1. Linear Algebra

- Let x denote a vector, and X denote a matrix.
Unless explicit, their norms are given by

$$\|x\| = \left(\sum_{i=1}^{n} x_i^2\right)^{1/2}, \quad \text{and}$$

Frobenius Norm $\longrightarrow$
$$\|X\| = [\operatorname{trace}(X'X)]^{1/2},$$

where
$$\operatorname{trace}(X) = \sum_{i=1}^{n} \underbrace{X_{ii}}_{\text{diagonal}}.$$

$$(*) \quad X'X = \Big[\underset{k\times 1}{X_1} \; \underset{k\times 1}{X_2} \; \cdots \; \underset{k\times 1}{X_n}\Big] \begin{bmatrix} \overbrace{X_1'}^{1\times k} \\ X_2' \\ \vdots \\ X_n' \end{bmatrix}$$

$$= \begin{bmatrix} X_{11} & X_{21} & \cdots & X_{n1} \\ X_{12} & X_{22} & \cdots & X_{n2} \\ \vdots & \vdots & & \vdots \\ X_{1k} & X_{2k} & & X_{nk} \end{bmatrix} \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1k} \\ X_{21} & X_{22} & \cdots & X_{2k} \\ \vdots & \vdots & & \vdots \\ X_{n1} & X_{n2} & \cdots & X_{nk} \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{i=1}^{n} X_{i1}^2 & \sum_{i=1}^{n} X_{i1} X_{i2} & \cdots & \\ \vdots & \sum^{n} X_{i2}^2 & & \\ & & \ddots & \\ & & & \sum^{n} X_{ik}^2 \end{bmatrix}$$

- Trace is invariant under cyclical permutations:

$$\operatorname{tr}(ABCD) = \operatorname{tr}(BCDA) = \operatorname{tr}(CDBA) = \operatorname{tr}(DABC).$$

- Cauchy-Schwarz: $\quad E\,\|\underset{k\times 1}{a}\,\underset{1\times \ell}{b^T}\| \leq \left(E\|a\|^2 \cdot E\|b\|^2\right)^{1/2}.$

- Matrix multiplication involves projecting rows onto columns:

$$\underset{k\times \ell}{A}\;\underset{\ell\times 1}{b} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{1\ell} \\ \vdots & \ddots & & \\ A_{k1} & A_{k2} & \cdots & A_{k\ell} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_\ell \end{bmatrix} = \begin{bmatrix} A_1^T \\ \vdots \\ A_\ell^T \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_\ell \end{bmatrix}$$

$$= \begin{bmatrix} A_1^T b \\ A_2^T b \\ \vdots \\ A_\ell^T b \end{bmatrix}$$

$$\underset{k\times \ell}{A}\;\underset{\ell\times p}{B} = \begin{bmatrix} A_{11} & A_{12} & \cdots & A_{k1} \\ \vdots & \ddots & & \\ A_{k1} & A_{k2} & \cdots & A_{k\ell} \end{bmatrix} \begin{bmatrix} B_{11} & B_{12} & \cdots & B_{1p} \\ \vdots & \ddots & & \\ B_{\ell 1} & B_{\ell 2} & \cdots & B_{\ell p} \end{bmatrix}$$

$$= \begin{bmatrix} A_1^T \operatorname{col}_1(B) & \cdots & A_1^T \operatorname{col}_p(B) \\ A_2^T \operatorname{col}_1(B) & & A_2^T \operatorname{col}_p(B) \\ \vdots & \ddots & \vdots \\ A_\ell^T \operatorname{col}_1(B) & \cdots & A_\ell^T \operatorname{col}_p(B) \end{bmatrix}$$

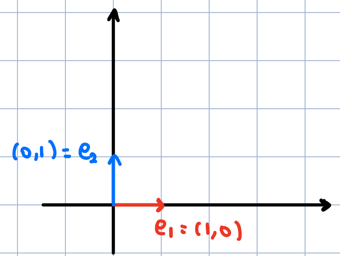
- Matrix A is full rank if

$$Ab = 0 \quad \text{iff} \quad b = 0.$$

- Example of A not having full rank:

$$A = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \quad \text{so that} \quad A : \underbrace{\mathbb{R}^2}_{\text{domain}} \to \underbrace{\mathbb{R}^2}_{\text{codomain}}$$

(where it can come out)



we can combine both to span any vector in $\mathbb{R}^2$

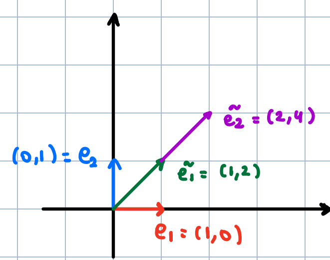
$$A \, e_1 = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \end{bmatrix} = \mathrm{Col}_1(A) := \tilde{e}_1$$

$$A \, e_1 = \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 \\ 4 \end{bmatrix} = \mathrm{Col}_2(A) := \tilde{e}_2$$

* Columns of A tell me where the basis vectors go to!

what happens to vector $(3,2)$?

$$A \begin{bmatrix} 3 \\ 2 \end{bmatrix} = A \left(3 \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{e_1} + 2 \underbrace{\begin{bmatrix} 0 \\ 1 \end{bmatrix}}_{e_2} \right) = 3 \,\mathrm{col}_1(A) + 2 \,\mathrm{col}_2(A)$$



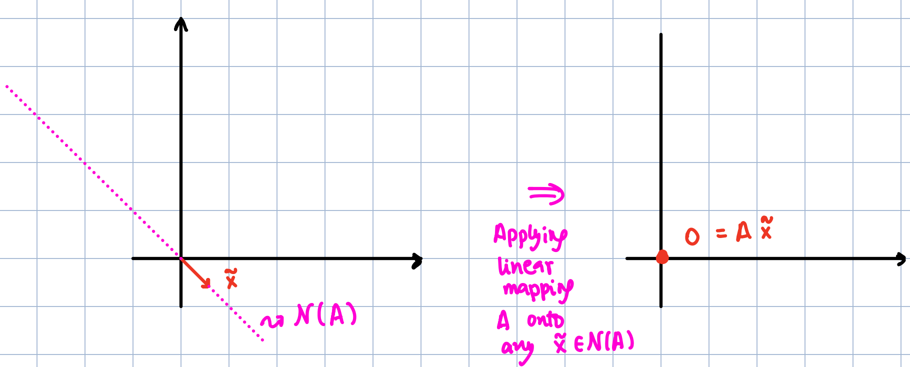
I can no longer linearly combine $\tilde{e}_1$ and $\tilde{e}_2$ to span every vector in $\mathbb{R}^2$.

To find the null space we solve $A\tilde{x} = 0$

$$\Rightarrow \tilde{x}_1 \begin{bmatrix} 1 \\ 2 \end{bmatrix} + \tilde{x}_2 \begin{bmatrix} 2 \\ 4 \end{bmatrix} = 0$$

$$\Rightarrow \tilde{x}_1 = -2\,\tilde{x}_2 \longrightarrow \tilde{x} = \begin{bmatrix} 1 \\ -1/2 \end{bmatrix}$$

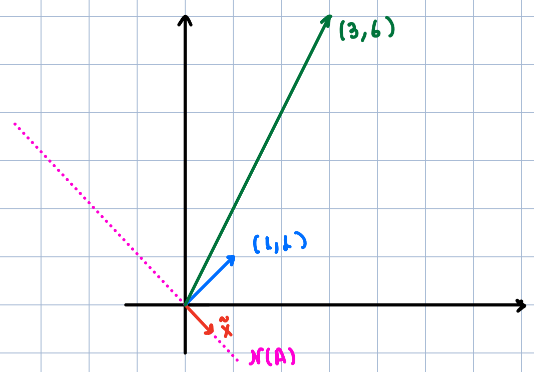
Any scaling of this vector satisfies $\tilde{x} \in N(A)$, where $N(A)$ is the null space of A.



Transformation squishes $\mathbb{R}^2$ space into a range given by rank(A).
Therefore, the rank (in this case rank(A)=1) is a measure of how much space get squished.

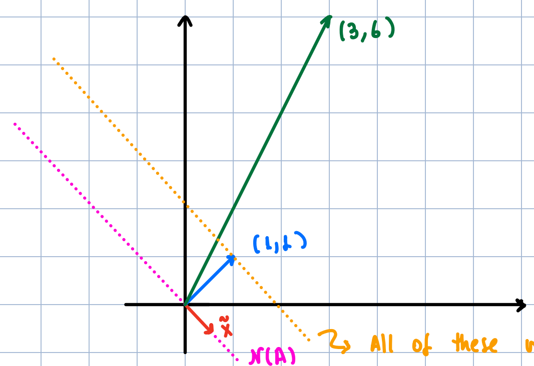
What if we apply A to $\begin{bmatrix}1\\1\end{bmatrix}$?

$$A\begin{bmatrix}1\\1\end{bmatrix} = 1\begin{bmatrix}1\\2\end{bmatrix} + 2\begin{bmatrix}2\\4\end{bmatrix} = \begin{bmatrix}3\\6\end{bmatrix}$$



The problem is that moving $\begin{bmatrix}1\\1\end{bmatrix}$ in the direction of $\mathcal{N}(A)$ also satisfies the same equation:

$$A\begin{bmatrix}1\\1\end{bmatrix} + \underbrace{A\tilde{x}}_{=0} = A\left(\begin{bmatrix}1\\1\end{bmatrix} + \tilde{x}\right) = \begin{bmatrix}3\\6\end{bmatrix}$$



↳ All of these vectors are $\begin{bmatrix}1\\1\end{bmatrix} + \tilde{x}$, $\tilde{x} \in \mathcal{N}(A)$ and also satisfy the linear equation.

The problem of finding a solution for $Ax = \begin{bmatrix}3\\6\end{bmatrix}$ is that it is not a bijective mapping.

That is, from $\begin{bmatrix}3\\6\end{bmatrix}$ we don't have enough information about which x in the orange line was the one that got transformed.

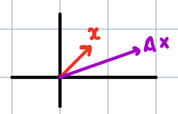
- Matrix A is positive semi-definite if
  $x'Ax \geq 0$ for any $x \neq 0$.

- If matrix A is symmetric and positive definite (e.g. $A = X^T X$), then

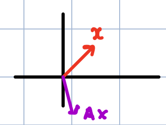
$$A = Q \Lambda Q^T \quad \text{where} \quad Q^T Q = I \quad (\text{i.e. } Q \text{ is orthonormal})$$

- Si $X_{n \times k}$ tiene full rank ($\text{rank}(X) = k$) entonces $\text{rank}(X^T X) = k$ también.

- El producto punto $\underset{1\times1}{\langle x, y \rangle} = x \cdot y = \underset{1\times k}{x^T} \underset{k\times k}{I} \underset{k\times 1}{y}$ puede usar otros pesos $\langle x, y \rangle_A = \langle x, Ay \rangle$



$x^T A x > 0$
$x \neq 0$



$x^T A x < 0$
$x \neq 0$

$\Rightarrow$ Si x y Ax apuntan en una dirección similar decimos que A es positiva definida.

2. Probability and Asymptotics

- The sequence of vectors $X_n = o(a_n)$ means that

$$\lim_{n \to \infty} \left\| \frac{X_n}{a_n} \right\| = 0. \quad \Rightarrow \quad a_n \text{ grows faster than } X_n.$$

- The sequence of vectors $X_n = O(a_n)$ means that $\exists M < \infty$ such that

$$\| X_n \| \leq M \cdot a_n, \quad \forall n \in \mathbb{N}.$$

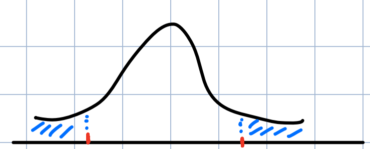
- The sequence of random vectors $X_n = o_p(a_n)$ means that

$$\lim_{n \to \infty} P\left(\left\| \frac{X_n}{a_n} \right\| > \varepsilon \right) = 0, \quad \text{for all } \varepsilon > 0.$$

- The sequence of random vectors $X_n = O_p(a_n)$ means that for all $\varepsilon > 0$ there exists $M_\varepsilon < \infty$ such that

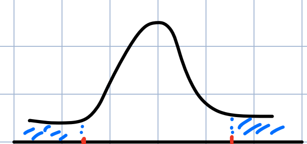
$$\lim_{n \to \infty} P\left(\left\| \frac{X_n}{a_n} \right\| > M_\varepsilon \right) < \varepsilon.$$

Example: $X_n = Z \sim N(0,1)$



$\varepsilon = 0.10$ $\quad -1.645 \quad M_\varepsilon = 1.645$

Prob of blue area is 0.10



$\varepsilon = 0.05$ $\quad -1.96 \quad 1.96$

Prob of blue area is 0.05

- Algebra of little o :
  (a) $o_p(1) + o_p(1) = o_p(1)$.
  (b) $O_p(1) + o_p(1) = O_p(1)$.
  (c) $O_p(1) \cdot o_p(1) = o_p(1)$.
  (d) $o_p(1) \cdot O(1) = o_p(1)$.
  (e) $o(1)$ sequence is also $o_p(1)$.
  (f) $O_p(a_n) \cdot o_p(b_n) = o_p(a_n \cdot b_n)$.
  (g) $O_p(a_n) + O_p(b_n) = O_p(\max\{a_n, b_n\})$.

- Let $X_n \xrightarrow[d]{} X$, for some random vector X, and $A_n \xrightarrow[p]{} a$ for some constant a. Then,

$$(X_n, a_n) \xrightarrow[d]{} (X, a).$$

- Continuous Mapping Theorem: let $\{X_n\}$ be a sequence of random vectors in $\mathbb{R}^d$ such that $X_n \to X$. Also, let $g(\cdot): \mathbb{R}^d \to \mathbb{R}^k$ be a continuous function on a set C such that $P\{X \in C\} = 1$ (almost everywhere). Then

$$g(X_n) \xrightarrow[d]{} g(X) \quad \text{as} \quad n \to \infty.$$

- Slutsky's Theorem : let $X_n \xrightarrow[d]{} X$ and $A_n \xrightarrow[p]{} a$ for some constant a. Then,

  (a) $X_n + A_n \xrightarrow[d]{} X + a$,

  (b) $A_n X_n \xrightarrow[d]{} aX$.

  * Trivial implication : $A_n X_n + B_n \xrightarrow[d]{} aX + b$ when $B_n \xrightarrow[p]{} b$.

- Weak Law of Large Numbers (iid) : let $\{X_i\}_1^n$ be a sequence of iid random vectors such that $E\|X_i\| < \infty$. Then

$$\frac{1}{n} \sum_{i=1}^{n} X_i \xrightarrow[p]{} E X_i,$$

which can also be written as

$$\frac{1}{n} \sum_{i=1}^{n} X_i = E X_i + o_p(1).$$

- <u>Central Limit Theorem (iid)</u>: Let $\{x_i\}_{i=1}^n$ be an iid sequence of random variables. Suppose $Var(X_i)$ is finite and bounded away from zero. Then

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{X_i - EX_i}{Var(X_i)} \xrightarrow[d]{} \mathcal{N}(0,1) .$$

For the case of iid random vectors we require $Var(X_i) = O(1)$ and positive definite. Then

$$Var(X_i)^{-1/2} \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - EX_i) \xrightarrow[d]{} \mathcal{N}(0, I) .$$

3. Objetivo 1 del curso

Vamos a mostrar que para un conjunto de estimadores que vienen de un problema de minimización / maximización

$$\hat{\theta}_n \in \underset{\theta}{\operatorname{argmin}} \underbrace{Q_n(\theta)}_{\substack{\text{función objetivo} \\ \text{(criterion function)}}}$$

se puede obtener la siguiente representación:

$$\underset{k\times 1}{\sqrt{n}(\hat{\theta}_n - \theta)} = \underbrace{\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \underset{k\times 1}{\xi_i} + o_p(1)}_{\text{Asymptotic Linear Representation}} ,$$

Asymptotic Linear Representation. Además ξ_i se le conoce como influence function.

$$\xrightarrow[d]{} \mathcal{N}(0, E\, \underset{k\times 1}{\xi_i}\, \underset{1\times k}{\xi_i^T}) .$$

Ejemplo: $\underset{1\times 1}{Q_n(\theta)} = \frac{1}{n}\sum_{i=1}^{n} (y_i - x_i^T\theta)^2 = \underset{1\times n}{(Y - X\theta)^T} \underset{n\times 1}{(Y - X\theta)}$ (OLS)

Nos da como resultado $\xi_i = \underbrace{\left(\frac{1}{n}\sum_{i=1}^{n} x_i x_i^T\right)^{-1}}_{\text{Hessiano de } Q_n(\theta)} \underbrace{x_i \varepsilon_i}_{\approx \text{ Gradiente de } Q_n(\theta)}$.

Es decir, la primera y segunda derivada de $Q_n(\theta)$ nos da la información suficiente para entender la distribución asintótica del estimador.

Deben entender bien como derivar con respecto a un vector.

4. Diferenciación

- La derivada representa una linealización:

$$\underbrace{f(x+\delta x)}_{\text{vector}} = f(x) + f'(x)\,\delta x + \underbrace{o(\delta x)}_{\substack{\text{Términos de mayor orden } \to 0 \\ \text{mientras } \|\delta x\| \to \infty}}$$

- En notación de diferenciales

$$df = f(x+dx) - f(x) = f'(x)\,dx ,$$

↳ arbitrariamente pequeño.

$\Rightarrow$ Δ output $=$ Operador Lineal (Matriz) $\cdot$ Δ input .

Ejemplo: $\quad \underset{1\times 1}{f(\underset{n\times 1}{x})} = \underset{1\times n}{x'}\,\underset{n\times n}{A}\,\underset{n\times 1}{x}$

$$\begin{aligned}
df &= f(x+dx) - f(x) \\
&= (x+dx)'A(x+dx) - x'Ax \\
&= \cancel{x'Ax} + x'A\,dx + dx'Ax + \underbrace{dx'A\,dx}_{\substack{\text{Término de} \\ \text{orden 2} \\ \to 0}} - \cancel{x'Ax} \\
&= x'A\,dx + dx'Ax \\
&= x'A\,dx + x'A'\,dx \\
&= \underbrace{x'(A+A')}_{\underset{1\times n}{f'(x)} := \nabla f^T := \left[\frac{\partial}{\partial x} f(x)\right]^T := \frac{\partial}{\partial x^T} f(x)} dx
\end{aligned}$$

↳ podemos transponer una matriz 1×1 y no afecta

- En dimensiones generales tenemos $\underset{l\times 1}{g(\underset{k\times 1}{x})} : \mathbb{R}^k \to \mathbb{R}^l$. Entonces,

$$dg = \underbrace{g'(x)}_{\substack{\text{Operador} \\ \text{Lineal} \\ l\times k \\ \text{(Matriz Jacobiana)}}} \underset{k\times 1}{dx} ,$$

donde su elemento (i,j) es $\dfrac{\partial g_i}{\partial x_j}$.

• Reglas de Diferenciación:

(a) SUMA:

$$\underset{\ell\times 1}{f(x)} = \underset{\ell\times 1}{g(x)} + \underset{\ell\times 1}{h(x)}$$

$$\Rightarrow \underset{\ell\times 1}{df} = dg + dh = \underset{\ell\times k}{g'(x)}\,\underset{k\times 1}{dx} + \underset{\ell\times k}{h'(x)}\,\underset{k\times 1}{dx} \quad \Big\}\ f'(x) = g'(x) + h'(x)$$

(b) PRODUCTO:

$$\underset{1\times 1}{f(x)} = \underset{1\times \ell}{g(x)^T}\,\underset{\ell\times 1}{h(x)}$$

$$\begin{aligned}\Rightarrow \underset{1\times 1}{df} &= dg^T\, h(x) + g(x)\, dh \\ &= dx^T\, g'(x)^T\, h(x) + g(x)^T\, h'(x)\, dx \\ &= h(x)^T\, g'(x)\, dx + g(x)^T\, h'(x)\, dx \\ &= [\, \underset{1\times\ell}{h(x)^T}\, \underset{\ell\times k}{g'(x)} + g(x)^T h'(x)\,]\, dx\end{aligned} \quad \Bigg\}\ f'(x) = h(x)^T g'(x) + g(x)^T h'(x)$$

$$\underset{k\times 1}{f(x)} = \underbrace{\underset{k\times\ell}{g(x)}}_{\substack{\text{Sabemos derivar un vector, pero} \\ \text{esto es una matriz.}}}\ \underset{\ell\times 1}{h(x)}$$

Def.- (Vec) El operador vec(A) vectoriza la matriz A en un vector que pone una columna encima de la otra. Dado A siendo $k\times\ell$, tenemos vec(A) es $k\ell\times 1$.

$$\underset{4\times 1}{\mathrm{vec}(\underset{2\times 2}{A})} = \mathrm{vec}\begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22}\end{pmatrix} = \begin{bmatrix} A_{11} \\ A_{21} \\ A_{12} \\ A_{22}\end{bmatrix}$$

Utilizando la siguiente propiedad de vec:

$$\underset{k\times\ell}{M^T}\,\underset{\ell\times 1}{v} = [\, I_k \otimes v^T\,]\ \mathrm{vec}(M)$$

$$= \underset{k\times k\ell}{\begin{bmatrix} \underset{1\times\ell}{v^T} & 0 & \cdots & 0 \\ 0 & \underset{1\times\ell}{v^T} & \cdots & 0 \\ \vdots & & \ddots & \\ \underbrace{\underset{1\times\ell}{0} \ \cdots\ \underset{1\times\ell}{0}}_{\substack{k \text{ veces} \\ \text{repetidos}}} & & \cdots & \underset{1\times\ell}{v^T}\end{bmatrix}}\ \underset{k\ell\times 1}{\mathrm{vec}(M)},$$

escribimos

$$\underset{k\times 1}{f(x)} = \underset{k\times\ell}{g(x)}\ \underset{\ell\times 1}{h(x)}$$

$$= g(x)\, h'(x)\, dx + \underbrace{dg\ h(x)}_{\text{Esto debemos entender}}.$$

Usando la propiedad

$$\underset{k\times\ell}{g(x)}\;\underset{\ell\times 1}{h(x)} = \left[I_k \otimes h(x)^T \right] \underbrace{vec(g(x))}$$

¡Esto ya es un vector!

Su derivada es

$$\underset{k\ell\times 1}{vec(dg)} = \underset{1\times k}{\frac{\partial}{\partial x^T}}\;\underset{k\ell\times 1}{vec(g(x))}\;\underset{k\times 1}{dx}$$

Nos da como respuesta final

$$\underset{k\times 1}{f(x)} = \underset{k\times\ell}{g(x)}\;\underset{\ell\times 1}{h(x)}$$

$$= g(x)\, h'(x)\, dx + \left[I_k \otimes h(x)^T \right] \frac{\partial}{\partial x^T} vec(g(x))\, dx$$

(c) CADENA: $f(x) = g(h(x))$

$$\Rightarrow \quad df = g'(h(x))\; dh$$
$$= g'(h(x))\; h'(x)\; dx$$

$\Big\}$ $f'(x) = g'(h(x))\, h'(x)$

Ejemplo: $x \in \mathbb{R}^k$, $h(x) \in \mathbb{R}^n$, $g(h) \in \mathbb{R}$. Por lo que

$$f(x): \mathbb{R}^k \xrightarrow[h]{} \mathbb{R}^n \xrightarrow[g]{} \mathbb{R}$$

$$\Rightarrow \quad \underset{1\times k}{f'(x)} = \underbrace{g'(h(x))}_{1\times n}\;\underbrace{h'(x)}_{n\times k}$$

5. Aplicaciones en OLS y GMM

- En OLS tenemos

$$Q_n(\theta) = \frac{1}{2}(\underbrace{y}_{n\times 1} - \underbrace{X}_{n\times k}\underbrace{\theta}_{k\times 1})^T (y - X\theta)$$

donde

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}_{n\times 1}, \quad X = \begin{bmatrix} x_{11} & \cdots & x_{1k} \\ \vdots & & \\ x_{n1} & \cdots & x_{nk} \end{bmatrix} = \begin{bmatrix} \underbrace{x_1^T}_{1\times k} \\ \vdots \\ \underbrace{x_n^T}_{1\times k} \end{bmatrix}_{n\times k}$$

(a) Podemos perturbar θ directamente

$$dQ = \frac{1}{2}(y - X(\theta + d\theta))^T (y - X(\theta + d\theta)) - \frac{1}{2}(y - X\theta)^T (y - X\theta)$$

$$= -\frac{1}{2}(y - X\theta)^T X\, d\theta - \underbrace{\frac{1}{2}(X\, d\theta)^T (y - X\theta)}_{1\times 1 \text{ se puede transponer y no cambia}}$$

* Recuerden que donde haya más de un $d\theta$ multiplicando se ignora por ser infinitesimal.

$$= \underbrace{-(y - X\theta)^T X}_{Q'(\theta) := \left[\frac{\partial Q_n(\theta)}{\partial \theta}\right]^T} d\theta$$

es decir, $\frac{\partial Q_n(\theta)}{\partial \theta} = -X^T(y - X\theta)$

(b) Podemos usar la regla de la cadena $Q_n(\theta) = \frac{1}{2} h^T h$, $h(\theta) = y - X\theta$

Tenemos primero

$$dQ = \frac{1}{2}(h + dh)^T (h + dh) - \frac{1}{2} h^T h$$

$$= \frac{1}{2} h^T dh + \frac{1}{2} dh^T h = h^T dh \quad (1)$$

Luego

$$dh = [y - X(\theta + d\theta)] - [y - X\theta]$$

$$= -X\, d\theta \quad (2)$$

Combinando (1) y (2) obtenemos:

$$dQ = -(y - X\theta)^T X\, d\theta,$$

que coincide con nuestra respuesta anterior.

Además, podemos perturbar nuevamente para obtener el Hessiano:

Paso 1: Tenemos previamente $dQ = \underbrace{-(y-X\theta)^T X}_{Q'(\theta)} d\theta$

Ahora vamos a diferenciar esto.

Paso 2:

$$\underbrace{dQ'}_{1\times K} = Q'(\theta + d\theta') - Q'(\theta)$$

$$= -(y - X(\theta + d\theta'))^T X - [-(y - X\theta)^T X]$$

$$= (X\, d\theta')^T X$$

$$= d\theta'^T X^T X$$

Paso 3: Lo expresamos como un vector columna, como lo hemos venido haciendo

$$\underbrace{dQ'^T}_{K\times 1} = \underbrace{X^T X}_{\frac{\partial Q'(\theta)^T}{\partial \theta} := \frac{\partial}{\partial \theta^T} Q'(\theta) := \frac{\partial Q(\theta)}{\partial \theta^T \partial \theta}} d\theta'$$

- En GMM de Variables Instrumentales tenemos

$$Q_n(\theta) = \frac{1}{2} \underbrace{g(\theta)^T}_{1\times \ell} \underbrace{W}_{\ell\times\ell} \underbrace{g(\theta)}_{\ell\times 1}, \quad \text{donde} \quad g(\theta) = \underbrace{Z'(y - X\theta)}_{\substack{\text{Momento / Condiciones} \\ \text{que debe cumplir:} \\ Z \text{ no correlacionado} \\ \text{con errores.}}}.$$

Además,

$$\underset{\ell\times\ell}{W} = \begin{bmatrix} w_{11} & \cdots & w_{1\ell} \\ \vdots & \ddots & \vdots \\ w_{\ell 1} & & w_{\ell\ell} \end{bmatrix}_{\ell\times\ell}$$

es una matriz de ponderación que es simétrica y definida positiva. Tiene $\text{rank}(W) = \ell$.

Aplicando las reglas que tenemos primero

$$dQ = \frac{1}{2} dg^T W g + \frac{1}{2} g^T W dg$$

$$= g^T W dg \quad ①$$

Luego

$$dg = Z'(y - X(\theta + d\theta)) - Z'(y - X\theta)$$

$$= -Z'X\, d\theta \quad ②$$

Combinamos 1 y 2 para obtener

$$dQ = \underbrace{-[z'(Y-X\theta)]^T W z'X}_{Q'(\theta)} \, d\theta$$

$$Q'(\theta) := \frac{\partial Q(\theta)}{\partial \theta^T} \Rightarrow \frac{\partial Q(\theta)}{\partial \theta} = -X'Z W Z' (Y - X\theta)$$

Para obtener el Hessiano :

Paso 1: Tenemos previamente $dQ = \underbrace{-[z'(Y-X\theta)]^T W z'X}_{Q'(\theta)} \, d\theta$

Ahora vamos a diferenciar esto.

Paso 2: Expresamos como vector columna

$$\underbrace{dQ'^T}_{K\times 1} = Q'(\theta + d\theta')^T - Q'(\theta)^T$$

$$= -X'Z W Z'(Y - X(\theta + d\theta')) + X'Z W Z'(Y - X\theta)$$

$$= \underbrace{X'Z W Z'X}_{\frac{\partial Q_n(\theta)}{\partial \theta^T \partial \theta}} \, d\theta'$$