MAT-269: Diferenciación matricial

Felipe Osorio

fosorios.mat.utfsm.cl

Departamento de Matemática, UTFSM

Notación:

Denotaremos por ϕ, $\boldsymbol{f}$ y $\boldsymbol{F}$ funciones escalar, vectorial y matricial, respectivamente mientras que ζ, $\boldsymbol{x}$ y $\boldsymbol{X}$ argumentos escalar, vectorial y matricial, respectivamente.

Ejemplo:

Podemos escribir los siguientes casos particulares:

$$\begin{array}{lll} \phi(\zeta) = \zeta^2, & \phi(\boldsymbol{x}) = \boldsymbol{a}^\top \boldsymbol{x}, & \phi(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{X}^\top \boldsymbol{X}), \\ \boldsymbol{f}(\zeta) = (\zeta, \zeta^2)^\top, & \boldsymbol{f}(\boldsymbol{x}) = \boldsymbol{A}\boldsymbol{x}, & \boldsymbol{f}(\boldsymbol{X}) = \boldsymbol{X}\boldsymbol{a}, \\ \boldsymbol{F}(\zeta) = \zeta^2 \boldsymbol{I}_n, & \boldsymbol{F}(\boldsymbol{x}) = \boldsymbol{x}\boldsymbol{x}^\top, & \boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{X}^\top. \end{array}$$

Considere $\phi : S \to \mathbb{R}$ con $S \subset \mathbb{R}^n$, se define la derivada de ϕ con relación a $\boldsymbol{x} \in S$ como

$$\frac{\partial \phi(\boldsymbol{x})}{\partial \boldsymbol{x}} = \Big(\frac{\partial \phi}{\partial x_1}, \dots, \frac{\partial \phi}{\partial x_n}\Big)^\top = \Big(\frac{\partial \phi}{\partial x_i}\Big) \in \mathbb{R}^n$$

de este modo, introducimos la notación

$$\mathsf{D}\, \phi(\boldsymbol{x}) = \frac{\partial \phi(\boldsymbol{x})}{\partial \boldsymbol{x}^\top} \in \mathbb{R}^{1\times n}.$$

Ahora, si $\boldsymbol{f} : S \to \mathbb{R}^m$, $S \subset \mathbb{R}^n$. Entonces la matriz $m \times n$,

$$\mathsf{D}\, \boldsymbol{f}(\boldsymbol{x}) = \begin{pmatrix} \mathsf{D}\, f_1(\boldsymbol{x}) \\ \vdots \\ \mathsf{D}\, f_m(\boldsymbol{x}) \end{pmatrix} = \frac{\partial \boldsymbol{f}(\boldsymbol{x})}{\partial \boldsymbol{x}^\top},$$

es la derivada o matriz Jacobiana de $\boldsymbol{f}$. La transpuesta de la matriz Jacobiana $\mathsf{D}\, \boldsymbol{f}(\boldsymbol{x})$ se denomina gradiente de $\boldsymbol{f}(\boldsymbol{x})$.

Considere la fórmula de Taylor de primer orden,

$$\phi(c+u) = \phi(c) + u\phi'(c) + r_c(u),$$

donde,

$$\lim_{u\to 0} \frac{r_c(u)}{u} = 0.$$

es de orden más pequeño que u conforme $u \to 0$. Note también que

$$\lim_{u\to 0} \frac{\phi(c+u) - \phi(c)}{u} = \phi'(c).$$

De este modo, se define

$$\mathsf{d}\,\phi(c; u) = u\phi'(c),$$

como el (primer) diferencial de ϕ en c con incremento u. Esto motiva la siguiente definición.

Definición 1:

Sea $\boldsymbol{f} : S \to \mathbb{R}^m$, $S \subset \mathbb{R}^n$, si existe una matriz $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, tal que

$$\boldsymbol{f}(\boldsymbol{c} + \boldsymbol{u}) = \boldsymbol{f}(\boldsymbol{c}) + \boldsymbol{A}(\boldsymbol{c})\boldsymbol{u} + \boldsymbol{r}_c(\boldsymbol{u}),$$

para todo $\boldsymbol{u} \in \mathbb{R}^n$ con $||\boldsymbol{u}|| < \delta$, y

$$\lim_{u \to 0} \frac{\boldsymbol{r}_c(\boldsymbol{u})}{||\boldsymbol{u}||} = \boldsymbol{0},$$

entonces la función $\boldsymbol{f}$ se dice diferenciable en $\boldsymbol{c}$. El vector $m \times 1$

$$\mathsf{d}\,\boldsymbol{f}(\boldsymbol{c}; \boldsymbol{u}) = \boldsymbol{A}(\boldsymbol{c})\boldsymbol{u},$$

se denomina primer diferencial de $\boldsymbol{f}$ en $\boldsymbol{c}$ con incremento $\boldsymbol{u}$.

Resultado 1 (Magnus y Neudecker, 1985)[1]:

Sea $\boldsymbol{f}: S \to \mathbb{R}^m$, $S \subset \mathbb{R}^n$ función diferenciable, $\boldsymbol{c} \in S$ y $\boldsymbol{u}$ un vector n-dimensional. Entonces

$$\mathrm{d}\,\boldsymbol{f}(\boldsymbol{c};\boldsymbol{u}) = (\mathrm{D}\,\boldsymbol{f}(\boldsymbol{c}))\boldsymbol{u}.$$

La matriz $\mathrm{D}\,\boldsymbol{f}(\boldsymbol{c}) \in \mathbb{R}^{m\times n}$ se denomina matriz Jacobiana. Tenemos también que

$$\nabla \boldsymbol{f}(\boldsymbol{c}) = (\mathrm{D}\,\boldsymbol{f}(\boldsymbol{c}))^\top$$

es la matriz gradiente de $\boldsymbol{f}$.

[1] Journal of Mathematical Psychology **29**, 474-492.
También es conocido como el "Primer teorema de identificación"

Definición 2:

Sea $\boldsymbol{A} \in \mathbb{R}^{n\times q}$ particionada como

$$\boldsymbol{A} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_q),$$

donde $\boldsymbol{a}_k \in \mathbb{R}^n$ es la k-ésima columna de $\boldsymbol{A}$. Entonces

$$\operatorname{vec}(\boldsymbol{A}) = \begin{pmatrix} \boldsymbol{a}_1 \\ \vdots \\ \boldsymbol{a}_q \end{pmatrix}.$$

Definición 3:

Sea $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ y $\boldsymbol{B} \in \mathbb{R}^{p\times q}$, entonces el producto Kronecker entre $\boldsymbol{A}$ y $\boldsymbol{B}$ denotado por $\boldsymbol{A} \otimes \boldsymbol{B}$ es la matriz $mp \times nq$ definida como

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{pmatrix} a_{11}\boldsymbol{B} & \dots & a_{1n}\boldsymbol{B} \\ \vdots & & \vdots \\ a_{m1}\boldsymbol{B} & \dots & a_{mn}\boldsymbol{B} \end{pmatrix}$$

Resultado 2:

Sean $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ y $\boldsymbol{D}$ matrices de órdenes apropiados y λ escalar. Entonces

(a) $\boldsymbol{A} \otimes \boldsymbol{B} \otimes \boldsymbol{C} = (\boldsymbol{A} \otimes \boldsymbol{B}) \otimes \boldsymbol{C} = \boldsymbol{A} \otimes (\boldsymbol{B} \otimes \boldsymbol{C})$,

(b) $(\boldsymbol{A} + \boldsymbol{B}) \otimes (\boldsymbol{C} + \boldsymbol{D}) = \boldsymbol{A} \otimes \boldsymbol{C} + \boldsymbol{B} \otimes \boldsymbol{C} + \boldsymbol{A} \otimes \boldsymbol{D} + \boldsymbol{B} \otimes \boldsymbol{D}$,

(c) $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{A}\boldsymbol{C} \otimes \boldsymbol{B}\boldsymbol{D}$,

(d) $\lambda \otimes \boldsymbol{A} = \lambda \boldsymbol{A} = \boldsymbol{A} \otimes \lambda$,

(e) $(\boldsymbol{A} \otimes \boldsymbol{B})^\top = \boldsymbol{A}^\top \otimes \boldsymbol{B}^\top$,

(f) $(\boldsymbol{A} \otimes \boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} \otimes \boldsymbol{B}^{-1}$,

(g) $(\boldsymbol{A} \otimes \boldsymbol{B})^{-} = \boldsymbol{A}^{-} \otimes \boldsymbol{B}^{-}$.

Resultado 3:

Sean $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ y $\boldsymbol{B} \in \mathbb{R}^{p\times p}$. Entonces

(a) $\operatorname{tr}(\boldsymbol{A} \otimes \boldsymbol{B}) = \operatorname{tr}(\boldsymbol{A}) \operatorname{tr}(\boldsymbol{B})$,

(b) $|\boldsymbol{A} \otimes \boldsymbol{B}| = |\boldsymbol{A}|^p |\boldsymbol{B}|^n$,

(c) $\operatorname{rg}(\boldsymbol{A} \otimes \boldsymbol{B}) = \operatorname{rg}(\boldsymbol{A}) \operatorname{rg}(\boldsymbol{B})$.

Observación:

Si $\boldsymbol{a} \in \mathbb{R}^n$ y $\boldsymbol{b} \in \mathbb{R}^p$, entonces

$$\boldsymbol{a}\boldsymbol{b}^\top = \boldsymbol{a} \otimes \boldsymbol{b}^\top = \boldsymbol{b}^\top \otimes \boldsymbol{a}.$$

Por otro lado, tenemos que

$$\operatorname{vec}(\boldsymbol{a}\boldsymbol{b}^\top) = \operatorname{vec}(\boldsymbol{a} \otimes \boldsymbol{b}^\top) = \operatorname{vec}(\boldsymbol{b}^\top \otimes \boldsymbol{a}) = \boldsymbol{b} \otimes \boldsymbol{a}.$$

Esto sugiere una conexión entre el operador de vectorización, el producto Kronecker y la traza.

Resultado 4:

(a) Si $\boldsymbol{A}$ y $\boldsymbol{B}$ son ámbas matrices de orden $m \times n$, entonces

$$\operatorname{tr} \boldsymbol{A}^\top \boldsymbol{B} = \operatorname{vec}^\top \boldsymbol{A} \operatorname{vec} \boldsymbol{B},$$

(b) Si $\boldsymbol{A}, \boldsymbol{B}$ y $\boldsymbol{C}$ son de órdenes adecuados, entonces

$$\operatorname{vec} \boldsymbol{ABC} = (\boldsymbol{C}^\top \otimes \boldsymbol{A}) \operatorname{vec} \boldsymbol{B},$$

donde $\operatorname{vec}^\top \boldsymbol{A} = (\operatorname{vec} \boldsymbol{A})^\top$.

Resultado 5:

Sean $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}$ y $\boldsymbol{D}$ matrices, tal que, el producto $\boldsymbol{ABCD}$ está definido y es cuadrado, entonces

$$\operatorname{tr} \boldsymbol{ABCD} = \operatorname{vec}^\top \boldsymbol{D}^\top (\boldsymbol{C}^\top \otimes \boldsymbol{A}) \operatorname{vec} \boldsymbol{B} = \operatorname{vec}^\top \boldsymbol{D} (\boldsymbol{A} \otimes \boldsymbol{C}^\top) \operatorname{vec} \boldsymbol{B}^\top.$$

Sea $\boldsymbol{F} : S \to \mathbb{R}^{m\times p}$, $S \subset \mathbb{R}^{n\times q}$ una función matricial, podemos notar que

$$\operatorname{vec} \boldsymbol{F}(\boldsymbol{X}) = \boldsymbol{f}(\operatorname{vec} \boldsymbol{X})$$

esto permite obtener el diferencial de una función matricial considerando la relación

$$\operatorname{vec} \mathsf{d}\, \boldsymbol{F}(\boldsymbol{C}; \boldsymbol{U}) = \mathsf{d}\, \boldsymbol{f}(\operatorname{vec} \boldsymbol{C}; \operatorname{vec} \boldsymbol{U})$$

en cuyo caso $\boldsymbol{F}$ tiene matriz Jacobiana

$$\mathsf{D}\, \boldsymbol{F}(\boldsymbol{C}) = \mathsf{D}\, \boldsymbol{f}(\operatorname{vec} \boldsymbol{C})$$

Resultado 6:

Sea $\boldsymbol{F} : S \to \mathbb{R}^{m\times p}$, $S \subset \mathbb{R}^{n\times q}$ función diferenciable, $\boldsymbol{C} \in S$ y $\boldsymbol{U}$ matriz $n \times q$. Entonces

$$\operatorname{vec} \mathsf{d}\, \boldsymbol{F}(\boldsymbol{C}; \boldsymbol{U}) = (\mathsf{D}\, \boldsymbol{F}(\boldsymbol{C})) \operatorname{vec} \boldsymbol{U}.$$

con $(\mathsf{D}\, \boldsymbol{F}(\boldsymbol{C}))^\top$ la matriz gradiente de $\boldsymbol{F}$.

Considere $\phi : S \to \mathbb{R}$ con $S \subset \mathbb{R}^n$, entonces se define la matriz Hessiana como la matriz de segundas derivadas, dada por

$$\mathsf{H}\,\phi(\boldsymbol{x}) = \frac{\partial^2 \phi(\boldsymbol{x})}{\partial \boldsymbol{x} \partial \boldsymbol{x}^\top} = \frac{\partial}{\partial \boldsymbol{x}^\top}\left(\frac{\partial \phi(\boldsymbol{x})}{\partial \boldsymbol{x}^\top}\right)^\top = \mathsf{D}(\mathsf{D}\,\phi(\boldsymbol{x}))^\top.$$

Evidentemente, el segundo diferencial de una función escalar está dado por

$$\mathsf{d}^2\,\phi = \mathsf{d}(\mathsf{d}\,\phi).$$

Resultado 7:

Sea $\phi : S \to \mathbb{R}$, $S \subset \mathbb{R}^n$ dos veces diferenciable, $\boldsymbol{c} \in S$ y $\boldsymbol{u}$ vector n-dimensional. Entonces

$$\mathsf{d}^2\,\phi(\boldsymbol{c}; \boldsymbol{u}) = \boldsymbol{u}^\top(\mathsf{H}\,\phi(\boldsymbol{c}))\boldsymbol{u}.$$

donde $\mathsf{H}\,\phi(\boldsymbol{c}) \in \mathbb{R}^{n \times n}$ es la matriz Hessiana de ϕ.

Observación:

Algunas ventajas (prácticas) importantes del cálculo de diferenciales son:

- Sea $\boldsymbol{f}(\boldsymbol{x})$ función vectorial $m \times 1$ con argumento $\boldsymbol{x}$, vector n-dimensional, entonces
$$\mathsf{D}\,\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{m\times n} \qquad \text{sin embargo,} \qquad \mathsf{d}\,\boldsymbol{f}(\boldsymbol{x}) \in \mathbb{R}^{m}$$
- Para funciones matriciales, $\mathsf{d}\,\boldsymbol{F}(\boldsymbol{X})$ tiene la misma dimensión que $\boldsymbol{F}$ sin importar la dimensión de $\boldsymbol{X}$.

Reglas fundamentales:

Considere u y v funciones escalares y α una constante, entonces:

$$
\begin{aligned}
&\mathsf{d}\,\alpha = 0, && \mathsf{d}(\alpha u) = \alpha\,\mathsf{d}\,u,\\
&\mathsf{d}(u+v) = \mathsf{d}\,u + \mathsf{d}\,v, && \mathsf{d}(uv) = (\mathsf{d}\,u)v + u(\mathsf{d}\,v)\\
&\mathsf{d}(u/v) = \frac{(\mathsf{d}\,u)v - u(\mathsf{d}\,v)}{v^2}, (v \neq 0), && \mathsf{d}\,u^{\alpha} = \alpha u^{\alpha-1}\,\mathsf{d}\,u,\\
&\mathsf{d}\,e^{u} = e^{u}\,\mathsf{d}\,u, && \mathsf{d}\log u = u^{-1}\,\mathsf{d}\,u, (u > 0)\\
&\mathsf{d}\,\alpha^{u} = \alpha^{u}\log\alpha\,\mathsf{d}\,u, (\alpha > 0).
\end{aligned}
$$

Aquí por ejemplo,

$$\phi(x) = u(x) + v(x).$$

Reglas fundamentales:

Análogamente para $\boldsymbol{U}$, $\boldsymbol{V}$ funciones matriciales, α un escalar (constante) y $\boldsymbol{A} \in \mathbb{R}^{m\times n}$ constante, tenemos

$$
\begin{aligned}
&\mathsf{d}\,\boldsymbol{A} = \boldsymbol{0}, && \mathsf{d}(\alpha\boldsymbol{U}) = \alpha\,\mathsf{d}\,\boldsymbol{U},\\
&\mathsf{d}(\boldsymbol{U}+\boldsymbol{V}) = \mathsf{d}\,\boldsymbol{U} + \mathsf{d}\,\boldsymbol{V}, && \mathsf{d}(\boldsymbol{U}\boldsymbol{V}) = (\mathsf{d}\,\boldsymbol{U})\boldsymbol{V} + \boldsymbol{U}\,\mathsf{d}\,\boldsymbol{V},\\
&\mathsf{d}(\boldsymbol{U}\otimes\boldsymbol{V}) = \mathsf{d}\,\boldsymbol{U}\otimes\mathsf{d}\,\boldsymbol{V}, && \mathsf{d}(\boldsymbol{U}\odot\boldsymbol{V}) = \mathsf{d}\,\boldsymbol{U}\odot\mathsf{d}\,\boldsymbol{V},\\
&\mathsf{d}\,\boldsymbol{U}^\top = (\mathsf{d}\,\boldsymbol{U})^\top, && \mathsf{d}\operatorname{vec}\boldsymbol{U} = \operatorname{vec}\mathsf{d}\,\boldsymbol{U},\\
&\mathsf{d}\operatorname{tr}\boldsymbol{U} = \operatorname{tr}\mathsf{d}\,\boldsymbol{U}.
\end{aligned}
$$

Otros diferenciales de uso frecuente en Estadística son:

$$
\mathsf{d}\,|\boldsymbol{F}| = |\boldsymbol{F}|\operatorname{tr}\boldsymbol{F}^{-1}\,\mathsf{d}\,\boldsymbol{F}, \qquad \mathsf{d}\log|\boldsymbol{F}| = \operatorname{tr}\boldsymbol{F}^{-1}\,\mathsf{d}\,\boldsymbol{F},
$$
$$
\mathsf{d}\,\boldsymbol{F}^{-1} = -\boldsymbol{F}^{-1}(\mathsf{d}\,\boldsymbol{F})\boldsymbol{F}^{-1}.
$$

Sea $\boldsymbol{e}_j = (0, \dots, 1, \dots, 0)^\top \in \mathbb{R}^n$ el j-ésimo vector unitario. Es fácil notar que

$$\boldsymbol{1}_n = \sum_{j=1}^{n} \boldsymbol{e}_j.$$

Considere $\boldsymbol{E}_{ij} = \boldsymbol{e}_i \boldsymbol{e}_j^\top \in \mathbb{R}^{m \times n}$ una matriz de ceros, salvo el elemento (i, j). Además,

$$\boldsymbol{I}_n = \sum_{i=1}^{n} \boldsymbol{E}_{ii} = \sum_{i=1}^{n} \boldsymbol{e}_i \boldsymbol{e}_i^\top.$$

Evidentemente, cualquier matriz $\boldsymbol{A} = (a_{ij}) \in \mathbb{R}^{m \times n}$ puede expresarse como:

$$\boldsymbol{A} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \boldsymbol{E}_{ij} = \sum_{i=1}^{m} \sum_{j=1}^{n} a_{ij} \boldsymbol{e}_i \boldsymbol{e}_j^\top.$$

Observación:

Sea $\boldsymbol{A} \in \mathbb{R}^{m \times n}$, los vectores $\operatorname{vec}(\boldsymbol{A})$ y $\operatorname{vec}(\boldsymbol{A}^\top)$ contienen los mismos elementos, aunque en posiciones diferentes.

Definición 4 (matriz de conmutación):

Existe una única matriz de permutación, $\boldsymbol{K}_{mn} \in \mathbb{R}^{mn \times mn}$ que transforma $\operatorname{vec}(\boldsymbol{A})$ en $\operatorname{vec}(\boldsymbol{A}^\top)$, definida mediante:

$$\boldsymbol{K}_{mn} \operatorname{vec}(\boldsymbol{A}) = \operatorname{vec}(\boldsymbol{A}^\top).$$

esta matriz es llamada matriz de conmutación. En efecto,

$$\boldsymbol{K}_{mn} = \sum_{i=1}^{m} \sum_{j=1}^{n} (\boldsymbol{E}_{ij} \otimes \boldsymbol{E}_{ij}^\top).$$

Ejemplo:

Tenemos

$$\boldsymbol{K}_{23} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

La función `commutation` disponible en la biblioteca fastmatrix permite obtener $\boldsymbol{K}_{mn}$.[2]

Observación:

Si $m = n$ anotaremos simplemente $\boldsymbol{K}_m$. Debemos destacar que $\boldsymbol{K}_{mn}$ y $\boldsymbol{K}_{nm}$ aunque del mismo orden son matrices diferentes.

[2] URL: https://faosorios.github.io/fastmatrix/.

Propiedades:

(a) $\boldsymbol{K}_{mn}^{\top} = \boldsymbol{K}_{nm}$.

(b) $\boldsymbol{K}_{mn}^{\top}\boldsymbol{K}_{mn} = \boldsymbol{K}_{mn}\boldsymbol{K}_{mn}^{\top} = \boldsymbol{I}_{mn}$, es decir $\boldsymbol{K}_{mn}^{-1} = \boldsymbol{K}_{mn}^{\top} = \boldsymbol{K}_{nm}$.

(c) $\boldsymbol{K}_{1n} = \boldsymbol{K}_{n1} = \boldsymbol{I}_{n}$.

(d) $\boldsymbol{K}_{nm}\boldsymbol{K}_{mn} \operatorname{vec} \boldsymbol{A} = \operatorname{vec} \boldsymbol{A}$.

Propiedades:

Sea $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, $\boldsymbol{B} \in \mathbb{R}^{p\times q}$ y $\boldsymbol{b} \in \mathbb{R}^{p}$. Entonces,

(a) $\boldsymbol{K}_{pm}(\boldsymbol{A} \otimes \boldsymbol{B}) = (\boldsymbol{B} \otimes \boldsymbol{A})\boldsymbol{K}_{qn}$.

(b) $\boldsymbol{K}_{pm}(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{K}_{nq} = \boldsymbol{B} \otimes \boldsymbol{A}$.

(c) $\boldsymbol{K}_{pm}(\boldsymbol{A} \otimes \boldsymbol{b}) = \boldsymbol{b} \otimes \boldsymbol{A}$.

(d) $\boldsymbol{K}_{mp}(\boldsymbol{b} \otimes \boldsymbol{A}) = \boldsymbol{A} \otimes \boldsymbol{b}$.

Relacionada con la matriz $\boldsymbol{K}_n$ es la matriz simetrizadora, definida como:

$$\boldsymbol{N}_n = \tfrac{1}{2}(\boldsymbol{I}_{n^2} + \boldsymbol{K}_n).$$

Propiedades:

Sea $\boldsymbol{N}_n = \frac{1}{2}(\boldsymbol{I}_{n^2} + \boldsymbol{K}_n)$. Entonces,

(a) $\boldsymbol{N}_n = \boldsymbol{N}_n^\top = \boldsymbol{N}_n^2$.

(b) $\operatorname{rg}(\boldsymbol{N}_n) = \operatorname{tr}(\boldsymbol{N}_n) = \frac{1}{2}n(n+1)$.

(c) $\boldsymbol{N}_n\boldsymbol{K}_n = \boldsymbol{N}_n = \boldsymbol{K}_n\boldsymbol{N}_n$.

Propiedades:

Para $\boldsymbol{A}$ y $\boldsymbol{B}$ matrices $n \times n$. Tenemos,

(a) $\boldsymbol{N}_n(\boldsymbol{A} \otimes \boldsymbol{B})\boldsymbol{N}_n = \boldsymbol{N}_n(\boldsymbol{B} \otimes \boldsymbol{A})\boldsymbol{N}_n$.

(b) $\boldsymbol{N}_n(\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{N}_n = \boldsymbol{N}_n(\boldsymbol{A} \otimes \boldsymbol{A}) = (\boldsymbol{A} \otimes \boldsymbol{A})\boldsymbol{N}_n$.

Definición 5 (matriz de duplicación):

Para $\boldsymbol{A}$ matriz simétrica $p \times p$, sea $\operatorname{vech}(\boldsymbol{A})$ la vectorización de los elementos distintos[3] de $\boldsymbol{A}$. Existe una única matriz $\boldsymbol{D}_p \in \mathbb{R}^{p^2 \times p(p+1)/2}$ que transforma $\operatorname{vech}(\boldsymbol{A})$ en $\operatorname{vec}(\boldsymbol{A})$, es decir:

$$\boldsymbol{D}_p \operatorname{vech}(\boldsymbol{A}) = \operatorname{vec}(\boldsymbol{A}), \qquad (\boldsymbol{A} = \boldsymbol{A}^\top),$$

y $\boldsymbol{D}_p$ es llamada matrix de duplicación de orden p.

Adicionalmente,

$$\operatorname{vech}(\boldsymbol{A}) = \boldsymbol{D}_p^+ \operatorname{vec}(\boldsymbol{A}), \qquad \boldsymbol{D}_p^+ = (\boldsymbol{D}_p^\top \boldsymbol{D}_p)^{-1} \boldsymbol{D}_p^\top.$$

[3] En efecto, tenemos $p(p+1)/2$ elementos distintos.

Ejemplo:

Considere una matriz 3×3,

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

De este modo,

$$\operatorname{vech} \boldsymbol{A} = \begin{pmatrix} a_{11} \\ a_{12} \\ a_{22} \\ a_{13} \\ a_{23} \\ a_{33} \end{pmatrix}, \quad \boldsymbol{D}_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

La función `duplication` disponible en la biblioteca fastmatrix permite obtener $\boldsymbol{D}_p$.

Ejemplo:

Considere la función,

$$\phi(\boldsymbol{X}) = \operatorname{tr}(\boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B})^\top(\boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}).$$

Se desea obtener $\widehat{\boldsymbol{X}}$ como solución del problema

$$\min_{X} \phi(\boldsymbol{X}).$$

Diferenciando $\phi(\boldsymbol{X})$ con relación a $\boldsymbol{X}$ tenemos

$$\mathsf{d}_X \, \phi(\boldsymbol{X}) = \operatorname{tr} \mathsf{d}_X (\boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B})^\top(\boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) + \operatorname{tr}(\boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B})^\top \mathsf{d}_X (\boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}).$$

Podemos notar que

$$\mathsf{d}_X (\boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) = -\boldsymbol{A}(\mathsf{d}\,\boldsymbol{X})\boldsymbol{B}.$$

De este modo,

$$\mathsf{d}_X \, \phi(\boldsymbol{X}) = -\operatorname{tr} \boldsymbol{B}^\top(\mathsf{d}\,\boldsymbol{X})^\top \boldsymbol{A}^\top(\boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B}) - \operatorname{tr}(\boldsymbol{C} - \boldsymbol{A}\boldsymbol{X}\boldsymbol{B})^\top \boldsymbol{A}\,\mathsf{d}(\boldsymbol{X})\boldsymbol{B}.$$

Diferenciación matricial

Es fácil notar que

$$\mathsf{d}_X\,\phi(\boldsymbol{X}) = -2\,\mathrm{tr}\,\boldsymbol{A}^\top(\boldsymbol{C}-\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})\boldsymbol{B}^\top(\mathsf{d}\,\boldsymbol{X})^\top.$$

De ahí que, podemos obtener la solución aproximada del problema $\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}=\boldsymbol{C}$, desde la ecuación

$$\boldsymbol{A}^\top(\boldsymbol{C}-\boldsymbol{A}\boldsymbol{X}\boldsymbol{B})\boldsymbol{B}^\top=\boldsymbol{0},$$

o equivalentemente,

$$\boldsymbol{A}^\top\boldsymbol{A}\boldsymbol{X}\boldsymbol{B}\boldsymbol{B}^\top=\boldsymbol{A}^\top\boldsymbol{C}\boldsymbol{B}^\top.$$

Es decir, la solución LS $\widehat{\boldsymbol{X}}$ adopta la forma:

$$\widehat{\boldsymbol{X}}=(\boldsymbol{A}^\top\boldsymbol{A})^{-1}\boldsymbol{A}^\top\boldsymbol{C}\boldsymbol{B}^\top(\boldsymbol{B}\boldsymbol{B}^\top)^{-1}$$

Por otro lado,

$$\begin{aligned}\mathsf{d}^2_X\,\phi(\boldsymbol{X}) &= 2\,\mathrm{tr}\,\boldsymbol{A}^\top\boldsymbol{A}(\mathsf{d}\,\boldsymbol{X})\boldsymbol{B}\boldsymbol{B}^\top(\mathsf{d}\,\boldsymbol{X})^\top\\ &= 2(\mathsf{d}\,\mathrm{vec}\,\boldsymbol{X})^\top(\boldsymbol{B}\boldsymbol{B}^\top\otimes\boldsymbol{A}^\top\boldsymbol{A})\,\mathsf{d}\,\mathrm{vec}\,\boldsymbol{X},\end{aligned}$$

como $\boldsymbol{B}\boldsymbol{B}^\top\otimes\boldsymbol{A}^\top\boldsymbol{A}$ es definida positiva, $\widehat{\boldsymbol{X}}$ es mínimo global.