



Estadística Multivariante

Complementos del álgebra matricial. Derivación matricial.

Trabajo B

Antonio R. Moya
Martín-Castaño
Elena Romero Contreras
Nuria Rodríguez
Barroso
Universidad de Granada
anmomar85@correo.ugr.es
elenaromeroc@correo.ugr.es
rbnuria6@gmail.com

Índice

1. Complementos del álgebra matricial.	2
1.1. Introducción	2
1.2. Operación Vec.	2
1.3. Producto Kronecker	3
1.3.1. Producto Kronecker y Vec. Relaciones de interés entre ambas operaciones.	5
1.3.2. Ejercicios	5
1.4. La matriz conmutación	5
1.5. Producto * de dos matrices	8
1.6. Operación Vech. Matrices de transición. La matriz duplicación	8
2. Derivación matricial.	9
2.1. Introducción	9
2.2. Diferencial primera y jacobianos	9
2.2.1. Diferencial de una función vectorial	9
2.2.2. Diferencial de una función matricial	10
2.3. Matrices jacobianas y derivadas matriciales	11
2.4. Derivadas matriciales de funciones escalares de un vector	13
2.4.1. Derivadas matriciales de funciones escalares de matrices	13
2.5. Diferencial segunda y hessianos	14

1. Complementos del álgebra matricial.

1.1. Introducción

A pesar de que la manipulación algebraica de vectores aleatorios puede ser abordada sin ninguna complicación, las matrices aleatorias suponen un esfuerzo extra ya que el resultado obtenido sobre las mismas debe generalizar al realizado sobre vectores. Además, el uso de matrices aleatorias en el Análisis Multivariante es muy importante dado que aparecen de forma natural, por ejemplo, al estimar la matriz de covarianzas de una población vectorial.

1.2. Operación Vec.

Para que el trato de las matrices aleatorias generalice a la manipulación de vectores aleatorios, debemos *vectorizar* las matrices, esto es, tratarlas como si se tratasen de vectores. Esta consideración se puede hacer teniendo en cuenta que los espacios $\mathbb{M}_{n \times n}$ y \mathbb{R}^{n^2} son isomorfos.

Vemos que el objetivo puede ser muy cómodo a la hora de manipular matrices aleatorias, sin embargo, debemos entender primero bien las propiedades que unen las expresiones matriciales y las *vectorizadas*. Introducimos formalmente la definición de *Vec*

Definición 1.1. Sea \mathbf{X} una matriz de orden $n \times q$. Se define $Vec(\mathbf{X})$ como el vector de dimensión $nq \times 1$ formado al apilar las columnas de X una tras otra, o sea, si notamos por columnas $X = [x_1, x_2, \dots, x_q]$, entonces

$$Vec(X) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_q \end{pmatrix}$$

Como ya habíamos comentado, la identificación entre matrices y las *vectorizadas* se basa en un resultado que formalizamos a continuación:

Teorema 1.1. La aplicación $Vec: \mathbb{M}_{n \times q} \rightarrow \mathbb{R}^{nq}$ es un isomorfismo de espacios vectoriales.

Este isomorfismo presentado en este teorema nos puede servir para calcular la esperanza matemática. Así, cuando sea más fácil calcular la esperanza de su vectorización, utilizando la propiedad que obtenemos el en ejemplo siguiente, bastaría con deshacer el cambio con el isomorfismo anterior.

EJEMPLO 1.1: Sean x_1, \dots, x_N vectores aleatorios p -dimensionales con igual media μ . Sea la matriz aleatoria $\mathbf{X}_{N \times X} = [x_1, \dots, x_N]^t$ y consideremos el vector

$$Vec(X^t) = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix}$$

Entonces, si notamos $\mathbf{1}_N$ al vector N dimensional cuyas componentes son todas iguales a uno, se verifica

$$E[Vec(X^t)] = \begin{pmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{pmatrix} = Vec([\mu, \mu, \dots, \mu]) = Vec(\mu \mathbf{1}_N^t)$$

1.3. Producto Kronecker

Como ya hemos visto en el ejercicio anterior, utilizar la vectorización puede facilitar ciertos cálculos, como el de la esperanza de una matriz aleatoria. Sin embargo, hay ocasiones en las que es obligatorio usar la vectorización de matrices, como en el ejemplo siguiente:

EJEMPLO 1.2: En las hipótesis del ejercicio anterior, suponemos además que x_1, \dots, x_N son independientes y con igual matriz de covarianzas Σ . Entonces se define la matriz de covarianzas de \mathbf{X}^t como $Cov[Vec(X^t)]$ ya que dicha matriz contiene todas las matrices de covarianzas entre las columnas de \mathbf{X}^t . A partir de la definición de la matriz de covarianzas de un vector aleatorio, se verifica

***** (poner definición del apéndice B)

$$Cov[Vec(X^t)] = E[[Vec(X^t) - E[Vec(X^t)]] [Vec(X^t) - E[Vec(X^t)]]^t] =$$

$$= E \begin{bmatrix} X_1 - \mu \\ X_2 - \mu \\ \vdots \\ (X_1 - \mu)^t \quad \dots \quad (X_N - \mu)^t \\ \vdots \\ X_N - \mu \\ \mu \end{bmatrix} = \begin{pmatrix} \Sigma & 0 & \dots & 0 \\ 0 & \Sigma & \dots & 0 \\ \dots & \dots & \ddots & \dots \\ 0 & \dots & 0 & \Sigma \mu \end{pmatrix}$$

A raíz de la expresión obtenida en el ejemplo, vamos a definir el producto de Kronecker de matrices:

Definición 1.2. Sean $A_{m \times n}$ y $B_{p \times q}$ dos matrices. Se define el producto Kronecker de ellas como la matriz de dimensiones $mp \times nq$ siguiente

$$A \otimes B = \begin{pmatrix} a_{11}B & a_{12}B & \dots & a_{1n}B \\ a_{21}B & a_{22}B & \dots & a_{2n}B \\ \dots & \dots & \dots & \dots \\ a_{m1}B & a_{m2}B & \dots & a_{mn}B \end{pmatrix} = (a_{ij}B)_{ij}; i = 1, \dots, m, j = 1, \dots, n$$

Observamos la utilidad de la definición anterior obteniendo que en el ejemplo anterior se tiene $Cov[Vec(X^t)] = I_N \otimes \Sigma$.

Aunque hemos utilizado el resultado anterior para justificar la introducción del producto Kronecker, no es una justificación formal. Otra de las justificaciones que podríamos haber dado para introducirlo podría ser el resolver el sistema de ecuaciones $x = (A \otimes B)y$, que es no singular.

Ahora bien, una vez presentada esta nueva operación nos proponemos estudiar sus propiedades. Al igual que el producto usual matricial estaba relacionado con la composición de aplicaciones lineales, este nuevo producto lo estará con el producto tensorial. Vemos algunas propiedades en el siguiente teorema.

Teorema 1.2. *Se verifican las siguientes propiedades:*

1. Dadas $\alpha, \beta \in \mathbb{R}$, $A_{m \times n}$ y $B_{p \times q}$, entonces

$$(\alpha A) \otimes (\beta B) = \alpha\beta(A \otimes B) = \alpha\beta A \otimes B = A \otimes (\alpha\beta)B$$

2. Dadas $A_{m \times n}$, $B_{m \times n}$, $C_{p \times q}$ y $D_{p \times q}$, entonces

$$a) (A \otimes C) + (B \otimes C) = (A + B) \otimes C$$

$$b) (A \otimes C) + (A \otimes D) = A \otimes (C + D)$$

$$c) (A + B) \otimes (C + D) = (A \otimes C) + (A \otimes D) + (B \otimes C) + (B \otimes D)$$

3. Dadas $A_{m \times n}$, $B_{p \times q}$, $C_{r \times s}$, entonces $(A \otimes B) \otimes C = A \otimes (B \otimes C)$

4. Dadas $A_{m \times n}$, $B_{n \times p}$, $C_{q \times r}$ y $D_{r \times s}$, entonces $(A \otimes C)(B \otimes D) = AB \otimes CD$

5. Dadas $A_{m \times m}$ y $B_{n \times n}$ son no singulares entonces $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$

6. Si $A_{m \times n}$ y $B_{p \times q}$, entonces $(A \otimes B)^t = A^t \otimes B^t$

7. Si $A_{m \times m}$ y $B_{n \times n}$ son ortogonales, entonces $A \otimes B$ es ortogonal

8. Si $A_{m \times m}$ y $B_{n \times n}$ son matrices triangulares (inferiores), entonces $A \otimes B$ es triangular superior (inferior).

9. Si $A_{m \times m}$ y $B_{n \times n}$ son definidas positivas, entonces $A \otimes B$ es definida positiva

10. Dadas $A_{m \times n} = [A_1, \dots, A_k]$ y $B_{p \times q}$, entonces $A \otimes B = [A_1 \otimes B, \dots, A_k \otimes B]$. En particular, si $a_{m \times 1}$ y $b_{p \times 1}$ son dos vectores se tiene que $a \otimes b^t = ab^t = b^t \otimes a$

11. Dadas $A_{m \times n} = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}$ y $B_{p \times q}$, entonces $A \otimes B = \begin{pmatrix} A_{11} \otimes B & A_{12} \otimes B \\ A_{21} \otimes B & A_{22} \otimes B \end{pmatrix}$

12. Dadas $A_{m \times m}$ y $B_{n \times n}$, entonces $tr[A \otimes B] = tr[A]tr[B]$

13. Sean $A_{m \times m}$ y $B_{n \times n}$ matrices reales con autovalores reales respectivos $\lambda_1, \dots, \lambda_m$ y μ_1, \dots, μ_n . Entonces $A \otimes B$ tiene como autovalores $\lambda_i \mu_j, i = 1, \dots, m; j = 1, \dots, n$. Como consecuencia $\text{rg}(A \otimes B) = \text{rg}(A)\text{rg}(B)$
14. Dadas $A_{m \times m}$ y $B_{n \times n}$, entonces $|A \otimes B| = |A|^n |B|^m$

1.3.1. Producto Kronecker y Vec. Relaciones de interés entre ambas operaciones.

En este apartado vamos a exponer varias propiedades que relacionan las dos operaciones recientemente introducidas.

Teorema 1.3. *Se verifican las siguientes afirmaciones:*

- Si $a_{n \times 1}$ y $b_{q \times 1}$ son dos vectores, entonces $\text{Vec}(ab^t) = b \otimes a$
- Sean $A_{n \times q}$, $B_{q \times p}$ y $C_{p \times r}$ tres matrices cualesquiera. Entonces $\text{Vec}(ABC) = (C^t \otimes A)\text{Vec}(B)$
- Sean $A_{n \times q}$ y $B_{q \times n}$. Entonces $\text{tr}[AB] = \text{Vec}'(A^t)\text{Vec}(B) = \text{Vec}'(B^t)\text{Vec}(A)$
- Sea $\{e_i : i = 1, \dots, n\}$ la base canónica de \mathbb{R}^n . Entonces

$$\text{Vec}(I_n) = \sum_{i=1}^n (e_i \otimes e_i)$$

- Sea $\{J_{ij} : i = 1, \dots, n : j = 1, \dots, q\}$ la base canónica del espacio de matrices $\mathbb{M}_{n \times q}$. Entonces

$$\sum_{i=1}^n \sum_{j=1}^q (J_{ij} \otimes J_{ij}) = \text{Vec}(I_n)\text{Vec}'(I_q)$$

1.3.2. Ejercicios

1.4. La matriz conmutación

En esta sección vamos a introducir un tipo de matriz que toma diversos nombres dado a que ha aparecido a lo largo de la historia de diferentes maneras. Nosotros tomaremos la notación de *matriz conmutación*.

Una característica común de todas las introducciones de esta matriz es la propiedad de reordenar los elementos de una matriz. Una propiedad importante es la de transformar $\text{Vec}(A)$ en $\text{Vec}(A^t)$ siendo A una matriz arbitraria. También sirve para invertir un producto Kronecker, siendo una propiedad muy utilizada en la derivación matricial.

Definamos formalmente lo que venimos introduciendo:

Definición 1.3. *Sea $A_{n \times q}$. Se define la matriz de conmutación como la matriz de permutación, K_{nq} , de dimensión $nq \times nq$ tal que*

$$K_{nq}\text{Vec}(A) = \text{Vec}(A^t)$$

Notar que dado que K_{nq} es una matriz de permutación, es ortogonal, esto es $K_{nq}^t = K_{nq}^{-1}$. Además, claramente se tiene que $K_{qn}K_{nq}\text{Vec}(A) = K_{qn}\text{Vec}(A^t) = \text{Vec}(A)$, por lo que se obtiene que $K_{qn}K_{nq} = I_{nq}$, por lo que $K_{nq}^{-1} = K_{qn} = K_{nq}^t$. Además, es fácilmente comprobable que $K_{n1} = I_n$ y $K_{1q} = I_q$.

Para saber cómo se expresa de forma explícita la matriz de conmutación introducimos el siguiente teorema.

Teorema 1.4. Sea $\{J_{ij} : i = 1, \dots, n; j = 1, \dots, q\}$ la base canónica del espacio de matrices $\mathbb{M}_{n \times q}$. Entonces

$$K_{nq} = \sum_{i=1}^n \sum_{j=1}^q J_{ij} \otimes J_{ij}^t$$

Otra introducción de la matriz de conmutación es de forma constructiva, es decir, introduciéndola como aquella matriz $A_{n \times q}$ que verifica que

$$Vec(A^t) = \sum_{i=1}^n \sum_{j=1}^q (J_{ij} \otimes J_{ij}^t) Vec(A)$$

*****TENEMOS QUE PENSAR QUE EJEMPLOS PONEMOS *****

Tengamos en cuenta ahora que el resultado de $I_n \otimes I_q$ es I_{nq} . Así, podemos dar una nueva definición de la matriz de conmutación como la n-permutación por filas de la matriz $I_n \otimes I_q$, con lo que se asegura que $Vec(A^t) = K_{nq} Vec(A)$.

Notemos que con vistas a esta definición se puede usar la notación $I_{(q,n)}$ para hacer referencia a K_{nq} . Además se guarda cierta relación con K_{qn} que no deja de ser la q-permutación por filas de $I_n \otimes I_q$. Veamos ahora el siguiente teorema:

Teorema 1.5. Sean $\{e_i : i = 1 \dots n\}$ y $\{u_j : j = 1 \dots q\}$ las bases canónicas de \mathbb{R}^n y \mathbb{R}^q respectivamente. Entonces:

$$K_{nq} = \sum_{j=1}^q (u_j^t \otimes I_n \otimes u_j) = \sum_{i=1}^n e_i \otimes I_q \otimes e_i^t$$

Veamos ahora algunas propiedades de esta matriz:

Teorema 1.6. Consideramos las matrices $A_{m \times n}$, $B_{p \times q}$, $C_{q \times s}$, $D_{n \times t}$, $E_{m \times n}$ y los vectores $a_{m \times 1}$, $b_{p \times 1}$ y z un vector cualquiera. Entonces:

1. $K_{pm}(A \otimes B) = (B \otimes A)K_{qn}$
2. $z^t \otimes A \otimes B = K_{mp}(bz^t \otimes A)$.
3. $tr[K_{mn}(A^t \otimes E)] = tr[A^t E] = Vec'(A^t)K_{mn}Vec(E)$
4. $tr[K_{mn}] = 1 + f(m-1, n-1)$, donde:

$$f(m-1, n-1) = \begin{cases} m.c.d(m-1, n-1) \\ f(0, n) = f(n, 0) = n \end{cases}$$

5. Los autovalores de K_{nn} son 1 con multiplicidad $\frac{n(n+1)}{2}$ y -1 con multiplicidad $\frac{n(n-1)}{2}$.
6. $|K_{nn}| = (-1)^{\frac{n(n-1)}{2}}$ y $|K_{mn}| = (-1)^{\frac{m(m-1)n(n-1)}{4}}$
7. Supongo que $rg(A) = r$ y $\lambda_1 \dots \lambda_r$ los autovalores de $A^t A$. Sea $P = K_{mn}(A^t \otimes A)$. Entonces P es una matriz simétrica con rango r^2 y que cumple: $tr[P] = tr[A^t A]$. Además se verifica que $P^2 = (AA^t) \otimes (A^t A)$ y sus autovalores son los anteriores más $\pm(\lambda_i \lambda_j)^{\frac{1}{2}}$ con $i \neq j$.

Ahora podemos plantearnos como realizar la vectorización de un producto de Kronecker. Para ello mostramos el siguiente resultado:

Teorema 1.7. Sean $A_{m \times n}$, $B_{p \times q}$. Entonces $Vec(A \otimes B) = [I_n \otimes K_{qm} I_p][Vec(A) \otimes Vec(B)]$.

Vamos a llevar la matriz de conmutación un paso más adelante. Para ello vamos a definirla para más de dos índices $K_{st,n}$ como se hacía anteriormente pero tomando $st = m$. Su actuación se ve como:

$$A \otimes B \otimes C = K_{ms,p}(C \otimes A \otimes B)K_{q,nt} = K_{m,sp}(B \otimes C \otimes A)K_{tq,n}$$

. Así, podemos mostrar el siguiente teorema en el que se muestran algunas propiedades:

Teorema 1.8. Se cumple:

1. $K_{mn,p} = K_{nm,p}$ y $K_{m,np} = K_{m,pn}$
2. $K_{mn,p} = (I_m \otimes K_{np})(K_{mp} \otimes I_n) = (I_n \otimes K_{mp})(K_{np} \otimes I_m)$
3. $K_{m,np} = (K_{mn} \otimes I_p)(I_n \otimes K_{mp}) = (K_{mp} \otimes I_n)(I_p \otimes K_{mn})$
4. $K_{mn,p} = K_{m,np}K_{n,pm} = K_{n,mp}K_{m,pn}$ y $K_{m,np} = K_{mn,p}K_{pm,n} = K_{mp,n}K_{nm,p}$
5. $I_{nmp} = K_{mn,p}K_{pm,n}K_{np,m} = K_{mn,p}K_{pn,m}K_{mp,n} = K_{m,np}K_{n,pm}K_{p,mn} = K_{m,np}K_{p,nm}K_{n,mp}$
6. Cualquier producto de dos matrices de este tipo con el mismo conjunto de índices conmuta.

****FALTAN EJERCICIOS POR AQUI ****

1.5. Producto * de dos matrices

Para simplificar cálculos en las derivadas matriciales, se va a introducir un nuevo producto:

Definición 1.4. Sean $X_{m \times n}$ e $Y_{mp \times nq}$. Particionamos Y como sigue:

$$Y = \begin{bmatrix} Y_{11} & \cdots & \cdots & Y_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ Y_{m1} & \cdots & \cdots & Y_{mn} \end{bmatrix} \text{ donde } Y_{ij} \text{ es de dimensión } p \times q$$

Así, se define $X * Y = \sum_{i=1}^m \sum_{j=1}^n x_{ij} Y_{ij}$

Podemos ver así algunas propiedades de esta operación:

Teorema 1.9. Sean $X_{m \times n}$, $Y_{mp \times nq}$, $W_{mp \times nq}$, $Z_{s \times t}$. Se cumple:

1. $(X * Y)^t = X * Y^t$
2. Si $p = q = 1$, entonces $X * Y = \text{tr}[XY^t] = \text{tr}[YX^t] = \text{tr}[X^t Y] = \text{tr}[Y^t X]$
3. $(X * Y) \otimes Z = X * (Y \otimes Z)$
4. $X * (Y + W) = X * Y + X * W$

Y podemos ver más propiedades:

Teorema 1.10. Sean $X_{m \times n}$, $Y_{n \times p}$, $Z_{p \times q}$.

1. $XYZ = Y * \text{Vec}(X) * \text{Vec}'(Z^t)$
2. $XYZ = Y^t * (Z \otimes I_m) K_{qm}(X \otimes I_q)$

Ejercicios

EJERCICIO 1.1.: Sean $X_{m \times n}$, $Y_{n \times q}$. Demostrar que $Y * K_{np} = Y^t$

1.6. Operación Vech. Matrices de transición. La matriz duplicación

*****EN ESTE APARTADO HABIA ALGO MAL QUE ERA? *****

Vamos a realizar un planteamiento que nos lleve a un nuevo operador (**Vech**). Para ello, si consideramos $A_{n \times n}$, se puede ver $\text{Vech}(A)$ como el vector obtenido a partir de $\text{Vec}(A)$ eliminando los elementos de la diagonal superior de A. Por ejemplo si A es simétrica, entonces $\text{Vech}(A)$ contiene sus elementos distintos.

Es claro que se puede realizar una transformación que lleve de $\text{Vec}(A)$ a $\text{Vech}(A)$ mediante el producto por una matriz D_n . A esta matriz la llamaremos la **matriz de duplicación**.

Así se puede ver esto como un nuevo operador. Sea $A_{m \times n}$ una matriz con s elementos repetidos y/o v elementos constantes, el número de elementos matemáticamente independientes se puede calcular como sigue: $r = mn - s - v$.

Algunos ejemplos son los siguientes:

EJEMPLO 1.3: ■ Si A es simétrica, $v = 0$, $s = \frac{m(m-1)}{2}$ y por tanto $r = \frac{m(m+1)}{2}$.

- Si es antisimétrica, $v = m$, $s = \frac{m(m-1)}{2}$ y $r = \frac{m(m-1)}{2}$
- Si A es diagonal, $v = \frac{m(m-1)}{2}$, $s = 0$ y $r = m$ *****NO ME CUADRA ESO*
- Si A es una matriz de correlación, entonces $v = m$, $s = \frac{m(m-1)}{2}$ y $r = \frac{m(m-1)}{2}$

Así en estas matrices $Vech(A)$ se podría ver como el vector eliminando los elementos repetidos y/o constantes. De este modo, podemos establecer una relación entre $Vec(A)$ y $Vech(A)$, pues ordenan los elementos de la matriz A salvo excluidos. Esta relación se consigue a través de una matriz que se denomina **matriz de transición**.

2. Derivación matricial.

2.1. Introducción

En este apartado se va a tratar la derivación respecto a vectores y matrices, que es muy necesaria en estadística multivariante sobre todo desde el punto de vista de la optimización. Así, permite calcular datos tales como estimador máximo verosímil, matrices de información de Fisher, o cotas tipo Crámer-Rao. Más importancia tiene todavía este tema si tenemos en cuenta que, si ya la derivación vectorial puede dar lugar a cálculos costosos, en el caso de la matricial se pueden generar un enorme número de derivadas que pueden resultar difícil de ordenar con sentido en una matriz.

Muchas son las aproximaciones que se han dado y en este caso lo que se hará es seguir una línea en la que las matrices se conviertan en vectores, mucho más fáciles de manejar. *****

2.2. Diferencial primera y jacobianos

2.2.1. Diferencial de una función vectorial

Definición 2.2.1: Consideramos una función vectorial $f : S \rightarrow \mathbb{R}^m$ con $S \subset \mathbb{R}^n$. Sea \mathbf{c} un punto interior de S y consideremos una bola cerrada con centro en \mathbf{c} y radio \mathbf{r} , $B(\mathbf{c}, r)$. Sea u un punto de \mathbb{R}^n tal que $\|u\| \leq r$ es decir, $\mathbf{c} + r \in B(\mathbf{c}, r)$.

Diremos que f es **diferenciable** en \mathbf{c} si existe una matriz real de orden $m \times n$ que depende de c y no de u y que cumple que $f(\mathbf{c} + u) - f(\mathbf{c}) = A(\mathbf{c})u + r_c(u)$ con $\lim_{u \rightarrow 0} \frac{r_c(u)}{\|u\|} = 0$. Además, se define la **primera diferencial** de f en el punto c con incremento u como: $df(\mathbf{c}; u) = A(\mathbf{c})u$.

Definición 2.2.2: Sea $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ y sea $f_i : S \rightarrow \mathbb{R}$ su i -ésima componente. Sea c un punto interior de S y e_j el j -ésimo vector de la base canónica de \mathbb{R}^n . Se define la **derivada parcial** de f respecto a la j -ésima coordenada como:

$$D_j f_i(c) = \lim_{t \rightarrow 0} \frac{f_i(c + te_j) - f_i(c)}{t}, t \in \mathbb{R}$$

Teorema 2.1. (Primer teorema de identificación para funciones vectoriales)

Sea $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un punto c interior de S y $u \in \mathbb{R}^n$. Entonces $df(\mathbf{c}; u) = (Df(\mathbf{c}))u$

donde $Df(c)$ es una matriz $m \times n$ cuyos elementos $D_j f_i(c)$ son las derivadas parciales de \mathbf{f} evaluadas en c y que recibe el nombre de matriz jacobiana. Recíprocamente, si $A(c)$ es una matriz que verifica que $df(c; u) = A(c)u \forall u \in \mathbb{R}^n$, entonces $A(c) = Df(c)$.

El siguiente teorema nos proporciona la regla de la cadena para funciones vectoriales.

Teorema 2.2. Sea $f : S \subset \mathbb{R}^n \rightarrow \mathbb{R}^m$ diferenciable en un punto c interior de S . Sea T un subconjunto de \mathbb{R}^m tal que $f(x) \in T \forall x \in S$ y supongamos que $g : T \rightarrow \mathbb{R}^p$ es diferenciable en un punto \mathbf{b} ($\mathbf{b} = f(c)$) de T . Entonces la función compuesta $h : S \rightarrow \mathbb{R}^p$ definida por $h(x) = g(f(x))$, es diferenciable en c y $Dh(c) = Dg(\mathbf{b})Df(c)$.

Teorema 2.3. (Regla de invarianza de Cauchy)

En el ambiente del teorema anterior, si \mathbf{f} es diferenciable en \mathbf{c} y \mathbf{g} lo es en $\mathbf{b} = \mathbf{f}(\mathbf{c})$, entonces la diferenciable $h = g \circ f$ es $dh(c; u) = dg(\mathbf{b}; df(c; u))$.

2.2.2. Diferencial de una función matricial

Basta con extrapolar lo dicho en el caso vectorial usando la operación Vec .

Consideramos una función matricial $F : S \rightarrow \mathbb{M}_{m \times p}$ donde $S \subseteq \mathbb{M}_{n \times q}$. Sea C un punto interior de S , $\mathbb{B}(C; r) \subseteq S$ una bola abierta y U un punto de $\mathbb{M}_{n \times q}$ con $\|U\| < r$, por lo que $C + U$ pertenece a $\mathbb{B}(C; r)$, donde hemos considerado la norma matricial $\|U\| = (\text{tr}[U^t U])^{\frac{1}{2}}$.

Definición 2.1. En las condiciones anteriores, se dice que F es diferenciable en C si existe una matriz A de dimensiones $mp \times nq$, que dependa de C y no de U y tal que

$$\text{Vec}(F(C + U)) - \text{Vec}(F(C)) = A(C)\text{Vec}(U) + \text{Vec}(R_C(U)) \quad \text{donde} \quad \lim_{U \rightarrow 0} \frac{R_C(U)}{\|U\|} = 0$$

Se define la **matriz diferencial** de F en C con incremento U como la matriz $dF(C; U)$ de dimensiones $m \times p$ que verifique $\text{Vec}(dF(C; U)) = A(C)\text{Vec}(U)$ y a la matriz $A(C)$ se le llama la **primera derivada** de F en C .

Todas las propiedades de cálculo para las funciones matriciales se deducen de las correspondientes propiedades de las funciones vectoriales.

Tenemos los siguientes resultados análogos a los del caso vectorial:

Teorema 2.4. (Primer teorema de identificación para funciones matriciales) Sea $F : S \subseteq \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ diferenciable en un punto interior C de S . Entonces se verifica $\text{Vec}(dF(C; U)) = A(C)\text{Vec}(U) \Leftrightarrow DF(C) = A(C)$.

Teorema 2.5. (Regla de la cadena para funciones matriciales) Sea $F : S \subseteq \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ diferenciable en un punto interior C de S . Sea T un subconjunto de $\mathbb{M}_{r \times s}$ tal que $F(X) \in T, \forall X \in S$ y supongamos que $G : T \rightarrow \mathbb{M}_{r \times s}$ es diferenciable en un punto B ($B = F(C)$) de T . Entonces la función compuesta $H : S \rightarrow \mathbb{M}_{r \times s}$ definida por $H(X) = G(F(X))$ es diferenciable en C y $DH(C) = DG(B)DF(C)$.

Teorema 2.6. (Regla de invarianza de Cauchy para funciones matriciales) En las condiciones del teorema anterior, $dH(C; U) = dG(B; dF(C; U)), \forall U \in \mathbb{M}_{n \times q}$.

Corolario 2.7. Sean A_{txm}, B_{pxr} y C_{sxu} matrices constantes y sean $G_{m \times p}$ y $H_{r \times s}$ dos funciones matriciales diferenciables. Si $F(X) = AG(X)BH(X)C$, entonces

$$dF(X) = AdG(X)BH(X)C + AG(X)BdH(X)C$$

$$y \, DF(X) = (C^t H^t(X) B^t \otimes A) DG(X) + (C^t \otimes AG(X) B) DH(X)$$

*****COPIAMOS LAS MILDOSCIENTAS34 PROPIEDADES????

EJERCICIO 2.1: Sea $h : \mathbb{R}^k \rightarrow \mathbb{R}$ definida por $h(\beta) = (y - X\beta)^t(y - X\beta)$ donde $y \in \mathbb{R}^n$ y $X \in \mathbb{M}_{n \times k}$. Haciendo uso de la regla de invarianza de Cauchy demostrar que

$$dh(c; u) = dg(y - Xc; df(c; u)) = dg(y - Xc; -Xu) = -2(y - Xc)^t Xu$$

y con ello $Dh(c) = -2(y - Xc)^t X$.

Solución: Para comenzar, tengamos en cuenta que h es composición de $f : \mathbb{R}^k \rightarrow \mathbb{R}^n$ y $g : \mathbb{R}^n \rightarrow \mathbb{R}$, con $f(c) = (y - Xc)$ con X una matriz $n \times k$ y con $y \in \mathbb{R}^n$; y $g(X) = X^t X$, de modo que $h(x) = g(f(x)) \forall x \in \mathbb{R}^k$. Así, es claro que $dh(c; u) = dg(y - Xc; df(c; u))$. ***** Parece evidente pero no se que escribir xd*****

EJERCICIO 2.2: Sea $F(X) = AG(X)B$, donde $A_{m \times r}$ y $B_{s \times p}$ son matrices constantes y $G(X)_{r \times s}$ es una función diferenciable. Calcular $DF(C)$ a partir de la definición de diferencial matricial.

Solución:

EJERCICIO 2.3: Si $X_{n \times n}$ es una matriz simétrica y $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$ es diferenciable, demostrar que $d\text{Vec}(F(X)) = D_n DF(X) d\text{Vec}(X)$, mientras que $d\text{Vec}(F(X)) = N_n DF(X) d\text{Vec}(X)$ donde $N_n = \frac{1}{2}[I_{n^2} + K_{nn}]$.

Solución:

2.3. Matrices jacobianas y derivadas matriciales

No existe una única definición para la derivada de una función de argumento matricial y esto supone un problema a la hora de usar el cálculo diferencial matricial.

Veamos las definiciones clásicas para funciones reales de argumento vectorial y vectoriales, tanto de argumento real como vectorial.

Definición 2.2. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}$. Se define la derivada de f respecto de $x \in \mathbb{R}^n$ como el vector $1 \times n$ dado por $\frac{\partial f(x)}{\partial x^t} = \left(\frac{\partial f(x)}{\partial x_1}, \dots, \frac{\partial f(x)}{\partial x_n} \right)$.

Definición 2.3. Sea $f : \mathbb{R} \rightarrow \mathbb{R}^m$. Se define la derivada de f respecto de $x \in \mathbb{R}$ como el vector $m \times 1$ dado por $\frac{\partial f(x)}{\partial x} = \left(\frac{\partial f_1(x)}{\partial x}, \dots, \frac{\partial f_m(x)}{\partial x} \right)^t$.

Definición 2.4. Sea $f : \mathbb{R}^n \rightarrow \mathbb{R}^m$. Se define la derivada de f respecto de $x \in \mathbb{R}^n$ como la matriz $m \times n$

$$\frac{\partial f(x)}{\partial x^t} = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x^t} \\ \vdots \\ \frac{\partial f_m(x)}{\partial x^t} \end{pmatrix} = \begin{pmatrix} \frac{\partial f_1(x)}{\partial x_1} & \dots & \dots & \frac{\partial f_1(x)}{\partial x_n} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial f_m(x)}{\partial x_1} & \dots & \dots & \frac{\partial f_m(x)}{\partial x_n} \end{pmatrix}$$

Definición 2.5. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{R}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n \times q}$ como la matriz $n \times q$

$$\frac{\partial F(X)}{\partial X} = \begin{pmatrix} \frac{\partial F(X)}{\partial x_{11}} & \dots & \dots & \frac{\partial F(X)}{\partial x_{1q}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F(X)}{\partial x_{n1}} & \dots & \dots & \frac{\partial F(X)}{\partial x_{nq}} \end{pmatrix}$$

Definición 2.6. Sea $F : \mathbb{R} \rightarrow \mathbb{M}_{m \times p}$. Se define la derivada de F respecto de $x \in \mathbb{R}$ como la matriz $m \times p$

$$\frac{\partial F(x)}{\partial x} = \begin{pmatrix} \frac{\partial F_{11}(x)}{\partial x} & \dots & \dots & \frac{\partial F_{1p}(x)}{\partial x} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{m1}(x)}{\partial x} & \dots & \dots & \frac{\partial F_{mp}(x)}{\partial x} \end{pmatrix}$$

*****CREO que las importantes son solo las dos ultimas, quizas haya que borrar algunas de estas aunque eso se decidirá en un futuro xd *****

Existen así muchas formas de definir la derivada de una función de argumento matricial, aunque ahora vamos a mostrar dos que generalizan a las demás:

Definición 2.7. (Definición de derivada matricial según Mac Rae)

Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n \times q}$ como la matriz $nm \times pq$

$$\frac{\partial F(x)}{\partial x} = \begin{pmatrix} \frac{\partial F_{11}(x)}{\partial x} & \dots & \dots & \frac{\partial F_{1p}(x)}{\partial x} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F_{m1}(x)}{\partial x} & \dots & \dots & \frac{\partial F_{mp}(x)}{\partial x} \end{pmatrix}$$

Definición 2.8. (Definición de derivada matricial según Dwyer)

Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n \times q}$ como la matriz $nm \times pq$

$$\frac{\partial F(x)}{\partial x} = \begin{pmatrix} \frac{\partial F(x)}{\partial x_{11}} & \dots & \dots & \frac{\partial F(x)}{\partial x_{1q}} \\ \vdots & \vdots & \vdots & \vdots \\ \frac{\partial F(x)}{\partial x_{n1}} & \dots & \dots & \frac{\partial F(x)}{\partial x_{nq}} \end{pmatrix}$$

Teorema 2.8. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Entonces se verifica que $K_{nm} \frac{\partial F(x)}{\partial x} K_{pq} = \frac{\partial F(x)}{\partial x}$

Definición 2.9. (Definición de derivada matricial según Magnus y Neudecker)

Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Se define la derivada de F respecto de $X \in \mathbb{M}_{n \times q}$ como la matriz $mp \times nq$ dada por $DF(X) = \frac{\partial \text{Vec}(F(x))}{\partial \text{Vec}^t(x)}$

*****MILLONES Y MILLONES DE PROPIEDADES XD XD XD XD XD- MUERO *****

Teorema 2.9. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{m \times p}$. Entonces

1. $\frac{\partial F(x)}{\partial x^t} = \left(\frac{\partial (F(x))^t}{\partial x} \right)^t$
2. $\frac{\partial (F(x))^t}{\partial x^t} = \left(\frac{\partial F(x)}{\partial x} \right)^t$

Corolario 2.10. Sean A_{txm} , B_{pxr} y C_{sxu} matrices constantes y sean $G_{m \times p}$ y $H_{r \times s}$ dos funciones de argumento matricial $X_{n \times q}$. Si $F(X) = AG(X)BH(X)C$, entonces

$$\frac{\partial F(\partial x)}{x} = (A \otimes I_n) \frac{\partial G(x)}{\partial x} (BH(\partial x)C \otimes I_q) + (AG(x)B \otimes I_n) \frac{\partial H(x)}{\partial x} (C \otimes I_q)$$

EJERCICIO 2.4: A partir de las relaciones existentes entre la derivada matricial y la matriz jacobiana, verificar las siguientes expresiones:

a) Sea $X_{n \times n}$ y $F(X) = \text{tr}[X]$. Entonces $DF(X) = \text{Vec}^t(I_n)$.

b) Sea ahora $X_{n \times q}$ y $F(X) = X$. Entonces $DF(X) = I_q \otimes I_n = I_{nq}$.

EJERCICIO 2.5: Sea $X_{n \times q}$. Demostrar las siguientes igualdades:

a) $\frac{\partial X^t}{\partial X} = K_{qn}$.

b) $\frac{\partial X}{\partial X^t} = K_{nq}$.

c) $\frac{\partial X^t}{\partial X^t} = \text{Vec}(I_q)\text{Vec}^t(I_n)$.

EJERCICIO 2.6: Demostrar que si $X_{n \times n}$ es no singular entonces $\frac{\partial X^{-1}}{\partial X} = -\text{Vec}((X^{-1})^t)\text{Vec}^t(X^{-1})$.

Solución:

2.4. Derivadas matriciales de funciones escalares de un vector

Las principales funciones que suelen aparecer en la práctica son:

- Formas lineales: $\phi(x) = a^t x$
- Formas cuadráticas: $\phi(x) = x^t A x$

De las que se pueden sacar una serie de **propiedades**:

- $d\phi(x) = a^t dx \rightarrow D\phi(x) = \frac{\partial \phi(x)}{\partial x^t} = a^t$
- $d\phi(x) = x^t (A + A^t) dx \rightarrow D\phi(x) = \frac{\partial \phi(x)}{\partial x^t} = x^t (A + A^t)$

Y otras tomando f y g como dos funciones vectoriales del vector x :

- Si $\phi(x) = a^t f(x) \rightarrow D\phi(x) = a^t Df(x)$
- Si $\phi(x) = f(x)^t g(x) \rightarrow D\phi(x) = g(x)^t Df(x) + f(x)^t Dg(x)$
- Si $\phi(x) = x^t A f(x) \rightarrow D\phi(x) = f(x)^t A^t + x^t A Df(x)$
- Si $\phi(x) = f(x)^t A f(x) \rightarrow D\phi(x) = f(x)^t (A + A^t) Df(x)$
- Si $\phi(x) = f(x)^t A g(x) \rightarrow D\phi(x) = g(x)^t A^t Df(x) + f(x)^t A Dg(x)$
- Si $x = (x_1^t, x_2^t)^t$ y Si $\phi(x) = x_1^t A x_2 \rightarrow D\phi(x) = x^t \begin{pmatrix} 0 & A \\ A^t & 0 \end{pmatrix}$

2.4.1. Derivadas matriciales de funciones escalares de matrices

*****PUEDEN FALTAR MUCHAS PARTES XD*****

Derivadas matriciales asociadas a trazas

Teorema 2.11. Sea $F : \mathbb{M}_{n \times q} \rightarrow \mathbb{M}_{p \times p}$ y $\phi(x) = \text{tr}[F(x)]$. Entonces $d\phi(x) = \text{tr}[dF(x)]$, $D\phi(x) = \text{Vec}^t(I_p)DF(x)$ y

$$\frac{\partial \phi(x)}{\partial x} = I_p * \frac{\partial F(x)}{\partial x} = \sum_{l=1}^p \frac{\partial F_{ll}(x)}{\partial x}$$

Teorema 2.12. Sean $F_{p \times m}$, $G_{m \times p}$ dos funciones matriciales de $X_{n \times q}$. Entonces se verifica:

- $d\text{tr}[FG(x)] = \text{tr}[F(x)dG(x)] + \text{tr}[F(x)dG(x)]$
- $D\text{tr}[(FG)(x)] = \text{Vec}^t(G^t(x))DF(x) + \text{Vec}^t(F^t(x))DG(x)$
- $\frac{\partial \text{tr}[(FG)(x)]}{\partial x} = G^t(x) * \frac{\partial F(x)}{\partial x} + F^t(x) * \frac{\partial G(x)}{\partial x}$

2.5. Diferencial segunda y hessianos