial \theta'} = \frac{\partial}{\partial \theta} \Big\{ \Big[\frac{1}{n} \sum_{i=1}^{n} (y_i - x_i'\theta) z_i' \Big] W \Big[\frac{1}{n} \sum_{i=1}^{n} x_i z_i' \Big] \Big\}$$

$$= \frac{1}{n} \sum_{i=1}^{n} x_i z_i' \quad W \quad \frac{1}{n} \sum_{i=1}^{n} x_i z_i'$$

Notice the connection with the asymptotic linear representation of $\hat{\theta}_n$:

$$\sqrt{n} \, (\hat{\theta}_n - \theta) = \underbrace{\Big(\Big[\frac{1}{n} \sum_{i=1} X_i Z_i' \Big] W \Big[\frac{1}{n} \sum_{i=1}^{n} x_i' z_i \Big] \Big)}_{\text{Hessian}}{}^{-1} \underbrace{\Big[\frac{1}{n} \sum_{i=1}^{n} X_i Z_i' \Big] W \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Z_i u_i}_{\text{gradient}} + o_p(1)$$

EXAMPLE 2: Non-linear GMM

Consider the criterion function $Q_n(\theta) = \frac{1}{2}\left[\frac{1}{n}\sum_{i=1}^{n} \underbrace{g(W_i,\theta)'}_{1\times \ell}\right] \underbrace{A'A}_{\substack{\ell\times\ell \\ \text{Symmetric}}} \left[\frac{1}{n}\sum_{i=1}^{n} \underbrace{g(W_i,\theta)}_{\ell\times 1}\right]$

where W_i' is the data (Y_i, X_i', Z_i').

- $$\frac{\partial Q_n(\theta)_{1\times1}}{\partial \theta_{k\times1}} = \left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)'_{1\times\ell}}{\partial\theta_{k\times1}}\right]_{k\times\ell} A'A \left[\frac{1}{n}\sum_{i=1}^{n} g(W_i,\theta)\right]_{\ell\times1}$$

- $$\frac{\partial Q_n(\theta)}{\partial\theta\,\partial\theta'} = \frac{\partial}{\partial\theta}\left\{ \underbrace{\left[\frac{1}{n}\sum_{i=1}^{n} g(W_i,\theta)'\right] A'A}_{V'} \underbrace{\left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)}{\partial\theta'}\right]}_{M'} \right\}$$

$$= \left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)'}{\partial\theta}\right] A'A \left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)}{\partial\theta'}\right] +$$

RULE 6

$$\frac{\partial}{\partial\theta}\left[\frac{1}{n}\sum_{i=1}^{n} \mathrm{vec}\left(\frac{\partial g(W_i,\theta)}{\partial\theta'}\right)'_{1\times k\ell}\right]\left[\underbrace{I_k}_{k\times k} \otimes A'A\ \frac{1}{n}\sum_{i=1}^{n} \underbrace{g(W_i,\theta)}_{\ell\times1}\right]$$

$$= \left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)'}{\partial\theta}\right] A'A \left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial g(W_i,\theta)}{\partial\theta'}\right] +$$

$$\left[\frac{1}{n}\sum_{i=1}^{n} \frac{\partial}{\partial\theta}\mathrm{vec}\left(\frac{\partial g(W_i,\theta)'}{\partial\theta}\right)'\right]\left[I_k \otimes A'A\ \underbrace{\frac{1}{n}\sum_{i=1}^{n} g(W_i,\theta)}_{\substack{\text{If correctly specified, this} \\ \text{term is } o_p(1).}}\right]$$

linear models are robust to misspecification in terms of asymp. variance.