

# Derivación Matricial

Fernando Lara  
Fundación Universitaria Konrad Lorenz

Director: Leonardo Jiménez Moscovitz  
Matemático  
Fundación Universitaria Konrad Lorenz

11 de junio de 2007

## Resumen

En este trabajo se hace una presentación de algunos aspectos teóricos y prácticos acerca de la derivación matricial. Este tema es especialmente importante, ya que las matrices no solo proveen una notación muy compacta para algunos desarrollos matemáticos, sino que además permiten una mayor generalización. Después de la presentación teórica, se expone un caso de aplicación a las redes neuronales, y se exponen algunas líneas de código en Matlab para la resolución de problemas que involucren la derivación matricial.

*This work develops a presentation of some theoretical and practical topics about matrix derivation. This area is specially important because matrix theory not only provides a compact notation for some mathematical developments, it also provides a greater generalization. After the theoretic presentation, this paper exposes an application to neural networks, and some code in Matlab to help in the solutions of problems involving matrix derivation.*

# Índice

|  |            |
|--|------------|
| <b>Introducción</b>  | <b>3</b>   |
| <b>1. Preliminares</b>   | <b>5</b>   |
| 1.1. Matrices . . . . .  | 5          |
| 1.2. Conceptos Básicos . . . . .   | 5          |
| 1.3. Matrices . . . . .  | 8          |
| 1.3.1. Operaciones con Matrices . . . . .  | 10         |
| 1.3.2. Matrices Especiales. . . . .  | 14         |
| <b>2. Derivación.</b>  | <b>23</b>  |
| 2.0.3. Interpretación de la derivada como la pendiente de una<br>tangente. . . . . | 24         |
| 2.0.4. Reglas de Derivacion. . . . .   | 25         |
| 2.0.5. Derivación Parcial . . . . .  | 33         |
| <b>3. Derivación Matricial</b>   | <b>35</b>  |
| 3.1. Funciones Matriciales . . . . .   | 35         |
| 3.1.1. Funciones de Variable Escalar . . . . .                                     | 35         |
| 3.1.2. Funciones de Variable Vectorial . . . . .                                   | 36         |
| 3.1.3. Funciones de Variable Matricial . . . . .                                   | 37         |
| 3.2. Otras Matrices Especiales . . . . .   | 37         |
| 3.3. Producto de Kronecker . . . . .   | 39         |
| 3.4. Fórmulas de Derivación Matricial. . . . .                                     | 49         |
| 3.5. Reglas de Derivación. . . . .   | 60         |
| 3.6. Propiedades de Derivación de Funciones Matriciales. . . . .                   | 70         |
| <b>4. Aplicaciones</b>   | <b>87</b>  |
| 4.1. Redes Neuronales . . . . .  | 87         |
| 4.1.1. Características del Algoritmo . . . . .                                     | 87         |
| 4.1.2. <b>Propagación hacia adelante.</b> . . . . .                                | 88         |
| 4.1.3. Propagación hacia atrás. . . . .  | 89         |
| 4.1.4. Resumen (Algoritmo) . . . . .   | 96         |
| 4.1.5. Propagación hacia adelante. . . . .   | 97         |
| 4.1.6. Propagación hacia atrás. . . . .  | 97         |
| 4.1.7. Propagación hacia atrás. . . . .  | 100        |
| 4.2. Derivadas Matriciales en Matlab. . . . .                                      | 101        |
| <b>5. Conclusiones</b>   | <b>106</b> |

## Introducción

El Álgebra matricial se ha convertido en los últimos años en una parte esencial de los conocimientos de matemáticas, necesarios en campos tan diversos como la ingeniería, la física, la pedagogía, la química, la sociología y en el campo científico, así como la estadística y la matemática pura. Este requerimiento refleja la importancia y la amplitud de sus aplicaciones.

Muchos de los temas tratados en los campos mencionados comparten métodos comunes para resolver algunos de los problemas específicos. Los programas de las materias de matemáticas han ido evolucionando para facilitar que el profesional domine las técnicas empleadas en el tratamiento de los problemas. Sin embargo, por razones obvias, la mayoría de los textos de matemáticas se han centrado en lo que es el núcleo básico de los programas de cálculo y álgebra. Por otra parte, asignaturas de corte matemático dirigidos a dotar de conocimientos adecuados para abordar problemas, con un nivel más riguroso, tienen carácter optativo. Con respecto a esas asignaturas, no es fácil encontrar manuales adecuados a sus contenidos específicos, lo que significa que un estudiante ha de confiar en las referencias bibliográficas recomendadas, que en múltiples ocasiones no se ajustan a sus necesidades de rigor de planteamiento, o bien, apuntes tomados en clase, con los conocidos inconvenientes que esta práctica conlleva. Como ejemplos, pensemos en temas como derivadas de vectores y matrices y derivadas de funciones matriciales.

Basado en la importancia que ha adquirido el cálculo matricial, el compendio que se ha proyectado, se presentará como suplemento a los textos sobre cálculo matricial, que entre otros, son bastante exigüos en nuestro medio.

El objetivo fundamental es fijar un criterio de derivación, indicar su relación con las aproximaciones vectoriales a la derivada y, a partir de ello, obtener resultados que permitan enfrentarse formalmente y con mayor comodidad al estudio de variables matriciales.

Antes de abordar el proceso de derivación matricial, se presenta un compendio con los elementos básicos del álgebra lineal, específicamente sobre matrices, y un compendio que incluye las propiedades básicas del cálculo diferencial, necesarios para comprender la manera como se aborda el estudio sobre cálculo matricial.

De la misma forma, dada la complejidad del problema, se hizo preciso incluir algunas nuevas operaciones entre matrices que completan el cálculo matricial tradicional; para estas nuevas operaciones, se incluyen conceptos y propiedades sobre producto de Kronecker, vectorización de matrices y se definen las matrices de permutación que se utilizan en algunos desarrollos.

En el presente trabajo se ha utilizado extensivamente la presentación de [Bar98]; este excelente libro es uno de los pocos escritos en español que tratan el tema de la derivación matricial. En particular, se ha utilizado mucho en la sección 3, donde se expone la presentación teórica y de las propiedades. Otros escritos que se han podido disponer tienen una presentación algo diferente; en particular, muchos de ellos utilizan el concepto de diferencial al trabajar con derivadas matriciales, lo que implica entre otros aspectos, cambios en la no-

tación.

En el capítulo 4, se presenta el desarrollo de un problema de redes neuronales que en su forma original es un aporte del Ing. Pervys Rengifo, profesor de la F.U.K.L., a quien se le agradece su amable colaboración. En este ejemplo se observa la aplicación de la derivación matricial en la solución de un problema de este tipo. Finalmente, se exponen algunas líneas de código en Matlab para ayudar en la resolución de problemas de derivación matricial o vectorial.

## 1. Preliminares

En este capítulo se exponen los conceptos y reglas más importantes que sirven de base a la comprensión de la diferenciación matricial, así como de los conceptos y procesos asociados. Por ello, se explican en primera instancia las definiciones y los conceptos básicos de matrices, así como las operaciones a las que se pueden someter estos objetos. Luego se presentan los tipos de matrices más importantes y las propiedades que éstos poseen.

### 1.1. Matrices

Aquí se introducirá el concepto de matriz y se presentarán ciertas operaciones algebraicas definidas sobre ellas. Las nociones expuestas serán la base para comprender los conceptos relativos al cálculo vectorial y matricial en general.

De la manera más general posible, se puede definir una matriz como una tabla o arreglo rectangular de objetos matemáticos que pueden someterse a las operaciones de suma y producto. Esto requiere por una parte, detallar cuales son esos objetos o elementos de la matriz que son de interés primordial para el matemático, y por otra parte especificar algunas de las propiedades asociadas con las operaciones sobre dichos elementos: éstos, junto con las operaciones de suma y producto, deben formar un cuerpo  $\mathbb{k}$  [Her86].

Los elementos de una matriz pueden ser o bien los números reales ( $\mathbb{k} = \mathbb{R}$ ) o bien los números complejos ( $\mathbb{k} = \mathbb{C}$ ); en estos caso, los elementos se llaman escalares. Pero los elementos de una matriz también pueden ser funciones definidas en los números reales o los complejos. En el presente trabajo se tratarán las matrices definidas sobre los reales, se definirá posteriormente la matriz de funciones.

El otro caso en que se considerarán las matrices y vectores es en la propia definición de las funciones. Para el presente trabajo el interés no radica en funciones de variable real por ejemplo, sino que la función puede ser de variable vectorial, o matricial y por otra parte su salida puede ser a su vez un valor escalar, vectorial o matricial.

### 1.2. Conceptos Básicos

En esta sección se expondrán los principales conceptos relacionados con la temática a tratar. Por una parte, se hace referencia a algunos conceptos del álgebra abstracta como es el concepto de cuerpo, y por otra parte se hace referencia al concepto de espacio vectorial. A través del concepto de cuerpo se describe el comportamiento de los elementos que conforman una matriz al realizar sobre ellos ciertas operaciones; el concepto de espacio vectorial es una generalización tanto de vectores en  $\mathbb{R}^n$  como del propio concepto de matriz.

**Definición 1.1 (Campo o Cuerpo)** *Un campo o cuerpo  $\mathbb{k}$  es una estructura algebraica conformada por un conjunto y dos operaciones definidas sobre los elementos del conjunto,  $\langle K, +, * \rangle$  que se comportan como un anillo conmutativo en el cual todo elemento diferente de 0 tiene inverso multiplicativo. [Her86].*

**Definición 1.2 (Espacio Vectorial)** *Un espacio vectorial definido sobre un cuerpo  $\mathbb{k}$ , es un conjunto  $V$  de objetos llamados vectores, junto con las operaciones cerradas de suma y producto por escalar, que cumplen ocho propiedades [San94]:*

*Si  $u, v, w \in V$  y  $\alpha, \beta$  son escalares cualesquiera, entonces*

1.  $(u + v) + w = u + (v + w)$
2.  $u + v = v + u$
3. Existe  $0 \in V$  tal que  $v + 0 = 0 + v = v$
4. Para cada  $v$  existe un vector  $-v$  tal que  $v + (-v) = (-v) + v = 0$
5.  $\alpha(u + v) = \alpha u + \alpha v$
6.  $(\alpha + \beta)u = \alpha u + \beta u$
7.  $(\alpha\beta)u = \alpha(\beta u)$
8.  $1 \cdot v = v$

**Definición 1.3 (Subespacios Vectoriales)** *Un subespacio  $U$  de un espacio vectorial  $V$  es un subconjunto de  $V$  que es en sí mismo, un espacio vectorial sobre el mismo cuerpo.*

Dentro de los espacios vectoriales más importantes, se tienen los espacios  $\mathbb{k} = \mathbb{R}^n$  así como  $\mathbb{k} = \mathbb{C}^n$ . Un vector en  $\mathbb{R}^n$  se puede expresar como  $v_1 = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n$ . Esto se puede generalizar a continuación.

**Definición 1.4 (Combinación Lineal)** *Sea  $V$  un espacio vectorial, y  $U = \{u_1, u_2, \dots, u_r\}$  un conjunto finito de vectores de  $V$ . Entonces un vector de la forma:*

$$v = \alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_r u_r = \sum_{i=1}^r \alpha_i u_i$$

*se llama una combinación lineal de vectores en  $U$ .*

Es importante considerar el caso en el que se generan todas las posibles combinaciones lineales con elementos de  $U$ , y con todos los posibles escalares  $\alpha_i \in \mathbb{k}$ .

**Definición 1.5 (Subespacio Generado)** *Sea  $V$  un espacio vectorial, y  $U = \{u_1, u_2, \dots, u_r\}$  un conjunto finito de vectores de  $V$ . El conjunto de todas las posibles combinaciones lineales de elementos de  $U$  se llama el subespacio generado por  $U$ , y se denota:*

$$S(U) = \{v | v = \sum_{i=1}^n \alpha_k u_i, \forall \alpha_k \in \mathbb{k}\}$$

Un concepto muy importante relacionado con estos vectores es el de dependencia o independencia lineal.

**Definición 1.6 (Dependencia Lineal)** Sea  $V$  un espacio vectorial, y  $U = \{u_1, u_2, \dots, u_r\}$  un conjunto finito de vectores de  $V$ . Entonces,  $U$  es linealmente **dependiente o ligado** si y solo si existen escalares  $\alpha_1, \alpha_2, \dots, \alpha_r$  no todos ceros, tal que:

$$\alpha_1 u_1 + \alpha_2 u_2 + \dots + \alpha_r u_r = 0$$

Se dice que  $U$  es linealmente **independiente o libre** si no es dependiente.

Un conjunto  $U$  linealmente independiente y de especial importancia, se define a continuación.

**Definición 1.7 (Base)** Sea  $V$  un espacio vectorial sobre  $\mathbb{k}$ . Si en  $V$  existe un subconjunto finito  $U$  de vectores linealmente independientes, tal que  $S(U) = V$ , o en otras palabras, que todo vector  $v \in V$  puede ser expresado como combinación lineal de vectores en  $U$ , entonces se dice que  $U$  es una base de  $V$ .

Si una base consiste de una cantidad finita de elementos, entonces se puede demostrar que toda base de un espacio vectorial  $V$  tiene la misma cantidad de elementos:

**Definición 1.8 (Dimensión)** Sea  $V$  un espacio vectorial con una base  $B$  con  $n$  elementos. Entonces, se define a  $n$  como la dimensión del espacio vectorial  $V$ , y se expresa como  $\dim(V) = n$ .

Para el caso  $V = \mathbb{R}^n$  con  $n = 3$  se tiene la dimensión del conjunto de vectores en el espacio.

Es necesario en muchos casos considerar funciones o morfismos entre espacios vectoriales diferentes, esto es,  $f : V \rightarrow W$  donde  $V, W$  son espacios vectoriales sobre el mismo cuerpo  $\mathbb{k}$ . Estas funciones, cuya descripción se puede encontrar en libros de álgebra abstracta tales como [Her86] son de especial importancia si poseen ciertas propiedades, tal como se observa en la siguiente definición.

**Definición 1.9 (Isomorfismo)** Sean  $V, W$  son espacios vectoriales sobre el mismo cuerpo  $\mathbb{k}$ , y sea  $f$  una función  $f : V \rightarrow W$  que cumple con las siguientes propiedades que para todo  $x_i \in V$  y para todo  $a, b \in \mathbb{k}$ :

1.  $f^{-1}$  existe. Esto es,  $f$  es invertible
2.  $f(ax_1 + bx_2) = f(ax_1) + f(bx_2) = af(x_1) + bf(x_2)$

A los espacios vectoriales  $V, W$  se les llama espacios isomorfos, es decir, que tienen la misma estructura.

**Definición 1.10 (Transformación Lineal)** Sea  $U$  un espacio vectorial  $n$ -dimensional y sea  $V$  un espacio vectorial  $m$ -dimensional, ambos sobre el mismo cuerpo  $\mathbb{K}$ . Sea  $B_U$  una base de  $U$  y  $B_V$  una base de  $V$ . Una transformación lineal es una función  $T : U \rightarrow V$  tal que, para escalares arbitrarios  $a, b \in \mathbb{K}$  y vectores  $x_1, x_2 \in U$ :

$$T(ax_1 + bx_2) = T(ax_1) + T(bx_2) = aT(x_1) + bT(x_2)$$

Una transformación lineal es entonces, una operación entre dos espacios vectoriales que preserva las operaciones. Son de especial importancia también las transformaciones lineales  $T : V \rightarrow V$  que reciben el nombre de **endomorfismos**.

### 1.3. Matrices

En esta sección se establecerán las condiciones para definir adecuadamente el concepto de matriz, sus diferentes tipos y examinar luego algunas de sus propiedades.

En el presente trabajo, las matrices se establecerán con referencia al espacio vectorial definido por  $M_{mn}$ , el conjunto de todas las matrices de tamaño  $m \times n$ , así:

**Definición 1.11 (Matriz)** Sea  $\mathbb{K}$  un cuerpo, y sea  $M_{mn}$  el espacio vectorial de todas las matrices de tamaño  $m \times n$  definidas sobre  $\mathbb{K}$ . Se define una matriz  $A \in M_{mn}$  con elementos en  $\mathbb{K}$ , y se expresa diciendo que  **$A$  es una matriz sobre  $\mathbb{K}$** , a un arreglo de  $mn$  elementos  $a_{ij} \in \mathbb{K}$ , ordenados de manera rectangular con  $m$  filas y  $n$  columnas encerrados entre corchetes, donde  $1 \leq i \leq m$  y  $1 \leq j \leq n$ :

$$\begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & \dots & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

Las líneas horizontales en una matriz se denominan **filas** y las líneas verticales se denominan **columnas**. El primero de los coeficientes indica la fila y el segundo la columna en que está ubicado el elemento. A una matriz con  $m$  filas y  $n$  columnas se le denomina matriz  $m$ -por- $n$  (escrito  $m \times n$ ), siendo los valores  $m$  y  $n$  sus **dimensiones**. La matriz anterior se denota también por  $(a_{ij})$ .

Bajo esta definición, entonces sobre las matrices se pueden definir dos operaciones: suma y producto por escalar, que deben cumplir con las propiedades mencionadas en la definición 1.2.

**Definición 1.12 (Matriz como Transformación Lineal)** Sean  $V, W$  espacios vectoriales de dimensión finita. Si se han determinado las bases  $B_V$  del espacio  $V$  con dimensión  $n$ , y  $B_W$  del espacio  $W$  con dimensión  $m$ , entonces cada transformación lineal  $T : V \rightarrow W$  puede ser representada por una matriz  $A$ .



**Ejemplo 1.1** Sean  $V, W$  espacios vectoriales. Sea  $\{v_1, v_2, \dots, v_n\}$  una base de  $V$  y  $\{w_1, w_2, \dots, w_m\}$  una base de  $W$ . Cada vector  $v \in V$  puede ser representado de manera única mediante los escalares  $\alpha_1, \alpha_2, \dots, \alpha_n \in \mathbb{k}$ . mediante la combinación lineal  $\alpha_1 v_1 + \alpha_2 v_2 + \dots + \alpha_n v_n$

Si  $f : V \rightarrow W$  es una aplicación lineal, entonces

$$f(a_1 v_1 + a_2 v_2 + \dots + a_n v_n) = a_1 f(v_1) + a_2 f(v_2) + \dots + a_n f(v_n)$$

La función queda completamente determinada por los valores  $f(v_1) + \dots + f(v_n)$ . Como  $\{w_1, w_2, \dots, w_m\}$  es una base de  $W$ , entonces cada  $f(v_j)$  puede representarse como:

$$f(v_j) = a_{1j} f(w_1) + a_{2j} f(w_2) + \dots + a_{mj} f(w_m)$$

Luego la función queda completamente determinada por los valores  $a_{ij}$ . Si estos valores se utilizan para construir una matriz  $A$  de tamaño  $m \times n$ , Entonces la matriz  $A$  se puede utilizar para calcular el valor de  $f(v)$  para todo vector en  $V$ .

Otra manera de representar la interpretación de una matriz como una aplicación lineal es tomando  $x \in \mathbb{k}^n$ ,  $y \in \mathbb{k}^m$ , luego el morfismo  $f : \mathbb{k}^n \rightarrow \mathbb{k}^m$  puede ser representado por  $y = Ax$ .

Las dimensiones de una matriz siempre se dan con el número de filas primero y el número de columnas después. La entrada de una matriz  $A$  que se encuentra en la fila  $i$ -ésima y la columna  $j$ -ésima se le llama entrada  $i, j$  o entrada  $(i, j)$ -ésima de  $A$ . Esto se escribe como  $A_{i,j}$  o  $A[i, j]$ . Entonces, el primero de los subíndices indica la fila y el segundo la columna a las que pertenece dicho elemento. Las filas de la matriz son las  $m$   $n$ -plas horizontales:

$$(a_{11} \ a_{12} \ \dots \ a_{1n}), (a_{21} \ a_{22} \ \dots \ a_{2n}), \dots, (a_{m1} \ a_{m2} \ \dots \ a_{mn})$$

y las columnas de la matriz son las  $n$   $m$ -plas verticales:

$$\begin{pmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{pmatrix}, \begin{pmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{pmatrix}, \dots, \begin{pmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{pmatrix}$$

Nótese que el elemento  $a_{ij}$  llamado entrada  $ij$ , aparece en la fila  $i$ -ésima y en la columna  $j$ -ésima. Una matriz con  $m$  filas y  $n$  columnas se denomina matriz  $m$  por  $n$ , o matriz  $m \times n$ ; el par de números  $(m, n)$  se llama su **tamaño o forma**.

Las matrices se denotan usualmente por letras mayúsculas  $A, B, \dots$  y los elementos del cuerpo  $\mathbb{k}$  por minúsculas,  $a, b, \dots$  con sus respectivos subíndices.

Una matriz con una sola columna o una sola fila se denomina a menudo **vector**, y se interpreta como un elemento del espacio euclídeo [Her86].

**Definición 1.13 (Vector)** *Un vector es una matriz que posee una sola fila o una sola columna. Una matriz  $1 \times n$  (una fila y  $n$  columnas) se denomina **vector fila**, y una matriz  $m \times 1$  (una columna y  $m$  filas) se denomina **vector columna**. En general, dada una matriz  $A$ , al vector  $(a_{i1}, a_{i2}, \dots, a_{in})$  que forma la  $i$ -ésima fila de  $A$  se le llama un vector fila, mientras que al vector  $(a_{1j}, a_{2j}, \dots, a_{mj})$  que forma la  $j$ -ésima columna de  $A$  se le llama vector columna. Nótese que en particular, un elemento del cuerpo  $\mathbb{K}$  puede verse como una matriz o vector de tamaño  $1 \times 1$ .*

**Ejemplo 1.2** *Se tiene la matriz  $A_{2 \times 3}$ :*

$$\begin{bmatrix} 1 & -3 & 4 \\ 0 & 5 & -2 \end{bmatrix}$$

*Sus filas son  $(1 \ -3 \ 4)$  y  $(0 \ 5 \ -2)$  y cada una de ellas es un vector fila, mientras que*

*sus columnas son  $\begin{pmatrix} 1 \\ 0 \end{pmatrix}, \begin{pmatrix} -3 \\ 5 \end{pmatrix}, \begin{pmatrix} 4 \\ -2 \end{pmatrix}$  y cada una de ellas es un vector columna.*

**Ejemplo 1.3** *La aserción  $\begin{bmatrix} x+y & 2z+w \\ x-y & z-w \end{bmatrix} = \begin{bmatrix} 3 & 5 \\ 1 & 4 \end{bmatrix}$  es equivalente al sigui-*

*ente sistema de ecuaciones:*

$$\begin{aligned} x + y &= 3 \\ x - y &= 1 \\ 2z + w &= 5 \\ z - w &= 4 \end{aligned}$$

**Definición 1.14 (Igualdad de Matrices)** *La condición necesaria y suficiente para que dos matrices  $A = (a_{ij})$  y  $B = (b_{ij})$  sean iguales ( $A = B$ ) es que tengan el mismo orden y que cada uno de los elementos de una de ellas sea igual al correspondiente de la otra. Esto es*

$$a_{ij} = b_{ij} \quad (i = 1, 2, \dots, m; \ j = 1, 2, \dots, n)$$

En otras palabras, dos matrices son iguales solo cuando una es copia de la otra.

### 1.3.1. Operaciones con Matrices

Las matrices se han definido dentro del espacio vectorial que se ha llamado el espacio vectorial  $M_{mn}$  de las matrices de tamaño  $m \times n$ . Por tanto, las dos primeras operaciones que son de interés son la suma y el producto por escalar. Dichas operaciones deben dar como resultado otra matriz del mismo tamaño. Además, para estas operaciones, más adelante, (def 1.19) se definirá la matriz  $0_{m \times n}$  tal que para cualquier matriz  $A_{m \times n}$ ,  $A + 0 = 0 + A = A$  y obviamente, se tendrá un escalar 1 tal que  $1A = A$

**Definición 1.15 (Suma de Matrices)** Sean  $A = (a_{ij})$  y  $B = (b_{ij})$  dos matrices del espacio vectorial  $M_{mn}$ , ambas de igual tamaño  $m \times n$ . La suma (o diferencia) de ellas, denotada como  $A \pm B$ , es otra matriz  $C = (c_{ij})$  de orden  $m \times n$ , en la cual cada elemento de  $C$  es la suma (o diferencia) de los elementos correspondientes de  $A$  y  $B$ . Esto es,  $(c_{ij}) = (a_{ij} + b_{ij})$

$$A = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}; B = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{m1} & b_{m2} & \cdots & b_{mn} \end{bmatrix}$$

$$C = A + B = \begin{bmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & a_{m2} + b_{m2} & \cdots & a_{mn} + b_{mn} \end{bmatrix}$$

Dos matrices del mismo orden se llaman **conformes respecto de la suma algebraica**. La suma de matrices no está definida para matrices no conformes.

**Definición 1.16 (Producto por Escalar)** Sea  $A$  una matriz del espacio vectorial  $M_{mn}$ , y  $k \in \mathbb{K}$  un escalar. El producto del escalar  $k$  por la matriz  $A$ , escrito  $kA$  es la matriz  $C$  obtenida multiplicando cada entrada de  $A$  por  $k$ :

$$C = kA = \begin{bmatrix} ka_{11} & ka_{12} & \cdots & ka_{1n} \\ ka_{21} & ka_{22} & \cdots & ka_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ ka_{m1} & ka_{m2} & \cdots & ka_{mn} \end{bmatrix}$$

Obsérvese que  $A + B$  y  $kA$  son también matrices de tamaño  $m \times n$ . Además se define:

$$-A = -1 \cdot A \text{ y } A - B = A + (-B)$$

**Ejemplo 1.4** Sean  $A = \begin{bmatrix} 1 & -2 & 3 \\ 4 & 5 & -6 \end{bmatrix}$ ,  $B = \begin{bmatrix} 3 & 0 & 2 \\ -7 & 1 & 8 \end{bmatrix}$  entonces

$$A + B = \begin{bmatrix} 1+3 & -2+0 & 3+2 \\ 4-7 & 5+1 & -6+8 \end{bmatrix} = \begin{bmatrix} 4 & -2 & 5 \\ -3 & 6 & 2 \end{bmatrix}$$

$$3A = \begin{bmatrix} 3x1 & 3x(-2) & 3x3 \\ 3x4 & 3x5 & 3x(-6) \end{bmatrix} = \begin{bmatrix} 3 & -6 & 9 \\ 12 & 15 & -18 \end{bmatrix}$$

$$2A - 3B = \begin{bmatrix} 2 & -4 & 6 \\ 8 & 10 & -12 \end{bmatrix} + \begin{bmatrix} -9 & 0 & -6 \\ 21 & -3 & -24 \end{bmatrix} = \begin{bmatrix} -7 & -4 & 0 \\ 29 & 7 & -36 \end{bmatrix}$$

**Proposición 1.1 (Propiedades de la Suma y el Producto por Escalar)**

Sea  $V$  el conjunto de todas las matrices  $m \times n$  sobre un campo  $\mathbb{K}$ . En tal caso, para matrices arbitrarias  $A, B, C, 0 \in V$  y escalares cualesquiera  $k_1, k_2 \in \mathbb{K}$  y los valores  $0, 1 \in \mathbb{K}$  se cumple:

1.  $(A + B) + C = A + (B + C)$
2.  $A + 0 = A$
3.  $A + (-A) = 0$
4.  $A + B = B + A$
5.  $k_1(A + B) = k_1A + k_1B$
6.  $(k_1 + k_2)A = k_1A + k_2A$
7.  $(k_1k_2)A = k_1(k_2A)$
8.  $1 \cdot A = A$  y  $0 \cdot A = 0$

Estas propiedades son de esperarse, dado que las matrices pertenecen a un espacio vectorial, y  $k$  es un escalar.

Si se supone que los vectores en  $R^n$  se representan por vectores fila se tiene

$$u = [a_1 \quad a_2 \quad \cdots \quad a_n] \quad v = [b_1 \quad b_2 \quad \cdots \quad b_n]$$

Entonces, vistos como matrices, la suma  $u + v$  y el producto  $ku$  son las siguientes:

$$u + v = [a_1 + b_1 \quad a_2 + b_2 \quad \cdots \quad a_n + b_n] \quad ku = [ka_1 \quad ka_2 \quad \cdots \quad ka_n]$$

**Definición 1.17 (Producto de vector fila por vector columna)** Sea  $A = (a_i)$  un vector fila, y  $B = (b_i)$  un vector columna con el mismo número de elementos. El producto de estos vectores  $A$  y  $B$ , escrito  $A \times B$  es un vector  $c$  de tamaño  $1 \times 1$  definido como:

$$[a_1 \quad a_2 \quad \cdots \quad a_n] \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} = a_1b_1 + a_2b_2 + \cdots + a_nb_n = \sum_{k=1}^n a_kb_k = c$$

Nótese que para este caso, el producto de vectores se puede ver como un escalar  $c \in \mathbb{k}$ . El producto  $A \times B$  no está definido si los vectores  $A$  y  $B$  tienen diferente cantidad de elementos.

**Ejemplo 1.5**  $\begin{bmatrix} 8 & -4 & 5 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ -1 \end{bmatrix} = 8 \cdot 3 + (-4) \cdot 2 + 5 \cdot (-1) = 24 - 8 - 5 = 11$

Ahora se puede ampliar este procedimiento para realizar el producto de matrices de tamaño  $m \times n$  en general.

**Definición 1.18 (Producto de Matrices)** Sean  $A = (a_{ij})$  y  $B = (b_{ij})$  matrices tales que el número de las columnas de  $A$  coincide con el mismo número de filas de  $B$ ; esto es,  $A$  es una matriz  $m \times p$  y  $B$  es una matriz  $p \times n$ . Entonces el producto de las dos matrices  $AB$ , en este orden, es otra matriz  $C$  de tamaño  $m \times n$  cuya entrada  $ij$  se obtiene multiplicando la fila  $i$ -ésima  $A_i$  de  $A$  por la columna  $j$ -ésima  $B_j$  de  $B$  (cada elemento de la fila se multiplica por el correspondiente de la columna y a continuación se suman los productos obtenidos según la definición 1.17):

$$C = A_{m \times p} B_{p \times n} = \begin{bmatrix} A_1 B^1 & A_1 B^2 & \cdots & A_1 B^n \\ A_2 B^1 & A_2 B^2 & \cdots & A_2 B^n \\ \vdots & \vdots & \vdots & \vdots \\ A_m B^1 & A_m B^2 & \cdots & A_m B^n \end{bmatrix}_{m \times n}$$

Donde  $c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{ip}b_{pj} = \sum_{k=1}^p a_{ik}b_{kj}$  para todo  $i, j$  en  $C$ .

Se observa que en general, el producto de matrices no es conmutativo.

Dos matrices en las cuales el número de columnas de  $A$  es igual al número de filas de  $B$  se llaman **matrices conformes respecto de la multiplicación**. Se debe tener en cuenta el hecho de que el producto  $AB$  sólo está definido para matrices conformes. Esto es, no está definido si  $A$  es una matriz  $m \times p$  y  $B$  una matriz  $q \times n$  con  $p \neq q$ .

**Ejemplo 1.6**  $\begin{bmatrix} r & s \\ t & u \end{bmatrix} \begin{bmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \end{bmatrix} = \begin{bmatrix} ra_1 + sb_1 & ra_2 + sb_2 & ra_3 + sb_3 \\ ta_1 + ub_1 & ta_2 + ub_2 & ta_3 + ub_3 \end{bmatrix}$

$$\begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 1 \times 1 + 2 \times 0 & 1 \times 1 + 2 \times 2 \\ 3 \times 1 + 4 \times 0 & 3 \times 1 + 4 \times 2 \end{bmatrix} = \begin{bmatrix} 1 & 5 \\ 3 & 11 \end{bmatrix}$$

$$\begin{bmatrix} 1 & 1 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} = \begin{bmatrix} 1 \times 1 + 1 \times 3 & 1 \times 2 + 1 \times 4 \\ 0 \times 1 + 2 \times 3 & 0 \times 2 + 2 \times 4 \end{bmatrix} = \begin{bmatrix} 4 & 6 \\ 6 & 8 \end{bmatrix}$$

Lo anterior muestra que el producto de matrices no es conmutativo, es decir los productos  $AB$  y  $BA$  de matrices no son necesariamente iguales.

**Proposición 1.2 (Propiedades del Producto de Matrices)** Sean  $A, B, C$  matrices conformes para la multiplicación o para la suma, según corresponda, y  $k$  un escalar. Entonces, se cumplen las siguientes propiedades:

1.  $(AB)C = A(BC)$
2.  $A(B + C) = AB + AC$
3.  $(B + C)A = BA + CA$
4.  $k(AB) = (kA)B = A(kB)$
5.  $0A = 0$  y  $B0 = 0$  donde  $0$  es la matriz nula conforme

### 1.3.2. Matrices Especiales.

**Definición 1.19 (Matriz Nula)** Una matriz que tenga nulos todos sus elementos se llama **matriz nula** o **matriz cero** y se denota por  $0_{m,n}$ . En el caso de que una matriz  $A$  sea nula y no haya lugar a confusiones con respecto a su orden, se escribe simplemente  $A = 0$  en lugar de la disposición  $m \times n$  con sus elementos iguales a cero.

**Ejemplo 1.7** La matriz cero de tamaño  $2 \times 3$  es:  $0_{2 \times 3} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$

Para cualquier matriz  $A_{m,n}$  y  $0_{m,n}$ , se tiene que  $A + 0 = 0 + A = A$ . Esto es de esperarse, ya que  $0$  cumple la función de vector nulo en el espacio vectorial  $M_{mn}$ .

**Definición 1.20 (Matriz Traspuesta)** La matriz traspuesta de una matriz  $A$  de orden  $m \times n$  es la matriz  $A^t$  de orden  $n \times m$  llamada la **traspuesta** de  $A$  que se obtiene permutando las filas por las columnas. Por tanto, el elemento  $a_{ij}$  de  $A$  corresponde al elemento  $a_{ji}$  de  $A^t$

**Ejemplo 1.8** La traspuesta de la matriz  $A = \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{bmatrix}$  es  $A^t = \begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix}$

Obsérvese que el elemento  $a_{ij}$  de  $A$  (fila  $i$ , columna  $j$ ) es  $a_{ji}$  de  $A^t$  (fila  $j$ , columna  $i$ ).

**Proposición 1.3** Sean  $A^t$  y  $B^t$ , respectivamente, las traspuestas de las matrices  $A$  y  $B$ , y sea  $k \in \mathbb{K}$  un escalar; en estas condiciones se cumple que:

1.  $(A^t)^t = A$

$$2. (kA)' = kA$$

$$3. (A + B)' = A' + B'$$

$$4. (A \times B)' = B' \times A'$$

**Ejemplo 1.9**  $A = \begin{bmatrix} 4 & -3 & 6 \\ -2 & 5 & -1 \\ 1 & 0 & 3 \end{bmatrix}, B = \begin{bmatrix} 7 & 8 & 9 \\ 3 & 2 & 1 \\ -5 & 6 & -4 \end{bmatrix}$

$$A + B = \begin{bmatrix} 11 & 5 & 15 \\ 1 & 7 & 0 \\ -4 & 6 & -1 \end{bmatrix}, (A + B)' = \begin{bmatrix} 11 & 1 & -4 \\ 5 & 7 & 6 \\ 15 & 0 & -1 \end{bmatrix}$$

$$A' = \begin{bmatrix} 4 & -2 & 1 \\ -3 & 5 & 0 \\ 6 & -1 & 3 \end{bmatrix}, B' = \begin{bmatrix} 7 & 3 & -5 \\ 8 & 2 & 6 \\ 9 & 1 & -4 \end{bmatrix}, A' + B' = \begin{bmatrix} 11 & 1 & -4 \\ 5 & 7 & 6 \\ 15 & 0 & -1 \end{bmatrix}$$

$$\text{Por tanto } (A + B)' = \begin{bmatrix} 11 & 1 & -4 \\ 5 & 7 & 6 \\ 15 & 0 & -1 \end{bmatrix} = A' + B' = \begin{bmatrix} 11 & 1 & -4 \\ 5 & 7 & 6 \\ 15 & 0 & -1 \end{bmatrix}$$

**Matrices Cuadradas.** Dentro de las matrices más utilizadas en álgebra, se encuentran las matrices cuadradas.

**Definición 1.21 (Matriz Cuadrada)** Una matriz  $A$  se llama matriz cuadrada si el número de filas es igual al número columnas. Es decir  $n = m$ . Se dice que una matriz cuadrada  $n \times n$  es de orden  $n$  y se le asigna el nombre de matriz  $n$ -cuadrada:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \cdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix}$$

Las matrices cuadradas son de especial importancia. Es en estas matrices donde surge el concepto de determinante y de traza. Además poseen algunas propiedades interesantes, por ejemplo toda matriz cuadrada se puede descomponer en la suma de una matriz simétrica y una matriz antisimétrica. Además, si  $A$  y  $B$  son matrices del mismo orden, entonces se pueden sumar entre sí y sus productos son válidos en ambos sentidos, esto es, tanto  $AB$  como  $BA$  están definidos.

**Definición 1.22 (Traza de una Matriz Cuadrada)** Sea  $A$  una matriz cuadrada de orden  $n$ . La diagonal principal es la línea formada por los elementos  $a_{11}, a_{22}, \dots, a_{nn}$ . La suma de los elementos de la diagonal principal de una matriz cuadrada  $A$ , se llama **traza** de la misma, y se denota como  $\text{tr}(A) = \sum_{i=1}^n a_{ii}$ .

**Ejemplo 1.10** Sean  $A = \begin{bmatrix} 1 & 2 & 3 \\ -4 & -4 & -4 \\ 5 & 6 & 7 \end{bmatrix}$   $B = \begin{bmatrix} 2 & -5 & 1 \\ 0 & 3 & -2 \\ 1 & 2 & -4 \end{bmatrix}$

Entonces  $A$  y  $B$  son matrices cuadradas de orden 3, y sus trazas son  $\text{tr}(A) = 4$  y  $\text{tr}(B) = 1$ .

**Definición 1.23 (Matriz Triangular)** Una matriz cuadrada  $A$  cuyos elementos  $a_{ij} = 0$  para  $i \geq j$  se llama triangular superior; una matriz cuadrada cuyos elementos  $a_{ij} = 0$  para  $i \leq j$  se denomina triangular inferior. Así, pues

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} & \cdots & a_{1n} \\ 0 & a_{22} & a_{23} & \cdots & a_{2n} \\ 0 & 0 & a_{33} & \cdots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix} \text{ es una matriz triangular superior}$$

$$\begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ a_{21} & a_{22} & 0 & \cdots & 0 \\ a_{31} & a_{32} & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & a_{n3} & \cdots & a_{nn} \end{bmatrix} \text{ es una matriz triangular inferior.}$$

**Definición 1.24 (Matriz Diagonal)** La matriz  $D$  que es **triangular superior e inferior**, se llama **matriz diagonal**. Se representa por:

$$D = \text{diag} [a_{11} \quad a_{22} \quad a_{33} \quad \cdots \quad a_{nn}] = \begin{bmatrix} a_{11} & 0 & 0 & \cdots & 0 \\ 0 & a_{22} & 0 & \cdots & 0 \\ 0 & 0 & a_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_{nn} \end{bmatrix}$$

El producto  $AB$  de una matriz diagonal cuadrada de orden  $m$ ,  $A = \text{diag}(a_{11}, a_{22}, \dots, a_{mm})$  por otra matriz cualquiera  $B$  de orden  $n \times m$  se obtiene multiplicando la primera fila de  $B$  por  $a_{11}$ , la segunda de  $B$  por  $a_{22}$  y así sucesivamente. Esto es:



$$\begin{bmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{n1} & b_{n2} & \cdots & b_{nn} \end{bmatrix} \\
= \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1n} \\ a_{22}b_{21} & a_{22}b_{22} & \cdots & a_{22}b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{nn}b_{n1} & a_{nn}b_{n2} & \cdots & a_{nn}b_{nn} \end{bmatrix}$$

**Definición 1.25 (Matriz Escalar y Matriz Unidad)** Si en una matriz diagonal  $D$  se verifica que  $a_{11} = a_{22} = \dots = a_{nn} = k$ , entonces  $D$  recibe el nombre de matriz escalar. Si además  $k = 1$ , la matriz se denomina **unidad o idéntica** y se representa por  $I_n$ .

Las matrices unitarias tienen algunas de las propiedades del entero 1. En todo caso, se verifica que:

$$I_m \times A_{m \times n} = A_{m \times n} \times I_n = I_m \times A_{m \times n} \times I_n = A$$

**Ejemplo 1.11**  $I_2 = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$        $I_3 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$

**Definición 1.26 (Matrices Conmutativas y Anticonmutativas)** Si  $A$  y  $B$  son dos matrices cuadradas y se verifica que  $AB = BA$  dichas matrices se llaman **permutables, conmutativas** o que conmutan. Es fácil demostrar que si  $A$  es una matriz cuadrada de orden  $n$ , conmuta consigo misma y también con  $I_n$ .

En las condiciones anteriores, si  $A$  y  $B$  son tales que  $AB = -BA$ , las matrices  $A$  y  $B$  se llaman **antipermutables o anticonmutativas**

**Ejemplo 1.12** Demostrar que las matrices  $\begin{bmatrix} a & b \\ b & a \end{bmatrix}$  y  $\begin{bmatrix} c & d \\ d & c \end{bmatrix}$  son permutables

para todos los valores de  $a, b, c, d$ . Esto se deduce de:

$$\begin{bmatrix} a & b \\ b & a \end{bmatrix} \times \begin{bmatrix} c & d \\ d & c \end{bmatrix} = \begin{bmatrix} ac + bd & ad + bd \\ bc + ad & bc + ad \end{bmatrix} = \\
\begin{bmatrix} ca + db & da + db \\ cb + da & cb + da \end{bmatrix} = \begin{bmatrix} c & d \\ d & c \end{bmatrix} \times \begin{bmatrix} a & b \\ b & a \end{bmatrix}$$

**Definición 1.27 (Matriz Idempotente)** Una matriz  $A$  de manera que  $A^{k+1} = A$ , se llama **matriz idempotente de grado  $k$**  la matriz  $A$ . Al grado  $k$  también se le llama **periodo**.

**Ejemplo 1.13**  $A = \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix}$  Verificar si la matriz  $A$  es idempotente:

$$A^2 = \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix} \times \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix} = \begin{bmatrix} 2 & -2 & -4 \\ -1 & 3 & 4 \\ 1 & -2 & -3 \end{bmatrix} = A$$

**Definición 1.28 (Matriz Nilpotente)** Una matriz  $A$  tal que  $A^p = 0$ , siendo  $p$  un número entero y positivo, se llama **nilpotente**. Si  $p$  es el menor número entero y positivo para la cual  $A^p = 0$ , la matriz  $A$  se llama **nilpotente de índice  $p$** .

**Ejemplo 1.14**  $A = \begin{bmatrix} 1 & 1 & 3 \\ 5 & 2 & 6 \\ -2 & -1 & -3 \end{bmatrix}$  Demostrar que  $A$  es una matriz nilpotente de índice 3.

$$A^2 = \begin{bmatrix} 1 & 1 & 3 \\ 5 & 2 & 6 \\ -2 & -1 & -3 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 3 \\ 5 & 2 & 6 \\ -2 & -1 & -3 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 3 & 3 & 9 \\ -1 & -1 & -3 \end{bmatrix}$$

$$A^3 = A^2 \times A = \begin{bmatrix} 0 & 0 & 0 \\ 3 & 3 & 9 \\ -1 & -1 & -3 \end{bmatrix} \times \begin{bmatrix} 1 & 1 & 3 \\ 5 & 2 & 6 \\ -2 & -1 & -3 \end{bmatrix} = 0$$

**Definición 1.29 (Matriz Involutiva)** Una matriz cuadrada  $A$  tal que  $A^2 = I$  se llama **involutiva**. Una matriz unidad, por ejemplo, es involutiva. La inversa de una matriz involutiva es ella misma.

**Proposición 1.4** La condición necesaria y suficiente para que una matriz  $A$  sea involutiva es que  $(I - A)(I + A) = 0$

**Demostración.** Supóngase  $(I - A)(I + A) = I - A^2 = 0$ ; luego  $A^2 = I$  y  $A$  es involutiva.

Supóngase que  $A$  es involutiva; entonces  $A^2 = I$  y  $(I - A)(I + A) = I - A^2 = I - I = 0$  ■

**Definición 1.30 (Matriz Simétrica)** Una matriz cuadrada  $A$  tal que  $A^t = A$  se llama **simétrica**. Por tanto, en una matriz cuadrada  $A = [a_{ij}]$  simétrica se verifica que  $a_{ij} = a_{ji}$  para todos los valores de  $i$  y de  $j$ .

**Ejemplo 1.15**  $A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & -5 \\ 3 & -5 & 6 \end{bmatrix}$  es simétrica y también  $kA$  para cualquier  $k \in \mathbb{K}$

Si  $A$  es una matriz cuadrada de orden  $n$ , la matriz  $A + A'$  es simétrica.

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & -5 \\ 3 & -5 & 6 \end{bmatrix}, A' = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & -5 \\ 3 & -5 & 6 \end{bmatrix} \text{ entonces}$$

$$A + A' = \begin{bmatrix} 2 & 4 & 6 \\ 4 & 8 & -10 \\ 6 & -10 & 12 \end{bmatrix}$$

**Definición 1.31 (Matriz Antisimétrica)** Una matriz cuadrada  $A$  tal que  $A' = -A$  se llama hemisimétrica o antisimétrica. Por tanto, en una matriz cuadrada  $A$  hemisimétrica se verifica que  $a_{ij} = -a_{ji}$  para todos los valores de  $i$  y de  $j$ . Evidentemente, los elementos de la diagonal principal deben ser nulos.

$$\textbf{Ejemplo 1.16} \quad A = \begin{bmatrix} 0 & -2 & 3 \\ 2 & 0 & 4 \\ -3 & -4 & 0 \end{bmatrix} \quad A' = \begin{bmatrix} 0 & 2 & -3 \\ -2 & 0 & -4 \\ 3 & 4 & 0 \end{bmatrix} = -A = \begin{bmatrix} 0 & 2 & -3 \\ -2 & 0 & -4 \\ 3 & 4 & 0 \end{bmatrix}$$

La anterior es una matriz hemisimétrica, así como  $kA$ , cualquiera que sea el escalar  $k \in \mathbb{K}$ .

**Proposición 1.5** Toda matriz cuadrada  $A$  se puede descomponer en la suma de una matriz simétrica  $B = \frac{1}{2}(A + A')$  y otra matriz antisimétrica  $C = \frac{1}{2}(A - A')$ .

La inversa de una matriz juega un papel muy importante en el desarrollo de diferentes procedimientos algebraicos y de cálculo matricial. Así como en  $\mathbb{R}$ , para todo  $a \in \mathbb{R}$  con  $a \neq 0$ , existe un  $b \in \mathbb{R}$  que es el inverso multiplicativo, para el caso de una matriz  $A$  análogamente se plantea la posibilidad de que exista una matriz  $B$  conformable, tal que  $AB = I$ .

**Definición 1.32 (Matriz Inversa)** Sean  $A$  y  $B$  dos matrices cuadradas del mismo orden, para las cuales se cumple que  $AB = BA = I$ . A la matriz  $B$  se llama inversa de  $A$  y se escribe  $B = A^{-1}$  ( $B$  igual a inversa de  $A$ ). Recíprocamente, la matriz  $A$  es la inversa de  $B$ , y se puede escribir  $A = B^{-1}$ .

$$\textbf{Ejemplo 1.17} \quad \text{Como} \quad \begin{bmatrix} 1 & 2 & 3 \\ 1 & 3 & 3 \\ 1 & 2 & 4 \end{bmatrix} \times \begin{bmatrix} 6 & -2 & -3 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = I$$

Cada una de las matrices del producto es inversa de la otra.

No todas las matrices poseen inversa. Se puede demostrar, sin embargo, que si  $A$  posee matriz inversa, ésta es única.

**Proposición 1.6 (Propiedades de la Inversa)** *La inversa de una matriz cumple con las siguientes propiedades*

1. Si  $A$  es una matriz que tiene inversa, ésta es única

**Demostración.** Sean  $A$ ,  $B$ ,  $C$  tres matrices cuadradas de forma que  $AB = BA = I$  y  $CA = AC = I$ . En estas condiciones,  $(CA)B = C(AB)$  y por tanto  $B = C$ . En otra forma,  $B = C = A^{-1}$  es la única inversa de  $A$ . ■

2.  $(AB)^{-1} = B^{-1}A^{-1}$

**Demostración.** Por definición  $(AB)^{-1}(AB) = (AB)(AB)^{-1} = I$ . Ahora bien:

$$(B^{-1}A^{-1})AB = B^{-1}(A^{-1}A)B = B^{-1}IB = B^{-1}B = I$$

$$AB(B^{-1}A^{-1}) = A(BB^{-1})A^{-1} = AA^{-1} = I$$

Como  $(AB)^{-1}$  es única, luego  $(AB)^{-1} = B^{-1}A^{-1}$  ■

Una matriz que posee inversa se llama **invertible** o **no singular**.

**Definición 1.33 (Matriz Ortogonal)** *Se dice que una matriz  $A$  definida sobre un cuerpo  $\mathbb{k}$ , es ortogonal si  $AA^T = A^T A = I$ . Una matriz ortogonal  $A$  es necesariamente cuadrada e invertible, con una inversa  $A^{-1} = A^T$*

$$\text{Sea } A = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix}$$

Si  $A$  es ortogonal, entonces se debe cumplir:

$$AA^T = \begin{pmatrix} a_1 & a_2 & a_3 \\ b_1 & b_2 & b_3 \\ c_1 & c_2 & c_3 \end{pmatrix} \begin{pmatrix} a_1 & b_1 & c_1 \\ a_2 & b_2 & c_2 \\ a_3 & b_3 & c_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Esto proporciona el siguiente conjunto de ecuaciones:

$$\begin{array}{lll} a_1^2 + a_2^2 + a_3^2 = 1 & a_1b_1 + a_2b_2 + a_3b_3 = 0 & a_1c_1 + a_2c_2 + a_3c_3 = 0 \\ b_1a_1 + b_2a_2 + b_3a_3 = 0 & b_1^2 + b_2^2 + b_3^2 = 1 & b_1c_1 + b_2c_2 + b_3c_3 = 0 \\ c_1a_1 + c_2a_2 + c_3a_3 = 0 & c_1b_1 + c_2b_2 + c_3b_3 = 0 & c_1^2 + c_2^2 + c_3^2 = 1 \end{array}$$

o en otras palabras:

$$\begin{array}{lll} u_1 \bullet u_1 = 1 & u_1 \bullet u_2 = 0 & u_1 \bullet u_3 = 0 \\ u_2 \bullet u_1 = 0 & u_2 \bullet u_2 = 1 & u_2 \bullet u_3 = 0 \\ u_3 \bullet u_1 = 0 & u_3 \bullet u_2 = 0 & u_3 \bullet u_3 = 1 \end{array}$$

donde  $u_1 = (a_1, a_2, a_3)$ ,  $u_2 = (b_1, b_2, b_3)$ ,  $u_3 = (c_1, c_2, c_3)$  son las filas de  $A$ . Así las filas  $u_1, u_2$  y  $u_3$  son ortogonales entre sí y tienen longitudes unidad o, dicho de otro modo, forman un **conjunto ortonormal de vectores**

**Definición 1.34 (Matriz Escalonada)** Sean  $[A_1 \ A_2 \ \cdots \ A_S]$  matrices cuadradas de órdenes  $[m_1 \ m_2 \ \cdots \ m_S]$ , respectivamente.

$$\text{La generalización } A = \begin{bmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_S \end{bmatrix} = \text{diag} [A_1 \ A_2 \ \cdots \ A_S]$$

de la matriz diagonal se llama **suma directa** o **matriz escalonada** de las matrices  $A_i$

**Ejemplo 1.18** Sean  $A_1 = [2]$ ,  $A_2 = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}$ ,  $A_3 = \begin{bmatrix} 1 & 2 & -1 \\ 2 & 0 & 3 \\ 4 & 1 & -2 \end{bmatrix}$

La suma directa de  $A_1, A_2, A_3$  es la matriz escalonada:

$$\text{diag} (A_1, A_2, A_3) = \begin{bmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 2 & 0 & 0 & 0 \\ 0 & 3 & 4 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & -1 \\ 0 & 0 & 0 & 2 & 0 & 3 \\ 0 & 0 & 0 & 0 & 1 & -2 \end{bmatrix}$$

También se puede ver a la matriz escalonada como una matriz particionada o una matriz de bloques en la cual las submatrices no nulas están sobre la diagonal

**Definición 1.35 (Determinante de una Matriz Cuadrada)** El determinante de una matriz cuadrada se puede definir recursivamente mediante desarrollos por columnas o por filas. Sea  $A = (a_{ij})$  una matriz  $n \times n$ , donde  $i$  es el índice de la fila y  $j$  es el índice de la columna. Se nota por  $A_{ij}$  la matriz  $(n-1) \times (n-1)$  que se obtiene al quitar la fila  $i$  y la columna  $j$  de la matriz  $A$ . Entonces

$$\text{Desarrollo por la fila } i : \det A = |A| = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}).$$

$$\text{Desarrollo por la columna } j : \det A = |A| = \sum_{i=1}^n (-1)^{i+j} a_{ij} \det(A_{ij}).$$

Aplicando repetidamente estas fórmulas, se va reduciendo el orden de las determinantes hasta llegar a determinantes de órdenes uno, dos o tres que se pueden calcular usando las reglas de Sarrus:

$$|a_{11}| = a_{11}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

$$\begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{21}a_{32}a_{13}$$

$$-a_{13}a_{22}a_{31} - a_{23}a_{32}a_{11} - a_{21}a_{12}a_{33}$$

El valor del determinante no depende de las filas o columnas escogidas, mientras que la dificultad del cálculo probablemente sí.

**Proposición 1.7 (Propiedades del Determinante)** *Las principales propiedades de los determinantes de matrices cuadradas son las siguientes.*

1. Si una columna es cero, el determinante es cero.
2. Si hay dos columnas iguales, el determinante es cero.
3. Si las columnas son ld, el determinante es cero.
4. El determinante cambia de signo al permutar dos columnas.
5. El determinante no cambia si a una columna se le suma una columna de las restantes.
6. El determinante es lineal respecto a cada columna:  
 $\det(\dots, c_i + c_j, \dots) = \det(\dots, c_i, \dots) + \det(\dots, c_j, \dots)$ .  
 $\det(\dots, c_i, \dots) = \det(\dots, c_i, \dots)$ .
7. Las filas también cumplen las anteriores propiedades.
8.  $\det(\lambda A) = \lambda^n \det(A)$ .
9. El determinante del producto es igual al producto de determinantes:  
 $\det(AB) = \det A \cdot \det B$ .
10. Una matriz  $A$  es invertible si y solo si  $\det A \neq 0$ . Además,  $\det(A^{-1}) = (\det A)^{-1}$ .
11. Una matriz y su traspuesta tienen el mismo determinante:  $\det(A^T) = \det A$ .
12. El determinante de una matriz triangular es igual al producto de los elementos diagonales.
13. El determinante de una matriz triangular por bloques es igual al producto de los determinantes de los bloques diagonales.

## 2. Derivación.

En esta sección se exponen los conceptos fundamentales de la derivación de funciones reales, y algunas de las reglas principales. Se consideran principalmente las funciones algebraicas, y al final de la sección se exponen las fórmulas de algunas funciones trigonométricas y logarítmicas que podrán ser utilizadas posteriormente.

**Definición 2.1 (Pendiente)** Se define la pendiente, en el punto donde  $x = a$ , de la tangente a una curva cuya ecuación es  $y = f(x)$ , como

$$m = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}$$

De hecho, los límites con esta forma surgen siempre al calcular una rapidez de cambio en cualquier ciencia o rama de la ingeniería, como la rapidez de reacción en química o un costo marginal en economía. Dado que este tipo de límite se presenta con suma frecuencia, se le da un nombre y una notación especial.

**Definición 2.2 (Derivada)** La derivada de la función  $f$  en un número  $a$  representada por  $f'(a)$  es

$$f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \text{ dado el caso de que el límite exista.}$$

Si se escribe  $x = a + h$ , entonces  $h = x - a$  y  $h \rightarrow 0$  si y solo si  $x \rightarrow a$ ; por consiguiente de acuerdo con la determinación de las tangentes, un modo equivalente de enunciar la definición de derivada, es

$$f'(a) = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}$$

**Ejemplo 2.1** Determinar la derivada de la función  $f(x) = x^2 - 8x + 9$  en el número  $a$ .

$$\begin{aligned} f'(a) &= \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h} \\ f'(a) &= \lim_{h \rightarrow 0} \frac{[(a+h)^2 - 8(a+h) + 9] - [a^2 - 8a + 9]}{h} \\ f'(a) &= \lim_{h \rightarrow 0} \frac{a^2 + 2ah + h^2 - 8a - 8h + 9 - a^2 + 8a - 9}{h} \\ f'(a) &= \lim_{h \rightarrow 0} \frac{2ah + h^2 - 8h}{h} = \lim_{h \rightarrow 0} (2a + h - 8) \\ f'(a) &= 2a - 8 \end{aligned}$$

El proceso de calcular la derivada de una función se llama derivación.

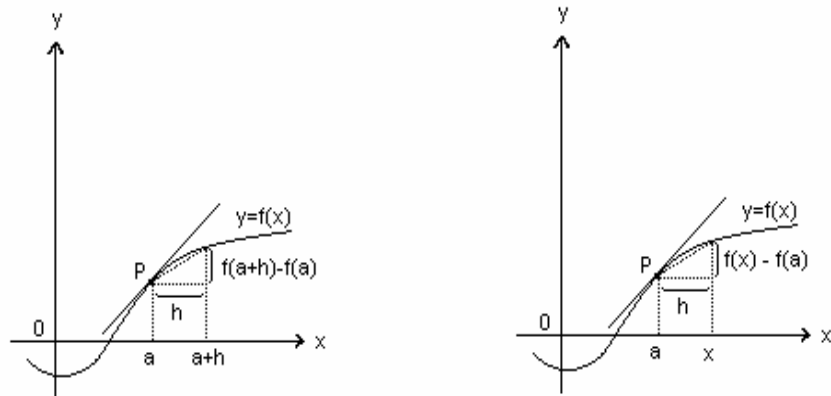


Figura 1:

**Notación 2.1 (Derivada)** Si se emplea la notación tradicional  $y = f(x)$  para indicar que la variable independiente es  $x$  y que la dependiente es  $y$ , hay otras notaciones alternativas comunes de la derivada:

$$f'(x) = y' = \frac{dy}{dx} = \frac{df}{dx} = \frac{d}{dx}f(x) = Df(x) = D_x f(x)$$

Los símbolos  $D$  y  $\frac{d}{dx}$  se denominan operadores de diferenciación porque indican la operación de diferenciación, que es el proceso de calcular una derivada.

El símbolo  $\frac{d}{dx}$  fue introducido por Leibniz y no se debe considerar como una relación, solo es un símbolo de  $f'(x)$ . No obstante, es una notación muy útil y sugerente, en especial cuando se usa con la notación de incrementos.

### 2.0.3. Interpretación de la derivada como la pendiente de una tangente.

Si se parte de la definición de que la línea tangente o recta tangente a la curva  $y = f(x)$  en el punto  $P(a, f(a))$  es la línea que pasa por  $P$  cuya pendiente es

$$m = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} \quad \text{siempre que exista ese límite}$$

Como, según la definición 2.2 es la misma que la derivada  $f'(a)$  ahora se puede decir que la recta tangente a  $y = f(x)$  en  $(a, f(a))$  es la línea que pasa por  $(a, f(a))$  cuya pendiente es igual a  $f'(a)$  la derivada de  $f$  en  $a$ . Así, la interpretación geométrica de una derivada (como se tiene en la definición 2.2) es lo que registra la figura ??.



$$f'(a) = \lim_{h \rightarrow 0} \frac{f(a+h) - f(a)}{h}, \text{ b) } f'(a) = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a}$$

= pendiente de tangente en P, = pendiente de tangente en P

Al emplear la forma punto-pendiente de la ecuación de la recta, se llega a:  
 Si existe  $f'(a)$  entonces una ecuación de la recta tangente a la curva  $y = f(x)$  en el punto  $(a, f(a))$  es la siguiente:

$$y - f(a) = f'(a)(x - a)$$

**Ejemplo 2.2** Deducir una ecuación de la tangente a la parábola  $y = x^2 - 8x + 9$  en el punto  $(3, -6)$ .

De acuerdo con el ejemplo anterior se sabe que la derivada de  $f(x) = x^2 - 8x + 9$ , en el número  $a$  es  $f'(a) = 2a - 8$ . Entonces la pendiente de la tangente en  $(3, -6)$  es  $f'(3) = 2(3) - 8 = -2$ . Así, la ecuación de la recta tangente  $y - (-6) = (-2)(x - 3)$  o sea  $y = -2x$ .

#### 2.0.4. Reglas de Derivacion.

Una función se dice diferenciable en un intervalo si lo es en cada uno de sus puntos. Las funciones del cálculo elemental son diferenciables, excepto posiblemente en puntos aislados, en sus intervalos de definición.

Si siempre fuera necesario determinar las derivadas directamente a partir de la definición, las operaciones serían tediosas y se requeriría mucho ingenio para evaluar algunos límites. Por fortuna, se han desarrollado varias reglas para hallar derivadas que obvian ese proceso y simplifican mucho la diferenciación. Las reglas siguientes suponen que se trabaja con funciones diferenciables.

**Regla 2.1 (Derivada de una Constante)** Si  $f$  es una función constante,  $f(x) = c$ , entonces  $f'(x) = 0$

Este resultado es geoméricamente evidente porque la gráfica de una función constante es una recta horizontal con pendiente 0; la demostración formal también es simple.

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{c - c}{h} = \lim_{h \rightarrow 0} 0 = 0$$

En notación de Leibniz, se escribe:  $\frac{d}{dx}c = 0$

**Regla 2.2 (Regla de Potencias)** Si  $f(x) = x^n$  en donde  $n$  es un entero positivo,  $f'(x) = nx^{n-1}$

En la notación de Leibniz la regla de potencias se expresa como:  $\frac{d}{dx}(x^n) = nx^{n-1}$

**Demostración.** La fórmula

$$x^n - a^n = (x - a)(x^{n-1} + x^{n-2}a + \dots + xa^{n-2} + a^{n-1})$$

Se comprueba multiplicando el lado derecho, o sumando el segundo factor como una serie geométrica. Así, si se usa la ecuación 2.2 para  $f'(a)$  y después se utiliza la expresión de arriba, se obtiene:

$$f'(a) = \lim_{x \rightarrow a} \frac{f(x) - f(a)}{x - a} = \lim_{x \rightarrow a} \frac{x^n - a^n}{x - a}$$

$$f'(a) = \lim_{x \rightarrow a} (x^{n-1} + x^{n-2}a + \dots + xa^{n-2} + a^{n-1})$$

$$f'(a) = a^{n-1} + a^{n-2} + \dots + xaa^{n-2} + a^{n-1} \blacksquare$$

**Demostración.**  $f'(a) = na^{n-1}$

$$f'(x) = \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} = \lim_{h \rightarrow 0} \frac{(x+h)^n - x^n}{h}$$

Al desarrollar  $(x+h)^n$  de acuerdo con el teorema del binomio, se llega a:

$$f'(x) = \lim_{h \rightarrow 0} \frac{\left[ x^n + nx^{n-1}h + \frac{n(n-1)}{2}x^{n-2}h^2 + \dots + nxh^{n-1} + h^n \right] - x^n}{h}$$

$$f'(x) = \lim_{h \rightarrow 0} \left[ nx^{n-1} + \frac{n(n-1)}{2}x^{n-2}h + \dots + nxh^{n-2} + h^{n-1} \right]$$

$$f'(x) = nx^{n-1}$$

Por cuanto todos los términos excepto el primero, tienen a  $h$  como factor; por consiguiente, tienden a 0.  $\blacksquare$

**Ejemplo 2.3** Los siguientes ejemplos

$$\blacksquare \text{ Si } y = t^5 \text{ entonces } \frac{dy}{dt} = 5t^4$$

$$\blacksquare D_u(u^m) = mu^{m-1}$$

$$\blacksquare \text{ Si } y = x^{20}, \text{ entonces } y' = 20x^{19}$$

$$\blacksquare \frac{d}{dr}(r^6) = 6r^5$$

- $\frac{d}{dx}\sqrt{x} = \frac{d}{dx}\left(x^{\frac{1}{2}}\right) = \frac{1}{2}x^{\left(\frac{1}{2}\right)-1} = \frac{1}{2\sqrt{x}}$

- Sea  $y = \frac{1}{\sqrt[3]{x^2}}$

Entonces  $\frac{dy}{dx} = \frac{d}{dx}\left(x^{-\frac{2}{3}}\right) = -\frac{2}{3}x^{-\left(\frac{2}{3}\right)-1} = -\frac{2}{3}x^{-\frac{5}{3}}$

- Derivar la función  $f(t) = \sqrt{t}(1-t)$ .

Aplicando la regla del producto se tiene:

$$f'(t) = \sqrt{t}\frac{d}{dt}(1-t) + (1-t)\frac{d}{dt}\sqrt{t}$$

$$f'(t) = \sqrt{t}(-1) + (1-t)\frac{1}{2}t^{-\frac{1}{2}}$$

$$f'(t) = -\sqrt{t} + \frac{1-t}{2\sqrt{t}} = \frac{1-3t}{2\sqrt{t}}$$

Si se utilizan primero las leyes de los exponentes, después se podrá proceder directamente, sin recurrir a la regla del producto.

$$f(t) = \sqrt{t} - t\sqrt{t} = t^{\frac{1}{2}} - t^{\frac{3}{2}}$$

$$f'(t) = \frac{1}{2}t^{-\frac{1}{2}} - \frac{3}{2}t^{\frac{1}{2}} \text{ que equivale a la respuesta en la solución anterior}$$

Las fórmulas de diferenciación que siguen indican que la derivada de una constante multiplicada por una función es igual a la constante multiplicada por la derivada de la función, y que la derivada de una suma (o resta) de funciones es igual a la suma (o resta) de las derivadas, siempre y cuando las derivadas existan.

Para definir las siguientes reglas, se supone que  $c$  es una constante y que tanto  $f'(x)$  como  $g'(x)$  existen.

**Regla 2.3** Si  $g(x) = cf(x)$ , entonces  $f'(x)$  existe y  $g'(x) = cf'(x)$

También se escribe:  $\frac{d}{dx}(cf) = c\frac{df}{dx}$

$$\begin{aligned}
\textbf{Demostración. } g'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
&= \lim_{h \rightarrow 0} \frac{cf(x+h) - cf(x)}{h} \\
g'(x) &= \lim_{h \rightarrow 0} c \left[ \frac{f(x+h) - f(x)}{h} \right] \\
g'(x) &= c \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
g'(x) &= cf'(x) \blacksquare
\end{aligned}$$

**Regla 2.4** Si  $H(x) = f(x) \pm g(x)$ , entonces  $H'(x)$  existe, y

$$\begin{aligned}
H'(x) &= f'(x) \pm g'(x) \\
\text{También se escribe: } \frac{d}{dx}(f \pm g) &= \frac{df}{dx} \pm \frac{dg}{dx}
\end{aligned}$$

Este resultado se puede ampliar para la suma de cualquier número de funciones.

$$\begin{aligned}
\textbf{Demostración. } H'(x) &= \lim_{h \rightarrow 0} \frac{H(x+h) - H(x)}{h} \\
H'(x) &= \lim_{h \rightarrow 0} \frac{[f(x+h) \pm g(x+h)] - [f(x) \pm g(x)]}{h} \\
H'(x) &= \lim_{h \rightarrow 0} \left[ \frac{f(x+h) - f(x)}{h} \pm \frac{g(x+h) - g(x)}{h} \right] \\
H'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \pm \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} \\
H'(x) &= f'(x) \pm g'(x) \blacksquare
\end{aligned}$$

De manera resumida, las dos reglas anteriores se pueden expresar como:

$$\begin{aligned}
(cf)' &= cf' \\
(f \pm g)' &= f' \pm g'
\end{aligned}$$

**Ejemplo 2.4**  $\frac{d}{dx}(x^8 + 12x^5 - 4x^4 + 10x^3 - 6x + 5)$

$$\begin{aligned}
&= \frac{d}{dx}(x^8) + 12 \frac{d}{dx}(x^5) - 4 \frac{d}{dx}(x^4) + 10 \frac{d}{dx}(x^3) - 6 \frac{d}{dx}(x) + \frac{d}{dx}(5) \\
&= 8x^7 + 12(5x^4) - 4(4x^3) + 10(3x^2) - 6(1) + 0 \\
&= 8x^7 + 60x^4 - 16x^3 + 30x^2 - 6
\end{aligned}$$

**Ejemplo 2.5** Si  $f(x) = x^4 - x^3 + x^2 - x + 1$ , deducir la ecuación de la tangente a la gráfica de  $f$  en el punto  $(1, 1)$ .

La pendiente de  $f'(1)$  que se calculara como sigue:

$$f'(x) = 4x^3 - 3x^2 + 2x - 1$$

$$f'(1) = 4 - 3 + 2 - 1 = 2$$

Por tanto, la ecuación de la tangente en  $(1, 1)$  es  $y - 1 = 2(x - 1)$  o bien  $2x - y - 1 = 0$

A continuación se necesita una fórmula para la derivada de un producto de dos funciones; es posible sentirse inclinado a suponer, como lo hizo Leibniz hace tres siglos, que la derivada de un producto es igual al producto de las derivadas; pero se puede ver que tal hipótesis no es correcta si se considera un ejemplo en particular. Sean  $f(x) = x$  y  $g(x) = x^2$ . Entonces, la regla de potencias establece que  $f'(x) = 1$  y  $g'(x) = 2x$  y se tendría  $f' \cdot g' = 2x$ . Sin embargo, si se toma  $h = f \cdot g = x^3$  y, por consiguiente,  $h' = (f \cdot g)' = 3x^2$ . Entonces  $(f \cdot g)' \neq f' \cdot g'$ . Leibniz descubrió la fórmula correcta, la cual se llama regla del producto.

**Regla 2.5 (Regla del Producto)** Si  $H(x) = f(x)g(x)$  y tanto  $f'(x)$  como  $g'(x)$  existen a la vez, entonces

$$H'(x) = f'(x)g(x) + f(x)g'(x)$$

$$\text{También se puede expresar como: } \frac{d}{dx}(fg) = f \frac{dg}{dx} + g \frac{df}{dx}$$

$$\text{y de manera abreviada: } (fg)' = fg' + f'g$$

$$\textbf{Demostración.} \quad H'(x) = \lim_{h \rightarrow 0} \frac{H(x+h) - H(x)}{h}$$

$$H'(x) = \lim_{h \rightarrow 0} \frac{f(x+h)g(x+h) - f(x)g(x)}{h}$$

Para evaluar este límite, se van a separar las funciones  $f$  y  $g$  sumando y restando el término  $f(x+h)g(x)$  en el numerador

$$\begin{aligned} H'(x) &= \lim_{h \rightarrow 0} \frac{f(x+h)g(x+h) - f(x+h)g(x) + f(x+h)g(x) - f(x)g(x)}{h} \\ &= \lim_{h \rightarrow 0} \left[ f(x+h) \frac{g(x+h) - g(x)}{h} + g(x) \frac{f(x+h) - f(x)}{h} \right] \end{aligned}$$

$$\begin{aligned}
&= \lim_{h \rightarrow 0} f(x+h) \lim_{h \rightarrow 0} \frac{g(x+h) - g(x)}{h} + \\
&+ \lim_{h \rightarrow 0} g(x) \lim_{h \rightarrow 0} \frac{f(x+h) - f(x)}{h} \\
H'(x) &= f(x)g'(x) + f'(x)g(x) \blacksquare
\end{aligned}$$

**Ejemplo 2.6** Determinar  $F'(x)$  si  $F(x) = (6x^3)(7x^4)$

Según la regla del producto:

$$\begin{aligned}
F'(x) &= (6x^3) \frac{d}{dx}(7x^4) + (7x^4) \frac{d}{dx}(6x^3) \\
F'(x) &= (6x^3)(28x^3) + (7x^4)(18x^2) \\
F'(x) &= 168x^6 + 126x^6 = 294x^6
\end{aligned}$$

Se podría comprobar la respuesta a este ejemplo en forma directa multiplicando primero los factores:

$$F(x) = (6x^3)(7x^4) = 42x^7 \Rightarrow F'(x) = 42(7x^6) = 294x^6$$

**Regla 2.6 (Regla del Cociente)** Si  $F(x) = \frac{f(x)}{g(x)}$  y existen  $f'(x)$  y  $g'(x)$  a la vez, entonces existe  $F'(x)$  y  $F'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2}$

$$\text{En la notación de Leibniz: } \frac{d}{dx} \left( \frac{f(x)}{g(x)} \right) = \frac{g(x) \frac{d}{dx} f(x) - f(x) \frac{d}{dx} g(x)}{[g(x)]^2}$$

$$\text{y de manera abreviada: } \left( \frac{f}{g} \right)' = \frac{gf' - fg'}{g^2}$$

$$\textbf{Demostración.} \quad F'(x) \lim_{h \rightarrow 0} \frac{F(x+h) - F(x)}{h} = \lim_{h \rightarrow 0} \frac{\frac{f(x+h)}{g(x+h)} - \frac{f(x)}{g(x)}}{h}$$

$$F'(x) \lim_{h \rightarrow 0} \frac{f(x+h)g(x) - f(x)g(x+h)}{hg(x+h)g(x)}$$

Para separar  $f$  y  $g$  en esta expresión se suma y resta el término  $f(x)g(x)$  al numerador:

$$F'(x) \lim_{h \rightarrow 0} \frac{f(x+h)g(x) - f(x)g(x) + f(x)g(x) - f(x)g(x+h)}{hg(x+h)g(x)}$$

$$f'(x) \lim_{h \rightarrow 0} \frac{g(x) \frac{f(x+h)-f(x)}{h} - f(x) \frac{g(x+h)-g(x)}{h}}{g(x+h)g(x)}$$

$$f'(x) \frac{\lim_{h \rightarrow 0} g(x) \lim_{h \rightarrow 0} \frac{f(x+h)-f(x)}{h} - \lim_{h \rightarrow 0} f(x) \lim_{h \rightarrow 0} \frac{g(x+h)-g(x)}{h}}{\lim_{h \rightarrow 0} g(x+h) \lim_{h \rightarrow 0} g(x)}$$

$$f'(x) = \frac{g(x)f'(x) - f(x)g'(x)}{[g(x)]^2} \blacksquare$$

En palabras, la derivada de un cociente es igual al denominador multiplicado por la derivada del numerador, menos el numerador multiplicado por la derivada del denominador, y todo ello se divide entre el cuadrado del denominador.

**Ejemplo 2.7** Sea  $y = \frac{x^2-x-2}{x^3+6}$

$$\text{Entonces } y' = \frac{(x^3+6)D(x^2+x-2) - (x^2+x-2)D(x^3+6)}{(x^3+6)^2}$$

$$y' = \frac{(x^3+6)(2x+1) - (x^2+x-2)(3x^2)}{(x^3+6)^2}$$

$$y' = \frac{(2x^4+x^3+12x+6) - (3x^4+3x^3-6x^2)}{(x^3+6)^2}$$

$$y' = \frac{-x^4-2x^3+6x^2+12x+6}{(x^3+6)^2}$$

También se puede emplear la regla del cociente para ampliar la regla de potencias al caso en que el exponente es un entero negativo.

**Regla 2.7** Si  $f(x) = x^{-n}$ , donde  $n$  es un entero positivo,  $f'(x) = -nx^{-n-1}$

$$\text{Demostración. } f'(x) = \frac{d}{dx}(x^{-n}) = \frac{d}{dx}\left(\frac{1}{x^n}\right)$$

$$f'(x) = \frac{x^n D(1) - 1 D(x^n)}{(x^n)^2}$$

$$f'(x) = \frac{-nx^{n-1}}{x^{2n}} = -nx^{n-1-2n} = -nx^{-n-1} \blacksquare$$

**Ejemplo 2.8** Si  $y = \frac{1}{x}$

$$\text{Entonces, } \frac{dy}{dx} = \frac{d}{dx}(x^{-1}) = -x^{-2} = -\frac{1}{x^2}$$

**Ejemplo 2.9**  $\frac{d}{dt}\left(\frac{6}{t^3}\right) = 6\frac{d}{dt}(t^{-3}) = 6(-3)t^{-4} = -\frac{18}{t^4}$

**Ejemplo 2.10** Derivar la función  $f(t) = \sqrt{t}(1-t)$  Aplicando la regla del producto se tiene:

$$f'(t) = \sqrt{t} \frac{d}{dt}(1-t) + (1-t) \frac{d}{dt}\sqrt{t}$$

$$f'(t) = \sqrt{t}(-1) + (1-t) \frac{1}{2}t^{-\frac{1}{2}}$$

$$f'(t) = -\sqrt{t} + \frac{1-t}{2\sqrt{t}} = \frac{1-3t}{2\sqrt{t}}$$

Si se utilizan primero las leyes de los exponentes, después se podrá proceder directamente, sin recurrir a la regla del producto.

$$f(t) = \sqrt{t} - t\sqrt{t} = t^{\frac{1}{2}} - t^{\frac{3}{2}}$$

$$f'(t) = \frac{1}{2}t^{-\frac{1}{2}} - \frac{3}{2}t^{\frac{1}{2}} \text{ que equivale a la respuesta en la solución anterior}$$

**Ejemplo 2.11** En qué puntos de la hipérbola  $xy = 12$  la tangente es paralela a la recta  $3x + y = 0$ ?

Como  $xy = 12$  se puede escribir en la forma  $y = \frac{12}{x}$

$$\frac{dy}{dx} = 12 \frac{d}{dx}(x^{-1}) = 12(-x^{-2}) = -\frac{12}{x^2}$$

Sea  $a$  la abscisa de uno de los puntos en cuestión. Entonces, la pendiente de la tangente en ese punto es  $-\frac{12}{a^2}$ . Esa tangente será paralela a la recta  $3x + y = 0$  o  $y = -3x$ , si tiene la misma pendiente, que es  $-3$ . Al igualar las pendientes se llega a  $-\frac{12}{a^2} = -3$ , o sea  $a^2 = 4$ , o sea  $a = \pm 2$ . Por consiguiente, los puntos buscados son  $(2, 6)$  y  $(-2, -6)$ .

**Regla 2.8 (Regla de la Cadena)** Sean  $f, g$  dos funciones. Si existen a la vez las derivadas  $g'$  y  $f'$  y si  $H = f \circ g$  es la función compuesta definida por  $H(x) = f(g(x))$ , entonces  $H'(x)$  existe y está dada por el producto  $H'(x) = f'(g(x))g'(x)$

En la notación de Leibnitz, si  $y = f(u)$  y  $u = g(x)$  son dos funciones diferenciables, entonces

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$$

**Demostración.** Se desarrollará mediante el examen de dos casos.

$$\frac{du}{dx} \neq 0$$

Aquí  $\Delta u \neq 0$  si  $\Delta x$  es suficientemente pequeña (porque si no  $du/dx$  sería 0) pero en este caso se puede dividir entre y multiplicar por  $\Delta u$  en



$$1. \frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta x}$$

por tanto

$$\frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta u} \frac{\Delta u}{\Delta x}$$

$$\frac{dy}{dx} = \lim_{\Delta x \rightarrow 0} \frac{\Delta y}{\Delta u} \bullet \lim_{\Delta x \rightarrow 0} \frac{\Delta u}{\Delta x} \quad \Delta u \rightarrow 0 \text{ cuando } \Delta x \rightarrow 0$$

porque  $g$  es continua

$$\frac{dy}{dx} = \lim_{\Delta u \rightarrow 0} \frac{\Delta y}{\Delta u} \bullet \lim_{\Delta x \rightarrow 0} \frac{\Delta u}{\Delta x}$$

$$\frac{dy}{dx} = \frac{dy}{du} \frac{du}{dx}$$

$$2. \frac{du}{dx} = 0$$

Aquí  $\Delta u = 0$  para algunos casos de  $\Delta x$ , y  $\Delta u \neq 0$  para otros valores de  $\Delta x$ . Si se tiene  $\Delta x \rightarrow 0$  pasando por valores tales que  $\Delta u \neq 0$ , podemos escribir

$$\frac{\Delta y}{\Delta x} = \frac{\Delta y}{\Delta u} \bullet \frac{\Delta u}{\Delta x} \rightarrow \frac{dy}{du} \bullet \frac{du}{dx} = \frac{dy}{du} \bullet 0 = 0$$

Si  $\Delta x \rightarrow 0$  pasando por valores tales que  $\Delta u = 0$ , entonces\

$$\Delta y = f(u + \Delta u) - f(u) = f(u) - f(u) = 0$$

$$\text{y así } \frac{\Delta y}{\Delta x} = 0$$

En ambos casos  $\frac{\Delta y}{\Delta x} \rightarrow 0$ ,

$$\text{y entonces } \frac{dy}{dx} = 0 = \frac{dy}{du} \frac{du}{dx}$$

En estas circunstancias se comprueba la regla de la cadena, porque ambos lados son 0. ■

### 2.0.5. Derivación Parcial

Los casos expuestos anteriormente se aplican para funciones de una sola variable. Cuando se trabaja con funciones de varias variables  $f(x, y, z, \dots)$  el proceso es similar, teniendo en cuenta que se puede derivar de manera separada la función  $f$  con respecto a cada una de sus variables independientes. Para simplificar la presentación, se considerará en las líneas siguientes el caso de una función de dos variables, pero cada caso se puede llevar por analogía a funciones de tres o más variables.

**Definición 2.3** Sea  $z = f(x, y)$  una función de dos variables independientes  $x, y$ . Entonces, las derivadas parciales de  $z$  son las funciones  $f_x$  y  $f_y$  que se definen como:

$$\begin{aligned} f_x(x, y) &= \lim_{h \rightarrow 0} \frac{f(x+h, y) - f(x, y)}{h} \\ f_y(x, y) &= \lim_{h \rightarrow 0} \frac{f(x, y+h) - f(x, y)}{h} \end{aligned}$$

Donde  $f_x$  se denomina la primera derivada parcial de  $f(x, y)$  con respecto a  $x$ , y  $f_y$  se denomina la primera derivada parcial de  $f(x, y)$  con respecto a  $y$ .

Por tanto,  $f_x$  permite calcular la variación de  $f$  a medida que cambia  $x$ , con todas las demás variables fijas, y equivalentemente para  $f_y$ .

**Notación 2.2** Sea  $z = f(x, y)$ . Entonces, las derivadas parciales de esta función con respecto a  $x$  y con respecto a  $y$  son:

$$\begin{aligned} f_x(x, y) &= f_x = \frac{\partial f}{\partial x} = \frac{\partial}{\partial x} f(x, y) = \frac{\partial z}{\partial x} = f_1 = D_1 f = D_x f \\ f_y(x, y) &= f_y = \frac{\partial f}{\partial y} = \frac{\partial}{\partial y} f(x, y) = \frac{\partial z}{\partial y} = f_2 = D_2 f = D_y f \end{aligned}$$

### 3. Derivación Matricial

El cálculo matricial es en gran manera una ampliación de la notación tradicional del cálculo univariado, que permite una notación más compacta a la hora de realizar cálculo multivariado, en el cual las funciones pueden ser de un tipo más amplio, ya que la variable puede ser vectorial o matricial. En general, el cálculo matricial trabaja sobre espacios de matrices  $M(m, n)$  de tamaño  $m \times n$  definidas sobre  $\mathbb{k} = \mathbb{R}$ .

Antes de abordar el proceso de la derivación matricial dada la complejidad técnica del problema, es preciso incorporar algunas nuevas operaciones entre matrices que completan el cálculo matricial tradicional. Por ello, se definen las matrices de permutación y los conceptos de producto Kronecker y vectorización junto con sus propiedades.

Primero se exponen nuevos tipos de matrices que son de especial interés para desarrollos del cálculo matricial.

**Nota 3.1** *Es importante resaltar en todo caso, que en este escrito se identificarán explícitamente los espacios de matrices  $M_{pn}$  con  $\mathbb{R}^{pn}$  y  $M_{mq}$  con  $\mathbb{R}^{mq}$ , respectivamente.*

#### 3.1. Funciones Matriciales

Dentro de esta sección se presentarán los diversos casos de funciones de acuerdo con su variable, hasta llegar al caso general de las funciones de variable matricial.

Para ello, sea  $M_{mn}$  el espacio vectorial de las matrices de tamaño  $m \times n$ , en el cual se definen las matrices  $X, Y$ . Dentro del espacio vectorial  $M_{1n}$  se definen los vectores fila  $A, B$ . Además se definen los escalares  $x, y \in \mathbb{k}$ . Se define además una función  $f$ .

##### 3.1.1. Funciones de Variable Escalar

Sea  $M_{mn}$  el espacio vectorial de las matrices de tamaño  $m \times n$ , y  $X, Y$  matrices de  $M_{mn}$ . Sea  $M_{1n}$  el espacio vectorial de los vectores fila de tamaño  $n$ , y sean  $A, B$  vectores de  $M_{1n}$ . Sean  $x, y \in \mathbb{k}$  escalares reales. Entonces se pueden dar los siguientes casos:

**Caso 3.1**  $f : \mathbb{R} \rightarrow \mathbb{R}$

$$f(x) = y$$

*Este es el caso usual, cuya derivación y propiedades de la derivación se ha expuesto en la sección 2.*

**Ejemplo 3.1**  $f(x) = x + 1$

**Caso 3.2**  $f : \mathbb{R} \rightarrow M_{1n}$

$$f(x) = A$$

*Es el caso de una función de variable escalar, cuya salida es un vector.*

**Ejemplo 3.2**  $f(x) = (2x, x^2)$

**Caso 3.3**  $f : \mathbb{R} \rightarrow M_{mn}$

$$f(x) = Y$$

**Ejemplo 3.3**  $f(x) = \begin{bmatrix} 1 & x \\ x+1 & x^2 \end{bmatrix}$

### 3.1.2. Funciones de Variable Vectorial

Sea  $M_{mn}$  el espacio vectorial de las matrices de tamaño  $m \times n$ , y  $X, Y$  matrices de  $M_{mn}$ . Sea  $M_{1n}$  el espacio vectorial de los vectores fila de tamaño  $n$ , y sean  $A, B$  vectores de  $M_{1n}$ . Sean  $x, y \in \mathbb{K}$  escalares reales. Entonces se pueden dar los siguientes casos:

**Caso 3.4**  $f : M_{1n} \rightarrow \mathbb{R}$

$$f(A) = y$$

**Caso 3.5**  $f : M_{1n} \rightarrow M_{1n}$

$$f(A) = B$$

*Es el caso de una función de variable vectorial, cuya salida es un vector.*

**Caso 3.6**  $f : M_{1n} \rightarrow M_{mn}$

$$f(A) = Y$$

### 3.1.3. Funciones de Variable Matricial

Sea  $M_{mn}$  el espacio vectorial de las matrices de tamaño  $m \times n$ , y  $X, Y$  matrices de  $M_{mn}$ . Sea  $M_{1n}$  el espacio vectorial de los vectores fila de tamaño  $n$ , y sean  $A, B$  vectores de  $M_{1n}$ . Sean  $x, y \in \mathbb{K}$  escalares reales. Entonces se pueden dar los siguientes casos:

**Caso 3.7**  $f : M_{mn} \rightarrow \mathbb{R}$

$$f(X) = y$$

**Caso 3.8**  $f : M_{mn} \rightarrow M_{1n}$

$$f(X) = B$$

*Es el caso de una función de variable vectorial, cuya salida es un vector.*

**Caso 3.9**  $f : M_{mn} \rightarrow M_{mn}$

$$f(X) = Y$$

## 3.2. Otras Matrices Especiales

Se definen las siguientes matrices:

**Definición 3.1 (Partición de una Matriz)** Una matriz es particionada, si es subdividida en matrices más pequeñas llamadas **submatrices** o **bloques** mediante líneas horizontales y verticales que demarcan filas y columnas completas:

$$\left[ \begin{array}{cc|cc} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & \cdots & \cdots & a_{2n} \\ \hline \vdots & \vdots & \vdots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array} \right]$$

Una matriz  $A_{m \times n}$  se puede particionar de  $2^{m+n-2} - 1$  maneras distintas, con al menos una línea de partición.

**Definición 3.2 (Matriz de Bloques)** Una matriz a la que se ha realizado una partición se llama **matriz de bloques** o **Matriz Particionada**.

**Ejemplo 3.4** Dada la matriz  $A_{3 \times 4}$ , se le pueden realizar  $2^5 - 1 = 31$  particiones diferentes. Una de ellas puede ser:

$$A_{3 \times 4} = \left[ \begin{array}{ccc|c} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ \hline a_{31} & a_{32} & a_{33} & a_{34} \end{array} \right] = \left[ \begin{array}{cc} B & C \\ D & E \end{array} \right]$$

Donde a las submatrices o bloques se les ha denominado  $B, C, D, E$  y corresponden a:

$$B = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \end{bmatrix}, C = \begin{bmatrix} a_{14} \\ a_{24} \end{bmatrix}, D = [a_{31} \quad a_{32} \quad a_{33}], E = [a_{34}]$$

**Definición 3.3 (Matriz de Permutación)** Una matriz de permutación  $P$  es una matriz cuadrada de orden  $n$ , donde todos los elementos son 0, a excepción de uno cualquiera por cada fila y columna que tiene el valor 1

Existen  $n!$  matrices de permutación de tamaño  $n \times n$ . Las matrices de conmutación de orden  $n$  forman un grupo [Her86], cuyo elemento neutro es la matriz identidad de orden  $n$ , mientras que el elemento inverso es la transpuesta de la matriz dada.

**Ejemplo 3.5** Para  $n = 3$  se tienen las siguientes matrices de permutación:

$$\begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{bmatrix}$$

Se les llama matrices de permutación por cuanto al multiplicar por  $P$  otra matriz conforme  $A$ , da como resultado una matriz con los mismos elementos de la matriz original  $A$ , pero con sus posiciones permutadas de acuerdo con la ubicación de los elementos con valor 1 de  $P$ .

**Definición 3.4 (Matrices de Permutación Par e Impar)** A una matriz de permutación  $P$  que tiene determinante igual a 1 se le denomina matriz de permutación par, y a la que tiene determinante igual a  $-1$  se le denomina matriz de conmutación impar. La mitad de las matrices de permutación de tamaño  $n \times n$  son matrices de permutación pares, y la otra mitad son impares.

**Definición 3.5 (Matrices de Permutación en Bloques)** Una matriz de permutación por bloques de orden  $mn$ , que se denota por  $P_{m,n}$ , es una matriz cuadrada de orden  $mn$ , dividida en  $n \times m$  bloques o cajas cada una de ellas de orden  $m \times n$ , de manera que el bloque  $(i, j)$ ,  $i = 1, \dots, n$ ,  $j = 1, \dots, m$  tiene todos sus elementos nulos salvo el que está situado en su  $j$ -ésima fila e  $i$ -ésima columna que es igual a la unidad

Esta definición muestra que tanto el primer elemento como el último elemento de una matriz de permutación por bloques es 1.

**Proposición 3.1 (Propiedades Matrices de Permutación por Bloques)**

Las matrices de permutación en bloques cumplen las siguientes propiedades:

1.  $P_{m,1} = P_{1,m} = I_m$
2.  $P'_{m,n} = P_{n,m}$
3.  $P_{m,n}P_{n,m} = I_{nm}$
4. La matriz  $P_{m,n}$  es ortogonal

**Demostración.** Se va a demostrar la propiedad 4; la demostración de las demás propiedades se hacen de manera análoga. Las demostraciones se pueden consultar en [Bar98].

$P'_{m,n} = P_{n,m}$  entonces de (3)., resulta

$$P_{m,n}P_{n,m} = P_{n,m}P'_{m,n} = I_{n,m}$$

de donde se deduce que  $P_{m,n}$  es ortogonal ya que  $P_{m,n}^{-1} = P'_{m,n}$  ■

### 3.3. Producto de Kronecker

En ocasiones el producto de matrices que se asocia a la composición de aplicaciones lineales es insuficiente. El producto de Kronecker que a continuación se define, en cierto sentido lo generaliza.

**Definición 3.6 (Producto de Kronecker)** Sea  $A$  una matriz  $m \times n$  y  $B$  una matriz  $p \times q$ . El producto de Kronecker de la matriz  $A$  por la matriz  $B$ , denotado como  $A \otimes B$ , es la matriz bloque  $C$  de tamaño  $mp \times nq$  definida como:

$$C = A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ \vdots & \ddots & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix}$$

y desarrollando las operaciones implícitas en cada bloque  $a_{ij}B$ , se tiene para  $C = A \otimes B$  que

$$C = \begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & \cdots & a_{11}b_{1q} & \cdots & \cdots & a_{1n}b_{11} & a_{1n}b_{12} & \cdots & a_{1n}b_{1q} \\ a_{11}b_{21} & a_{11}b_{22} & \cdots & a_{11}b_{2q} & \cdots & \cdots & a_{1n}b_{21} & a_{1n}b_{22} & \cdots & a_{1n}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{11}b_{p1} & a_{11}b_{p2} & \cdots & a_{11}b_{pq} & \cdots & \cdots & a_{1n}b_{p1} & a_{1n}b_{p2} & \cdots & a_{1n}b_{pq} \\ \vdots & \vdots & & \vdots & \ddots & & \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots & & \ddots & \vdots & \vdots & & \vdots \\ a_{m1}b_{11} & a_{m1}b_{12} & \cdots & a_{m1}b_{1q} & \cdots & \cdots & a_{mn}b_{11} & a_{mn}b_{12} & \cdots & a_{mn}b_{1q} \\ a_{m1}b_{21} & a_{m1}b_{22} & \cdots & a_{m1}b_{2q} & \cdots & \cdots & a_{mn}b_{21} & a_{mn}b_{22} & \cdots & a_{mn}b_{2q} \\ \vdots & \vdots & \ddots & \vdots & & & \vdots & \vdots & \ddots & \vdots \\ a_{m1}b_{p1} & a_{m1}b_{p2} & \cdots & a_{m1}b_{pq} & \cdots & \cdots & a_{mn}b_{p1} & a_{mn}b_{p2} & \cdots & a_{mn}b_{pq} \end{bmatrix}$$

El producto de Kronecker también recibe el nombre de **Producto Tensorial** o **Producto Directo**.

Se llama producto de Kronecker, denotado con  $\otimes$ , a una operación sobre dos matrices de tamaño arbitrario que da como resultado una matriz bloque. El producto de Kronecker no debe confundirse con el producto de matrices habitual, que es una operación totalmente diferente. Debe su nombre al matemático alemán Leopold Kronecker.

$$\textbf{Ejemplo 3.6} \quad \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \\ a_{31} & a_{32} \end{bmatrix} \otimes \begin{bmatrix} b_{11} & b_{12} & b_{13} \\ b_{21} & b_{22} & b_{23} \end{bmatrix} =$$

$$\begin{bmatrix} a_{11}b_{11} & a_{11}b_{12} & a_{11}b_{13} & a_{12}b_{11} & a_{12}b_{12} & a_{12}b_{13} \\ a_{11}b_{21} & a_{11}b_{22} & a_{11}b_{23} & a_{12}b_{21} & a_{12}b_{22} & a_{12}b_{23} \\ a_{21}b_{11} & a_{21}b_{12} & a_{21}b_{13} & a_{22}b_{11} & a_{22}b_{12} & a_{22}b_{13} \\ a_{21}b_{21} & a_{21}b_{22} & a_{21}b_{23} & a_{22}b_{21} & a_{22}b_{22} & a_{22}b_{23} \\ a_{31}b_{11} & a_{31}b_{12} & a_{31}b_{13} & a_{32}b_{11} & a_{32}b_{12} & a_{32}b_{13} \\ a_{31}b_{21} & a_{31}b_{22} & a_{31}b_{23} & a_{32}b_{21} & a_{32}b_{22} & a_{32}b_{23} \end{bmatrix}$$

$$\textbf{Ejemplo 3.7} \quad \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \otimes \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} =$$

$$\begin{bmatrix} 1,5 & 1,6 & 2,5 & 2,6 \\ 1,7 & 1,8 & 2,7 & 2,8 \\ 3,5 & 3,6 & 4,5 & 4,6 \\ 3,7 & 3,8 & 4,7 & 4,8 \end{bmatrix} = \begin{bmatrix} 5 & 6 & 10 & 12 \\ 7 & 8 & 14 & 16 \\ 15 & 18 & 20 & 24 \\ 21 & 24 & 28 & 32 \end{bmatrix}$$

$$\textbf{Ejemplo 3.8} \quad \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \otimes \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} =$$

$$\begin{bmatrix} 5,1 & 5,2 & 6,1 & 6,2 \\ 5,3 & 5,4 & 6,3 & 6,4 \\ 7,1 & 7,2 & 8,1 & 8,2 \\ 7,3 & 7,4 & 8,3 & 8,4 \end{bmatrix} = \begin{bmatrix} 5 & 10 & 6 & 12 \\ 15 & 20 & 18 & 24 \\ 7 & 14 & 8 & 16 \\ 21 & 28 & 24 & 32 \end{bmatrix}$$

Con el ejemplo 3.8 se muestra que el producto de Kronecker no es conmutativo. En general,  $A \otimes B$  y  $B \otimes A$  son matrices diferentes. Sin embargo disfruta de otras propiedades, algunas de las cuales se recogen en la siguiente proposición.

**Proposición 3.2 (Propiedades del Producto Kronecker)** Sean,  $A_1, A_2 \in M_{m \times n}$  y  $B_1, B_2 \in M_{p \times q}$ . Para el producto Kronecker se verifican las siguientes propiedades:



$$1. (A_1 \otimes B_1) + (A_2 \otimes B_1) = (A_1 + A_2) \otimes B_1$$

$$(A_1 \otimes B_1) + (A_1 \otimes B_2) = A_1 \otimes (B_1 + B_2).$$

$$2. \text{ Dadas } A_{m \times n}, B_{p \times q} \text{ y } \alpha \in R, \text{ se verifica } (\alpha A \otimes B) = (A \otimes \alpha B) = \alpha (A \otimes B).$$

$$3. \text{ Dadas las matrices } A_{m \times n}, B_{p \times q}, C_{r \times s}, \text{ se verifica que } [(A \otimes B) \otimes C] = [A \otimes (B \otimes C)].$$

$$4. \text{ Dadas } A_1 \in M_{m \times n}, A_2 \in M_{m \times p}, B_1 \in M_{q \times r}, B_2 \in M_{r \times s}, \text{ si se consideran las matrices}$$

$$C_1 = A_1 \otimes B_1 \text{ y } C_2 = A_2 \otimes B_2 \text{ de ordenes } mq \times nr \text{ y } nr \times ps \text{ respectivamente, entonces}$$

$$C_1 C_2 = (A_1 \otimes B_1) (A_2 \otimes B_2) = A_1 A_2 \otimes B_1 B_2.$$

$$5. \text{ Dadas } A \text{ y } B \text{ matrices de ordenes } m \times n \text{ y } p \times q, \text{ respectivamente, en general } A \otimes B \neq B \otimes A.$$

$$\text{Esto quiere decir, el producto de Kronecker no es conmutativo, si bien se verifica que } P_{m,p} (A \otimes B) P_{q,n} = (B \otimes A)$$

$$6. \text{ Si } A \in M_n \text{ y } B \in M_m \text{ son dos matrices invertibles, entonces se verifica que } A \otimes B \text{ es invertible y su inversa}$$

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}.$$

$$7. \text{ Supuestas } A \text{ y } B \text{ dos matrices cualesquiera, se verifica que } (A \otimes B)^t = A^t \otimes B^t.$$

$$8. \text{ Dadas } A \text{ y } B \text{ dos matrices cuadradas de órdenes } m \text{ y } n, \text{ respectivamente, se verifica que } \text{tr}(A \otimes B) = \text{tr}(A) \cdot \text{tr}(B).$$

$$9. \text{ Sean, } A \in M_{n \times n} \text{ y } B \in M_{m \times m} \text{ entonces se verifica que } |A \otimes B| = |A|^n |B|^m.$$

$$10. \text{ Dadas } A \text{ y } B \text{ matrices cualesquiera, se tiene que } \text{rg}(A \otimes B) = \text{rg}(A) \cdot \text{rg}(B).$$

11. Si  $A$  es una matriz  $m \times n$  particionada en cuatro bloques  $A_{ij}$ ,  $i, j = 1, 2$  de dimensiones  $m_i \times n_j$ ,  $i, j = 1, 2$  con  $m_1 + m_2 = m$  y  $n_1 + n_2 = n$  y  $B$  es una matriz de orden  $p \times q$ , entonces

$$A \otimes B = \begin{pmatrix} A_{11} \otimes B & A_{12} \otimes B \\ A_{21} \otimes B & A_{22} \otimes B \end{pmatrix}$$

Se presenta ahora la demostración de algunas de las anteriores proposiciones.

**Demostración.** Si para  $k = 1, 2$  se tiene  $A_k = (a_{ij}^k)$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, n$ ,  $B_k = (b_{ij}^k)$ ,  $i = 1, \dots, p$ ,  $j = 1, \dots, q$  entonces como

$$A_1 \otimes B_1 = (a_{ij}^1 B_1)_{ij},$$

$$A_2 \otimes B_1 = (a_{ij}^2 B_1)_{ij}$$

resulta que

$$(A_1 \otimes B_1) + (A_2 \otimes B_1) = ((a_{ij}^1 + a_{ij}^2) B_1)_{ij} = (A_1 + A_2) \otimes B_1$$

Análogamente se comprueba la otra igualdad. ■

**Demostración.** Dado que

$$A \otimes B = \begin{bmatrix} a_{11}B & \cdots & a_{1n}B \\ a_{21}B & \cdots & a_{2n}B \\ \vdots & & \vdots \\ a_{m1}B & \cdots & a_{mn}B \end{bmatrix},$$

por definición de matriz traspuesta y producto de Kronecker resulta que

$$(A \otimes B)^t = \begin{bmatrix} a_{11}B^t & a_{21}B^t & \cdots & a_{m1}B^t \\ \vdots & \vdots & & \vdots \\ a_{1n}B^t & a_{2n}B^t & \cdots & a_{mn}B^t \end{bmatrix},$$

tal como se quería probar. ■

**Nota 3.2** Si se comparan las propiedades del producto ordinario de matrices y del producto Kronecker se tiene:

| Producto Matricial                                      | Producto Kronecker   |
|---|--|
| $(AB)^t = B^t A^t$                                      | $(A \otimes B)^t = A^t \otimes B^t$                        |
| $(AB)^{-1} = B^{-1} A^{-1}$                             | $(A \otimes B)^{-1} = A^{-1} \otimes B^{-1}$               |
| $\text{tr}(AB) \neq \text{tr}(A) \text{tr}(B)$          | $\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B)$       |
| $ AB  =  A   B $  | $ A \otimes B  =  A ^m  B ^n$                              |
| $\text{rg}(AB) \leq \min\{\text{rg}(A), \text{rg}(B)\}$ | $\text{rg}(A \otimes B) = \text{rg}(A) \cdot \text{rg}(B)$ |

**Ejemplo 3.9** Las matrices:

$$A = \begin{bmatrix} 1 & -1 \\ 2 & 1 \end{bmatrix} \quad B = \begin{bmatrix} -1 & 1 & 1 \\ 1 & 0 & -1 \\ 0 & 2 & 4 \end{bmatrix}$$

son tales que  $\text{tr}(A) = 2, \text{tr}(B) = 3; \text{rg}(A) = 2, \text{rg}(B) = 3; |A| = 3, |B| = -4$

Entonces para la matriz:

$$A \otimes B = \begin{bmatrix} -1 & 1 & 1 & 1 & -1 & -1 \\ 1 & 0 & -1 & -1 & 0 & 1 \\ 0 & 2 & 4 & 0 & -2 & -4 \\ -2 & 2 & 2 & -1 & 1 & 1 \\ 2 & 0 & -2 & 1 & 0 & -1 \\ 0 & 4 & 8 & 0 & 2 & 4 \end{bmatrix}$$

Sin necesidad de trabajar con ella directamente, en virtud de las propiedades del producto de Kronecker se pueden calcular los siguientes valores:

$$\text{tr}(A \otimes B) = \text{tr}(A) \text{tr}(B) = 6$$

$$\text{rg}(A \otimes B) = \text{rg}(A) \cdot \text{rg} B = 6$$

$$|A \otimes B| = |A|^3 |B|^2 = 3^2 (-4)^2 = 432$$

Además como  $\text{rg}(A \otimes B) = 6$ , la matriz  $A \otimes B$  es invertible, siendo su inversa, de acuerdo con la propiedad 6:

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

por tanto, como:

$$A^{-1} = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} \\ -\frac{2}{3} & \frac{1}{3} \end{bmatrix} \quad y \quad B^{-1} = \begin{bmatrix} -\frac{1}{2} & \frac{1}{2} & \frac{1}{4} \\ 1 & 1 & 0 \\ -\frac{1}{2} & -\frac{1}{2} & \frac{1}{4} \end{bmatrix}$$

se tiene que:

$$(A \otimes B)^{-1} = \begin{bmatrix} -\frac{1}{6} & \frac{1}{6} & \frac{1}{12} & -\frac{1}{6} & \frac{1}{6} & \frac{1}{12} \\ \frac{1}{3} & \frac{1}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ -\frac{1}{6} & -\frac{1}{6} & \frac{1}{12} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{12} \\ \frac{1}{3} & -\frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{6} & \frac{1}{12} \\ -\frac{2}{3} & -\frac{2}{3} & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ \frac{1}{3} & \frac{1}{3} & -\frac{1}{6} & -\frac{1}{6} & -\frac{1}{6} & \frac{1}{12} \end{bmatrix}$$

Un ámbito en el que puede resultar útil expresar una matriz como un vector columna es el de la derivación de matrices o expresiones matriciales. Por ello, a continuación se define lo que se entiende por vectorización de una matriz y se analizan algunas de sus propiedades.

**Definición 3.7 (Vectorización de una Matriz)** *La vectorización de una matriz es una transformación lineal que convierte una matriz en un vector columna. Dada una matriz  $A$  de orden  $m \times n$ , la vectorización de  $A$  es el vector columna de  $mn$  elementos que se obtiene escribiendo las columnas de  $A$  una a continuación de otra y se denota por  $\text{vec}(A)$ . Entonces:*

$$\text{vec}(A) = \begin{pmatrix} a_{\bullet 1} \\ a_{\bullet 2} \\ \vdots \\ a_{\bullet n} \end{pmatrix}, \text{ con } a_{\bullet j} = \begin{pmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{pmatrix}, j = 1, \dots, n.$$

**Ejemplo 3.10** *Dada la matriz*

$$A = \begin{pmatrix} 1 & -2 & -1 \\ -5 & 1 & 0 \\ 3 & 0 & 2 \end{pmatrix} \text{ se tiene que } \text{vec}(A) = \begin{pmatrix} 1 \\ -5 \\ 3 \\ -2 \\ 1 \\ 0 \\ -1 \\ 0 \\ 2 \end{pmatrix}$$

**Proposición 3.3 (Propiedades de la Vectorización)** *La vectorización de matrices verifica las siguientes propiedades:*

1. Sean  $A, B \in M_{m \times n}$  dos matrices cualesquiera, y sea  $\alpha \in R$ . Se verifica las siguientes propiedades:  $\text{vec}(A + B) = \text{vec}(A) + \text{vec}(B)$

$$\text{vec}(\alpha A) = \alpha \text{vec}(A)$$

2. Dadas las matrices  $A_{m \times n}$  y  $B_{n \times q}$  se verifica que:

$$\text{vec}(AB) = (B^T \otimes I_m) \text{vec}(A) = (I_p \otimes A) \text{vec}(B) = (B^T \otimes A) \text{vec}(I_n)$$

3. Dada la matriz  $A$  de orden  $m \times n$ , se verifica que:

$$\text{vec}(A) = (I_n \otimes A) \text{vec}(I_n) = (A^T \otimes I_m) \text{vec}(I_m)$$

4. Dadas las matrices  $A_{m \times n}$ ,  $B_{n \times p}$  y  $C_{p \times q}$ , se verifica que:

$$\text{vec}(ABC) = (C' \otimes A) \text{vec}(B) = (I_q \otimes AB) \text{vec}(C) = (C' B' \otimes I_m) \text{vec}(A)$$

5. Dada la matriz  $A$  de orden  $m \times n$ , se verifica:

$$\text{vec}(A) = P_{m,n} \text{vec}(A'), \text{vec}(A') = P_{n,m} \text{vec}(A)$$

6. Dada la matriz  $A$  de orden  $m \times n$ , se verifica:

$$[(\text{vec} I_n)' \otimes I_m] [I_n \otimes \text{vec}(A)] = A$$

$$[\text{vec}(A)' \otimes I_m] [I_n \otimes \text{vec}(I_m)] = A$$

$$[I_m \otimes (\text{vec}(A'))'] [\text{vec}(I_m) \otimes I_n] = A. [I_m \otimes (\text{vec}(I_n))'] [\text{vec}(A') \otimes I_n] = A$$

7. Dadas las matrices  $A$  y  $B$  de orden  $m \times n$  y  $C, D$  de orden  $m \times p$ , se tiene que:

$$\text{vec}[(A+B)(C+D)] = [(I_p \otimes A) + (I_p \otimes B)] [\text{vec}(C) + \text{vec}(D)]$$

$$\text{vec}[(A+B)(C+D)] = [(C' \otimes I_m) + (D' \otimes I_m)] [\text{vec}(A) + \text{vec}(B)]$$

8. Sean las matrices  $A_{m \times n}$  y  $B_{n \times m}$ , entonces se tiene que

$$\text{tr}(AB) = (\text{vec}(A'))' \text{vec}(B) = (\text{vec}(B))' \text{vec}(A)$$

En particular si  $n = m$  y  $B = I_n$ :

$$\text{tr}(A) = (\text{vec}(A'))' \text{vec}(I_n) = (\text{vec}(I_n))' \text{vec}(A)$$

**Demostración.** La  $j$ -ésima columna del producto  $AB$  es  $Ab_{\bullet j}$ , entonces por definición de vectorización y el producto de matrices particionadas de tiene

$$\text{vec} AB = \begin{pmatrix} Ab_{\bullet 1} \\ Ab_{\bullet 2} \\ \vdots \\ Ab_{\bullet p} \end{pmatrix} = \begin{pmatrix} A & 0_{m \times n} & \cdots & 0 \\ 0 & A & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A \end{pmatrix} \begin{pmatrix} b_{\bullet 1} \\ b_{\bullet 2} \\ \vdots \\ b_{\bullet p} \end{pmatrix}$$

Ahora bien, teniendo en cuenta el producto Kronecker de matrices, esto se puede expresar como

$$vec(AB) = (I_p \otimes A) \cdot vec(B)$$

Por otra parte, la  $j$ -ésima columna de  $AB$  es  $\sum_{i=1}^n a_{\bullet i} b_{ij}$

pues la componenete  $k$ -ésima de este vector coincide con el elemento de la  $k$ -ésima fila y la  $j$ -ésima columna de  $AB$  que es  $\sum_{i=1}^n a_{ki} b_{ij}$

Por tanto:

$$vec(AB) = \begin{pmatrix} \sum_{i=1}^n a_{\bullet i} b_{i1} \\ \sum_{i=1}^n a_{\bullet i} b_{i2} \\ \vdots \\ \sum_{i=1}^n a_{\bullet i} b_{ip} \end{pmatrix} = \begin{pmatrix} B' & 0_{p \times n} & \cdots & 0 \\ 0 & B' & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B' \end{pmatrix} \begin{pmatrix} a_{\bullet 1} \\ a_{\bullet 2} \\ \vdots \\ a_{\bullet n} \end{pmatrix}$$

y por definición de producto de Kronecker y vectorización, resulta:

$$vec(AB) = (B' \otimes I) vec(A) \blacksquare$$

**Demostración.** Como  $A = I_m A$  por la proposición 3.3,2:

$$vec(A) = vec(I_m A) = (A' \otimes I_m) vec(I_m)$$

Análogamente, como  $A = A I_n$ , de nuevo por la proposición 3.3,2:

$$vec(A) = vec(A I_n) = (I_n \otimes A) vec(I_n) \blacksquare$$

**Demostración.** De acuerdo con la proposición 3.3,2 se verifica que:

$$vec(ABC) = vec[(AB)C] = (I_q \otimes AB) vec(C)$$

$$vec(ABC) = vec[(AB)C] = ((BC)' \otimes I_m) vec(A)$$

$$vec(ABC) = vec[(AB)C] = (C'B' \otimes I_m) vec(A)$$

Además, como en virtud de la proposición 3.3, 1, se tiene también que

$$vec(ABC) = vec[(AB)C] = (C' \otimes AB) vec(I_p)$$

entonces por la proposición 3.3, 4 del producto de Kronecker,

$$(C' \otimes AB) = (C' \otimes A)(I_p \otimes B)$$

y, teniendo en cuenta la proposición 3.3, 3, se obtiene finalmente

$$\text{vec}(ABC) = (C' \otimes A) (I_p \otimes B) \text{vec}(I_p) = (C' \otimes A) \text{vec}(B) \blacksquare$$

Las proposiciones 3.3, 2 y 8 pueden generalizarse para el producto de un número finito de matrices. Así, para matrices  $A_1, A_2, A_3$  y  $A_4$  de dimensiones adecuadas, se tiene que:

$$\begin{aligned} \text{vec}(A_1 A_2 A_3 A_4) &= (I \otimes A_1 A_2 A_3) \text{vec}(A_4) \\ \text{vec}(A_1 A_2 A_3 A_4) &= (A_4' \otimes A_1 A_2) \text{vec}(A_3) \\ \text{vec}(A_1 A_2 A_3 A_4) &= (A_4 A_3' \otimes A_1) \text{vec}(A_2) \\ \text{vec}(A_1 A_2 A_3 A_4) &= (A_4 A_3 A_2' \otimes I) \text{vec}(A_1) \end{aligned}$$

y

$$\begin{aligned} \text{tr}(A_1 A_2 A_3) &= (\text{vec}(A_1'))' (A_3' \otimes I) \text{vec}(A_2) \\ \text{tr}(A_1 A_2 A_3) &= (\text{vec}(A_1'))' (I \otimes A_2) \text{vec}(A_3) \\ \text{tr}(A_1 A_2 A_3) &= (\text{vec}(A_2'))' (I \otimes A_3) \text{vec}(A_1) \\ \text{tr}(A_1 A_2 A_3) &= (\text{vec}(A_2'))' (A_1' \otimes I) \text{vec}(A_3) \\ \text{tr}(A_1 A_2 A_3) &= (\text{vec}(A_3'))' (A_2' \otimes I) \text{vec}(A_1) \\ \text{tr}(A_1 A_2 A_3) &= (\text{vec}(A_3'))' (I \otimes A_1) \text{vec}(A_2) \end{aligned}$$

que se pueden demostrar fácilmente, pues son consecuencia de la vectorización y traza del producto de dos matrices y de las propiedades del producto de Kronecker. Resultados análogos existen para el producto de un número finito de matrices.

En algunas ocasiones, cuando se vectoriza una matriz, conviene hacerlo a partir de sus filas.

**Definición 3.8 (Vectorización por Filas)** Dada  $A$  una matriz de orden  $m \times n$ , se define:

$$\overline{\text{vec}}(A) = \begin{pmatrix} a_{1\bullet} \\ a_{2\bullet} \\ \vdots \\ a_{m\bullet} \end{pmatrix} \text{ con } a_{i\bullet} = (a_{i1}, \dots, a_{in}), i = 1, \dots, m$$

Es obvio y fácil de verificar que  $\text{vec}(A) = \overline{\text{vec}}(A')$ . Por tanto, pueden enunciarse las propiedades de  $\overline{\text{vec}}(A)$ . Por ejemplo:

$$\overline{\text{vec}}(AB) = \overline{\text{vec}}((AB)') = \text{vec}(B'A) = [(A \otimes I) \text{vec}(B)]$$

$$\overline{\text{vec}}(AB) = (A \otimes I) \overline{\text{vec}}(B)$$

y

$$tr(AB) = (vec(A))' vec(B) = (\overline{vec}(A))' vec(B')$$

$$tr(AB) = (vec(A))' \overline{vec}(B)$$

**Definición 3.9** ( $\widetilde{vec}$ ) Dada una matriz simétrica  $A$  de orden  $n$ , puede definirse una vectorización diferente de las ya indicadas de manera que en dicha vectorización, solo se recojan los elementos distintos de la matriz  $A$ . Así, se denota por  $\widetilde{vec}(A)$  el vector columna de orden  $r = \frac{n(n+1)}{2}$  dado por:

$$\widetilde{vec}(A) = \begin{pmatrix} a_{11} \\ a_{12} \\ a_{22} \\ a_{13} \\ a_{23} \\ a_{33} \\ \vdots \\ a_{1n} \\ a_{2n} \\ \vdots \\ a_{nn} \end{pmatrix}$$

En  $\widetilde{vec}(A)$  se incluyen los elementos situados por encima de la diagonal principal de cada una de las columnas de  $A$ . Por otra parte, la relación entre esta vectorización y la indicada en la definición 3.7 se establece por la igualdad  $\widetilde{vec}(A) = Qvec(A)$ , (donde  $Q$  es la matriz diagonal por bloques de orden  $r \times n$ ):

$$Q = \begin{bmatrix} Q_1 & & & \\ & Q_2 & & \\ & & \ddots & \\ & & & Q_n \end{bmatrix}$$

donde  $i = 1, \dots, n$ , y el bloque  $Q_i$  una matriz de orden  $i \times n$  dada por

$$Q_i = \begin{bmatrix} 1 & 0 & \cdots & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \ddots & & \ddots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \end{bmatrix}$$

i-ésima columna



**Nota 3.3** La vectorización se utiliza en conjunto con el producto Kronecker para expresar la multiplicación de matrices como una transformación lineal sobre las matrices. Por ejemplo, si  $A, B, X$  son matrices conformes:

$$\text{vec}(AXB) = (B^T \otimes A)\text{vec}(X)$$

### 3.4. Fórmulas de Derivación Matricial.

La derivada de una función matricial de variable matricial no difiere sustancialmente del concepto de derivada habitual. La diferencia y dificultad se presenta cuando es preciso manipular estas derivadas matriciales, debido a su dimensión y la ubicación de sus elementos [Bar98].

En esta sección se indican las definiciones de funciones escalares, vectoriales y matriciales de variable escalar, vectorial o matricial, adoptando un criterio común para la colocación de los distintos elementos que componen cada una de estas derivadas. En las definiciones de las derivadas de funciones escalar o vectorial, bien sean de variable escalar o vectorial, existe casi unanimidad en la literatura en cuanto a su ordenación.

Se acostumbra usar la variable en negrilla cuando se trata de una variable de tipo vectorial o matricial; sin embargo, en las siguientes secciones no se utilizará esta notación ya que el contexto hará claridad. Así, si  $x \in R^n$  e  $y \in R^m$ , entonces se puede observar que una notación bastante usual se expone así:

- Si  $y = y(x_1, \dots, x_n)$  su derivada es  $\frac{\partial y}{\partial x} = \left( \frac{\partial y}{\partial x_1} \dots \frac{\partial y}{\partial x_n} \right)$ , que es el caso de las funciones escalares de variable vectorial.

- Si  $y = y(x) = \begin{pmatrix} y_1(x) \\ \vdots \\ y_n(x) \end{pmatrix}$  su derivada es  $\frac{\partial y}{\partial x} = \begin{pmatrix} \frac{\partial y_1(x)}{\partial x} \\ \vdots \\ \frac{\partial y_n(x)}{\partial x} \end{pmatrix}$ , que es el caso de las funciones vectoriales de variable escalar.

- Si  $y = y(x) = \begin{pmatrix} y_1(x_1, \dots, x_n) \\ \vdots \\ y_n(x_1, \dots, x_n) \end{pmatrix}$  su derivada es  $\frac{\partial y}{\partial x} = \begin{pmatrix} \frac{\partial y_1}{\partial x_1} & \dots & \frac{\partial y_1}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_m}{\partial x_1} & \dots & \frac{\partial y_m}{\partial x_n} \end{pmatrix}$ , que es el caso de las funciones matriciales de variable vectorial

Cuando se ha de calcular el gradiente de  $y = f(x)$ , se respeta la ordenación del vector fila  $x = (x_1, \dots, x_n)$  de las variables explicativas; luego parece lógico que este mismo criterio de ubicación de elementos se adopte para definir la derivada de una función escalar de variable matricial.[Bar98]. Siguiendo este acuerdo se exponen las siguientes definiciones.

**Definición 3.10** Sea  $y = y(x)$  una función real de variable vectorial, es decir,  $y$  es una función definida de  $\mathbb{R}^n$  en  $\mathbb{R}$ . La derivada de  $y$  respecto de  $x$  es el vector fila  $\frac{\partial y(x)}{\partial x} = \left( \frac{\partial y(x)}{\partial x_1}, \dots, \frac{\partial y(x)}{\partial x_n} \right)$

Se adopta esta notación ya que se está derivando un escalar  $y(x)$  con respecto a un vector fila  $x = (x_1, \dots, x_n)$ . Obsérvese que  $\frac{\partial y(x)}{\partial x}$  es el vector gradiente de la función  $y(x)$ .

Si se precisa que el vector de variables  $x$  respecto a las que se efectúa la derivada de  $y$  se toma como columna, entonces se denotará por

$$\frac{\partial y(x)}{\partial x} = \left( \frac{\partial y(x)}{\partial x} \right)',$$

es decir, este vector coincide con el traspuesto del vector gradiente de  $y$ .

**Definición 3.11** Sea  $y = y(x)$  una función vectorial de variable real, es decir

$$y(x) = \begin{pmatrix} y_1(x) \\ \vdots \\ y_n(x) \end{pmatrix}$$

es una función definida de  $\mathbb{R}$  en  $\mathbb{R}^m$ . La derivada de  $y$  respecto del escalar  $x$  es el vector columna

$$\frac{\partial y(x)}{\partial x} = \begin{pmatrix} \frac{\partial y_1(x)}{\partial x} \\ \vdots \\ \frac{\partial y_m(x)}{\partial x} \end{pmatrix}$$

**Definición 3.12** Sea  $y = y(x)$  una función vectorial de variable vectorial, es decir, una función definida de  $\mathbb{R}^n$  en  $\mathbb{R}^m$ . Como

$$y(x) = \begin{pmatrix} y_1(x) \\ \vdots \\ y_m(x) \end{pmatrix} \text{ entonces:}$$

$$\frac{\partial y(x)}{\partial x} = \begin{pmatrix} \frac{\partial y_1(x)}{\partial x} \\ \vdots \\ \frac{\partial y_m(x)}{\partial x} \end{pmatrix} = \begin{pmatrix} \frac{\partial y_1(x)}{\partial x_1} & \dots & \frac{\partial y_1(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_m(x)}{\partial x_1} & \dots & \frac{\partial y_m(x)}{\partial x_n} \end{pmatrix}$$

Obsérvese que esta matriz de orden  $m \times n$  es el Jacobiano de la función  $y(x)$ , ya que cada una de sus filas es el gradiente de las funciones  $y_1(x), \dots, y_n(x)$ , que son las componentes de  $y(x)$ .

**Definición 3.13** Sea  $y = y(X)$  una función real de variable matricial, es decir, la función  $y$  está definida de  $M_{pn}$  en  $\mathbb{R}$ . La derivada de  $y$  respecto de la matriz  $X$  de orden  $p \times n$ , es la matriz del mismo orden dada por

$$\frac{\partial y(X)}{\partial X} = \begin{pmatrix} \frac{\partial y(X)}{\partial x_{11}} & \dots & \frac{\partial y(X)}{\partial x_{1n}} \\ \vdots & & \vdots \\ \frac{\partial y(X)}{\partial x_{p1}} & \dots & \frac{\partial y(X)}{\partial x_{pn}} \end{pmatrix} = \left( \frac{\partial y(X)}{\partial x_{ij}} \right)_{\substack{i=1,\dots,p \\ j=1,\dots,n}}$$

**Definición 3.14** Sea  $Y = Y(x)$  una función matricial de variable escalar, es decir la función  $Y$  está definida de  $\mathbb{R}$  en  $M_{mq}$ . La derivada de  $Y = Y(x)$  respecto de la variable  $x$  está dada por la matriz de orden  $m \times q$

$$\frac{\partial Y(x)}{\partial x} = \begin{pmatrix} \frac{\partial y_{11}(x)}{\partial x} & \dots & \frac{\partial y_{1q}(x)}{\partial x} \\ \vdots & & \vdots \\ \frac{\partial y_{m1}(x)}{\partial x} & \dots & \frac{\partial y_{mq}(x)}{\partial x} \end{pmatrix} = \left( \frac{\partial y_{ij}(x)}{\partial x} \right)_{\substack{i=1,\dots,m \\ j=1,\dots,q}}$$

**Definición 3.15** Sea  $y = y(X)$  una función vectorial de variable matricial definida de  $M_{pn}$  en  $\mathbb{R}$ . La derivada de  $y(X)$  respecto de la matriz  $A$  de orden  $p \times n$  es la matriz de orden  $pn \times n$  dada por

$$\frac{\partial y(X)}{\partial X} = \begin{pmatrix} \frac{\partial y_1(X)}{\partial X} \\ \vdots \\ \frac{\partial y_m(X)}{\partial X} \end{pmatrix},$$

ya que, de acuerdo con la definición 3.13, para cada  $k = 1, \dots, m$

$$\frac{\partial y_k(X)}{\partial X} = \left( \frac{\partial y_k(X)}{\partial x_{ij}} \right)_{\substack{i=1,\dots,p \\ j=1,\dots,n}}$$

**Definición 3.16** Sea  $Y = Y(x)$  una función matricial de variable vectorial definida de  $\mathbb{R}^n$  en  $M_{mq}$ . La derivada de  $Y(x)$  respecto del vector  $x$  es la matriz de orden  $m \times qn$  cuya expresión es

$$\frac{\partial Y(x)}{\partial x} = \begin{pmatrix} \frac{\partial y_{11}(x)}{\partial x} & \dots & \frac{\partial y_{1q}(x)}{\partial x} \\ \vdots & & \vdots \\ \frac{\partial y_{m1}(x)}{\partial x} & \dots & \frac{\partial y_{mq}(x)}{\partial x} \end{pmatrix}$$

donde, teniendo en cuenta la definición 3.10,

$$\frac{\partial y_{ij}(x)}{\partial x} = \left( \frac{\partial y_{ij}(x)}{\partial x_1}, \dots, \frac{\partial y_{ij}(x)}{\partial x_n} \right)$$

con  $i = 1, \dots, m, j = 1, \dots, q$

**Definición 3.17 (Derivada Funcion Matricial de Variable Matricial)**

Sea  $Y = Y(X)$  una función matricial de variable matricial, es decir, definida de  $M_{pn}$  en  $M_{mq}$ . La derivada de  $Y(X)$  respecto de la matriz  $X$  está dada por la matriz de orden  $pm \times nq$

$$\frac{\partial Y(X)}{\partial X} = \begin{pmatrix} \frac{\partial y_{11}(X)}{\partial X} & \dots & \frac{\partial y_{1q}(X)}{\partial X} \\ \vdots & & \vdots \\ \frac{\partial y_{m1}(X)}{\partial X} & \dots & \frac{\partial y_{mq}(X)}{\partial X} \end{pmatrix}$$

ya que, teniendo en cuenta la definición 3.14, para cada  $i = 1, \dots, m, j = 1, \dots, q$

$$\frac{\partial y_{ij}(X)}{\partial X} = \left( \frac{\partial y_{ij}(X)}{\partial x_{k\ell}} \right)_{\substack{k=1, \dots, p \\ \ell=1, \dots, n}}$$

La definición 3.17 hace referencia al caso general que se puede tratar en el caso de la derivación matricial. Este caso se convierte en los casos particulares de las definiciones 3.10 hasta la definición 3.16 si se dan las siguientes condiciones:

1.  $p = q = m = q$  se obtiene la definición 3.10
2.  $p = n = q = 1$  se obtiene la definición 3.11
3.  $p = q = 1$  se obtiene la definición 3.12
4.  $q = m = 1$  se obtiene la definición 3.13
5.  $p = n = 1$  se obtiene la definición 3.14
6.  $q = 1$  se obtiene la definición 3.15
7.  $p = 1$  se obtiene la definición 3.16

Existen múltiples formas de dar la expresión de derivadas matriciales, ya que pueden considerarse distintas ordenaciones de los elementos de la matriz final que se obtiene al derivar una matriz respecto de otra. Algunas derivadas matriciales se pueden definir a través de la vectorización de la matriz de variables dependientes e incluso de la de variables independientes.

1. En el presente trabajo, empleando el concepto de vectorización de una matriz, pueden obtenerse las siguientes expresiones alternativas de las definiciones 3.13, 3.14, 3.15, 3.16 y 3.17 en términos de vectorización:

$$\bullet \frac{\partial y(X)}{\partial \text{vec}(X)} = \begin{pmatrix} \frac{\partial y(X)}{\partial x_{11}} \\ \vdots \\ \frac{\partial y(X)}{\partial x_{p1}} \\ \vdots \\ \frac{\partial y(X)}{\partial x_{1n}} \\ \vdots \\ \frac{\partial y(X)}{\partial x_{pn}} \end{pmatrix} = \text{vec} \left( \frac{\partial y(X)}{\partial (X)} \right)$$

$$\bullet \frac{\partial \text{vec} Y(x)}{\partial x} = \begin{pmatrix} \frac{\partial y_{11}(x)}{\partial x_{11}} \\ \vdots \\ \frac{\partial y_{1m}(x)}{\partial x} \\ \vdots \\ \frac{\partial y_{1q}(x)}{\partial x} \\ \vdots \\ \frac{\partial y_{mq}(x)}{\partial x} \end{pmatrix} = \text{vec} \left( \frac{\partial Y(x)}{\partial x} \right)$$

$$\bullet \frac{\partial y(X)}{\partial \text{vec}(X)} = \begin{pmatrix} \frac{\partial y_1(X)}{\partial \text{vec}(X)} \\ \vdots \\ \frac{\partial y_m(X)}{\partial \text{vec}(X)} \end{pmatrix},$$

siendo  $\frac{\partial y_i(X)}{\partial \text{vec}(X)}$  para  $i = 1, \dots, m$  lo indicado en la primera expresión alternativa.

El vector columna  $\frac{\partial y(X)}{\partial \text{vec}(X)}$  de orden  $mnp$ , teniendo en cuenta propiedades del producto de Kronecker y de la vectorización, se puede expresar también como

$$\frac{\partial y(X)}{\partial \text{vec}(X)} = \left( \frac{\partial y(X)}{\partial X'} \otimes I_p \right) (\text{vec} I_p), \text{siendo}$$

$$\frac{\partial y(X)}{\partial X'} = \left( \frac{\partial y(X)}{\partial X} \right)'$$

Nótese que en este caso  $\frac{\partial y(X)}{\partial \text{vec}(X)} \neq \text{vec}\left(\frac{\partial y(X)}{\partial X}\right)$

$$\bullet \frac{\partial \text{vec} Y(x)}{\partial x} = \begin{pmatrix} \frac{\partial y_{11}(x)}{\partial x} \\ \vdots \\ \frac{\partial y_{m1}(x)}{\partial x} \\ \vdots \\ \frac{\partial y_{1q}(x)}{\partial x} \\ \vdots \\ \frac{\partial y_{mq}(x)}{\partial x} \end{pmatrix} = \begin{pmatrix} \frac{\partial y_{11}(x)}{\partial x_1} & \dots & \frac{\partial y_{11}(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_{m1}(x)}{\partial x_1} & \dots & \frac{\partial y_{m1}(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_{1q}(x)}{\partial x_1} & \dots & \frac{\partial y_{1q}(x)}{\partial x_n} \\ \vdots & & \vdots \\ \frac{\partial y_{mq}(x)}{\partial x_1} & \dots & \frac{\partial y_{mq}(x)}{\partial x_n} \end{pmatrix}$$

Se observa que  $\frac{\partial \text{vec}(Y(x))}{\partial x}$  es una matriz de orden  $mq \times n$ , mientras que

$\frac{\partial Y(x)}{\partial x}$  es de orden  $m \times qn$ . Entre estas dos matrices puede establecerse la siguiente relación:

$$\frac{\partial \text{vec}(Y(x))}{\partial x} = \left( I_q \otimes \frac{\partial Y(x)}{\partial x} \right) (\text{vec}(I_q) \otimes I_n)$$

Mediante las definiciones anteriores y las expresiones en términos de la vectorización y las propiedades del producto de Kronecker se obtienen las siguientes formas alternativas para el cálculo de  $\frac{\partial Y(X)}{\partial X}$ :

$$\bullet \frac{\partial \text{vec}(Y(X))}{\partial X} = \begin{pmatrix} \frac{\partial y_{11}(X)}{\partial X} \\ \vdots \\ \frac{\partial y_{m1}(X)}{\partial X} \\ \vdots \\ \frac{\partial y_{1q}(X)}{\partial X} \\ \vdots \\ \frac{\partial y_{mq}(X)}{\partial X} \end{pmatrix} = \left( I_q \otimes \frac{\partial Y(X)}{\partial X} \right) (\text{vec}(I_q) \otimes I_n)$$

$$\bullet \frac{\partial Y(X)}{\partial \text{vec} X} = \begin{pmatrix} \frac{\partial y_{11}(X)}{\partial \text{vec}(X)} & \dots & \frac{\partial y_{1q}(X)}{\partial \text{vec}(X)} \\ \vdots & & \vdots \\ \frac{\partial y_{m1}(X)}{\partial \text{vec}(X)} & \dots & \frac{\partial y_{mq}(X)}{\partial \text{vec}(X)} \end{pmatrix} = \left( \frac{\partial Y(X)}{\partial X'} \otimes I_p \right) (I_q \text{vec}(I_p))$$

$$\bullet \frac{\partial \text{vec}(Y(X))}{\partial \text{vec}(X)} = \text{vec} \left( \frac{\partial Y(X)}{\partial \text{vec} X} \right) = \left( I_q \otimes \frac{\partial Y(X)}{\partial X'} \otimes I_p \right) (\text{vec}(I_q) \otimes \text{vec}(I_p))$$

Alternativamente, es posible otra formulación para la derivada de expresiones matriciales. Para ello, es necesario considerar para cualquier  $m, n, p, q \in N$  y cualquier función:

$$Y(X) : M_{pn} \longrightarrow M_{mq}$$

la definición

$$\frac{\partial Y(X)}{\partial X} \equiv \frac{\partial [Y(X)]}{\partial (\text{vec} X)'}.$$

Retomando lo expresado anteriormente, en este trabajo se identificarán explícitamente los espacios de matrices  $M_{pn}$  y  $M_{mq}$  con  $\mathbb{R}^{pn}$  y  $\mathbb{R}^{mq}$ , respectivamente. Lo que se está considerando generalmente es  $Y(X)$  como función vectorial de variable vectorial definida de  $\mathbb{R}^{pn}$  en  $\mathbb{R}^{mq}$ , por lo que la expresión

$$\frac{\partial \text{vec}[Y(X)]}{\partial (\text{vec} X)'}$$

es el Jacobiano de dicha función vectorial. La elección de esta definición de derivada puede resultar útil para aplicar a funciones matriciales resultados matemáticos relativos a funciones vectoriales.

**Ejemplo 3.11** Dada la función  $y = y(x)$  de  $\mathbb{R}^2$  en  $\mathbb{R}^2$  definida por

$$y(x) = \begin{pmatrix} y_1(x_1, x_2) \\ y_2(x_1, x_2) \end{pmatrix} = \begin{pmatrix} x_1^2 + x_2^3 \\ 2x_1x_2 \end{pmatrix},$$

de acuerdo con la Definición 3.12 se tiene que

$$\frac{\partial y(x)}{\partial x} = \begin{pmatrix} \frac{\partial y_1(x)}{\partial x} \\ \frac{\partial y_2(x)}{\partial x} \end{pmatrix} = \begin{pmatrix} \frac{\partial y_1(x)}{\partial x_1} & \frac{\partial y_1(x)}{\partial x_2} \\ \frac{\partial y_2(x)}{\partial x_1} & \frac{\partial y_2(x)}{\partial x_2} \end{pmatrix} = \begin{pmatrix} 2x_1 & 3x_2^2 \\ 2x_2 & 2x_1 \end{pmatrix}$$

Por otra parte, teniendo en cuenta la Definición 3.16

$$\frac{\partial}{\partial x} = \begin{pmatrix} \frac{\partial y(x)}{\partial x} \end{pmatrix} = \begin{pmatrix} 2 & 0 & 0 & 6x_2 \\ 0 & 2 & 2 & 0 \end{pmatrix}$$

Observese que el resultado anterior difiere de

$$\frac{\partial \left[ \text{vec} \left( \frac{\partial y(x)}{\partial x} \right) \right]}{\partial x} = \frac{\partial}{\partial x} \left[ \begin{pmatrix} 2x_1 \\ 2x_2 \\ 3x_2^2 \\ 2x_1 \end{pmatrix} \right] = \begin{pmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 6x_2 \\ 2 & 0 \end{pmatrix}$$

**Ejemplo 3.12** Dada la matriz  $A = (a_{ij})$ ,  $i = 1, \dots, m$ ,  $j = 1, \dots, q$ , se verifica que la derivada de  $A$  respecto del primer vector fila que se denota por  $a_{1\bullet}$  es la matriz de orden  $m \times q^2$

$$\begin{aligned} \frac{\partial A}{\partial a_{1\bullet}} &= \begin{pmatrix} \frac{\partial a_{11}}{\partial a_{1\bullet}} & \dots & \frac{\partial a_{1q}}{\partial a_{1\bullet}} \\ \vdots & \vdots & \vdots \\ \frac{\partial a_{m1}}{\partial a_{1\bullet}} & \dots & \frac{\partial a_{mq}}{\partial a_{1\bullet}} \end{pmatrix} \\ &= \left( \begin{array}{cccc|cccc|ccc} 1 & 0 & \dots & 0 & 0 & 1 & \dots & 0 & \dots & 0 & 0 & \dots & 1 \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots & \vdots & \vdots & & \vdots & & \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 0 & 0 & 0 & \dots & 0 & \dots & 0 & 0 & \dots & 0 \end{array} \right) \end{aligned}$$

Haciendo uso de la expresión alternativa se obtiene

$$\frac{\partial \text{vec}(A)}{\partial a_{1\bullet}} = \left( I_q \otimes \frac{\partial A}{\partial a_{1\bullet}} \right) (\text{vec}(I_q) \otimes I_q) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ \hline 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0 \\ \hline \vdots & \vdots & \vdots & & \vdots \\ \hline 0 & 0 & 0 & \dots & 1 \\ 0 & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

La derivada de la matriz  $A$  respecto de sí misma viene dada por la matriz de orden  $m^2 \times q^2$



$$\frac{\partial A}{\partial A} = \begin{pmatrix} \frac{\partial a_{11}}{\partial A} & \cdots & \frac{\partial a_{1q}}{\partial A} \\ \vdots & & \vdots \\ \frac{\partial a_{m1}}{\partial A} & \cdots & \frac{\partial a_{mq}}{\partial A} \end{pmatrix} = (\text{vec} I_m)(\text{vec} I_q)'$$

para cada  $i = 1, \dots, m$ ,  $j = 1, \dots, q$  se tiene que

$$\frac{\partial a_{ij}}{\partial A} = \begin{pmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \leftarrow i - \text{ésima fila}$$

$\uparrow$   
 y  $j$ -ésima columna.

La derivada de la matriz  $A'$  respecto de  $A$  es una matriz cuadrada de orden  $mq$  definida como

$$\frac{\partial A'}{\partial A} = \begin{pmatrix} \frac{\partial a_{11}}{\partial A} & \cdots & \frac{\partial a_{m1}}{\partial A} \\ \vdots & & \vdots \\ \frac{\partial a_{1q}}{\partial A} & \cdots & \frac{\partial a_{mq}}{\partial A} \end{pmatrix} = P, m, q,$$

para cada  $i = 1, \dots, m$ ,  $j = 1, \dots, q$

$$\frac{\partial a_{ij}}{\partial A} = \begin{pmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \leftarrow j - \text{ésima fila}$$

$\uparrow$   
 y la  $j$ -ésima columna

**Ejemplo 3.13** Sea  $y = y(A)$  una función definida de  $M_{2 \times 2}$  en  $\mathbb{R}^2$  dada por

$$y = y(A) = \begin{pmatrix} y_1(A) \\ y_2(A) \end{pmatrix} = \begin{pmatrix} a_{11}^2 + 2a_{12} - a_{22}^2 \\ (a_{11} + a_{22})^2 - 2a_{12}a_{21} \end{pmatrix}$$

$$\text{donde } A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

La derivada de la función  $y = y(A)$  respecto de la matriz  $A$  de variables independientes, de acuerdo con la definición 3.15 es

$$\begin{aligned}\frac{\partial y(A)}{\partial A} &= \left( \frac{\frac{\partial y_1(A)}{\partial A}}{\frac{\partial y_2(A)}{\partial A}} \right) = \left( \frac{\frac{\frac{\partial y_1(A)}{\partial a_{11}}}{\frac{\partial y_1(A)}{\partial a_{21}}} \quad \frac{\frac{\partial y_1(A)}{\partial a_{12}}}{\frac{\partial y_1(A)}{\partial a_{22}}}}{\frac{\frac{\partial y_2(A)}{\partial a_{11}}}{\frac{\partial y_2(A)}{\partial a_{21}}} \quad \frac{\frac{\partial y_2(A)}{\partial a_{12}}}{\frac{\partial y_2(A)}{\partial a_{22}}}} \right) \\ &= \begin{pmatrix} 2a_{11} & 2 \\ 0 & -2a_{22} \\ \hline 2a_{11} + 2a_{22} & -2a_{21} \\ -2a_{12} & 2a_{11} + 2a_{22} \end{pmatrix},\end{aligned}$$

y la derivada de  $y(A)$  respecto a la vectorización de  $A$ , teniendo en cuenta la definición 3.17, tiene la expresión

$$\begin{aligned}\frac{\partial y(A)}{\partial \text{vec}(A)} &= \left( \frac{\frac{\partial y_1(A)}{\partial \text{vec}(A)}}{\frac{\partial y_2(A)}{\partial \text{vec}(A)}} \right) = \begin{pmatrix} \frac{\frac{\partial y_1(A)}{\partial a_{11}}}{\frac{\partial y_1(A)}{\partial a_{21}}} \\ \frac{\frac{\partial y_1(A)}{\partial a_{12}}}{\frac{\partial y_1(A)}{\partial a_{22}}} \\ \frac{\frac{\partial y_2(A)}{\partial a_{11}}}{\frac{\partial y_2(A)}{\partial a_{21}}} \\ \frac{\frac{\partial y_2(A)}{\partial a_{12}}}{\frac{\partial y_2(A)}{\partial a_{22}}} \end{pmatrix} = \begin{pmatrix} 2a_{11} \\ 0 \\ 2 \\ -2a_{22} \\ 2a_{11} + 2a_{22} \\ -2a_{12} \\ -2a_{21} \\ 2a_{11} + 2a_{22} \end{pmatrix} \\ &= \begin{pmatrix} 2a_{11} & 0 & 0 & 0 \\ 0 & 2a_{11} & 0 & 0 \\ 2 & 0 & -2a_{22} & 0 \\ 0 & 2 & 0 & -2a_{22} \\ 2a_{11} + 2a_{22} & 0 & -2a_{12} & 0 \\ 0 & 2a_{11} + 2a_{22} & 0 & -2a_{12} \\ -2a_{21} & 0 & 2a_{11} + 2a_{22} & 0 \\ 0 & -2a_{21} & 0 & 2a_{11} + 2a_{22} \end{pmatrix} \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \left[ \begin{pmatrix} 2a_{11} & 0 \\ 2 & -2a_{22} \\ 2a_{11} + 2a_{22} & -2a_{21} \\ -2a_{21} & 2a_{11} + 2a_{22} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right] \begin{pmatrix} 1 \\ 0 \\ 0 \\ 1 \end{pmatrix} \\ &= \left[ \left( \frac{\partial y(A)}{\partial A'} \otimes I_2 \right) \right] \text{vec}(I_2)\end{aligned}$$

**Ejemplo 3.14** Sea la función  $Y = Y(X)$  definida de  $M_{2 \times 2}$  en  $M_{3 \times 2}$  por

$$Y = Y(X) = \begin{pmatrix} y_{11}(X) & y_{12}(X) \\ y_{21}(X) & y_{22}(X) \\ y_{31}(X) & y_{32}(X) \end{pmatrix} = \begin{pmatrix} x_{11} + x_{12} & 2x_{12} \\ x_{22}^2 & e^{x_{21}} \\ x_{11} + x_{12} & x_{21}x_{12} \end{pmatrix},$$

siendo

$$X = \begin{pmatrix} \frac{y_{11}(X)}{\partial X} & \frac{y_{12}(X)}{\partial X} \\ \frac{y_{21}(X)}{\partial X} & \frac{y_{22}(X)}{\partial X} \\ \frac{y_{31}(X)}{\partial X} & \frac{y_{32}(X)}{\partial X} \end{pmatrix} = \left( \begin{array}{cc|cc} 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 \\ 0 & 2x_{22} & e^{x_{21}} & 0 \\ \hline x_{22} & 0 & 0 & x_{21} \\ 0 & x_{11} & x_{12} & 0 \end{array} \right).$$

Ahora bien, de acuerdo con la expresión alternativa

$$\begin{aligned} \frac{\partial \text{vec} Y(X)}{\partial X} &= \left( I_2 \otimes \frac{\partial Y(X)}{\partial X} \right) (\text{vec}(I_2) \otimes I_2) \\ &= \begin{pmatrix} 1 & 1 & 0 & 2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2x_{22} & e^{x_{21}} & 0 & 0 & 0 & 0 & 0 \\ x_{22} & 0 & 0 & x_{21} & 0 & 0 & 0 & 0 \\ 0 & x_{11} & x_{12} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 2x_{22} & e^{x_{21}} & 0 \\ 0 & 0 & 0 & 0 & x_{22} & 0 & 0 & x_{21} \\ 0 & 0 & 0 & 0 & 0 & x_{11} & x_{12} & 0 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \begin{pmatrix} 1 & 1 \\ 0 & 0 \\ 0 & 0 \\ 0 & 2x_{22} \\ x_{22} & 0 \\ 0 & x_{11} \\ 0 & 2 \\ 0 & 0 \\ 0 & 0 \\ e^{x_{21}} & 0 \\ 0 & x_{21} \\ x_{12} & 0 \end{pmatrix} \end{aligned}$$

### 3.5. Reglas de Derivación.

La derivada de un vector o de una matriz respecto de un escalar, vector o matriz, no difiere sustancialmente de la derivación habitual de funciones. Sin embargo esto no conlleva que sean válidas, en general las reglas usuales para derivar, por ejemplo, las de suma, el producto o la composición de funciones.

Ahora bien, como se verá a continuación, si se hace uso del producto Kronecker en lugar del producto matricial estandar, en algunos casos es posible obtener reglas de derivación análogas a las habituales para funciones.

**Proposición 3.4** *Sea  $C$  una matriz de variables independientes de orden  $p \times n$ , esto es,  $C \in M_{pn}$  :*

1. *Derivada del Producto por Escalar. Sea  $A$  una matriz de orden  $m \times q$  cuyos elementos dependen de los de la matriz  $C$  de orden  $p \times n$ . Entonces, para cualquier  $\lambda \in R$ , se verifica que*

$$\frac{\partial (\lambda A)}{\partial C} = \lambda \cdot \frac{\partial A}{\partial C}$$

2. *Derivada de la Suma. Sean  $A$  y  $B$  dos matrices de orden  $m \times q$ , cuyos elementos dependen de los de la matriz  $C$  de orden  $p \times n$ . Se verifica que*

$$\frac{\partial (A + B)}{\partial C} = \frac{\partial A}{\partial C} + \frac{\partial B}{\partial C}.$$

3. *Derivada de Producto. Sean  $A, B$  y  $C$  matrices de ordenes  $m \times q$ ,  $q \times r$ ,  $p \times n$ , respectivamente, tales que los elementos de  $A$  y  $B$  son funciones de los elementos de la matriz  $C$ . Entonces se verifica que*

- a) *Si  $p = n = 1$  ó  $m = q = 1$*

$$\frac{\partial AB}{\partial C} = \frac{\partial A}{\partial C} \cdot B + A \cdot \frac{\partial B}{\partial C}$$

- b) *Si  $p = 1$ , es decir,  $C$  es un vector fila de  $\mathbb{R}^n$ , entonces*

$$\frac{\partial AB}{\partial C} = \frac{\partial A}{\partial C} (B \otimes I_n) + A \cdot \frac{\partial B}{\partial C}$$

- c) *Si  $p \neq 1$  y  $n \neq 1$*

$$\frac{\partial AB}{\partial C} = \frac{\partial A}{\partial C} (B \otimes I_n) + (A \otimes I_p) \cdot \frac{\partial B}{\partial C}$$

4. *Derivada del Producto de Kronecker.* Sean  $A, B$  y  $C$  matrices de ordenes  $m \times q$ ,  $t \times r$ ,  $p \times n$ , respectivamente, tales que los elementos de  $A$  y  $B$  son función de los elementos de la matriz  $C$ . Entonces se verifica que

$$\frac{\partial A \otimes B}{\partial C} = \left( A \otimes \frac{\partial B}{\partial C} \right) + (I_m \otimes B \otimes I_p) \frac{\partial (A \otimes I_r)}{\partial C}$$

donde

$$\frac{\partial (A \otimes I_r)}{\partial C} = (P_{r,m} \otimes I_p) \left( I_r \otimes \frac{\partial A}{\partial C} \right) (P_{q,r} \otimes I_n)$$

5. *Regla de la Cadena.* Sea  $y = y(A)$  una función real de variable matricial, donde  $A$  es una matriz de orden  $t \times r$  cuyos elementos a su vez son funciones de la matriz  $C$  de orden  $p \times n$ . Se verifica

$$\frac{\partial y(A)}{\partial C} = \left( \frac{\partial y(A)}{\partial \text{vec}(A)} \otimes I_p \right)' \left( \frac{\partial \text{vec}(A)}{\partial C} \right)$$

o también

$$\frac{\partial y(A)}{\partial C} = [(\text{vec}(I_t))' \otimes I_p] \left( \frac{\partial y(A)}{\partial A} \otimes \frac{\partial A}{\partial C} \right) (\text{vec}(I_r) \otimes I_n).$$

**Demostración.** La derivada  $\frac{\partial \lambda A}{\partial C}$  de acuerdo con la definición 3.17 es una matriz de orden  $mp \times nq$  dada por

$$\frac{\partial \lambda A}{\partial C} = \left( \frac{\partial \lambda a_{ij}}{\partial C} \right)_{\substack{i=1,\dots,m \\ j=1,\dots,q}}$$

y como para cada  $i = 1, \dots, m$ ;  $j = 1, \dots, q$ , según la definición 3.10,

$$\frac{\partial \lambda a_{ij}}{\partial C} = \frac{\partial \lambda a_{ij}}{\partial c_{k\ell}} = \lambda \left( \frac{\partial a_{ij}}{\partial c_{k\ell}} \right)_{\substack{k=1,\dots,p \\ \ell=1,\dots,n}}$$

se obtiene que  $\frac{\partial \lambda A}{\partial C} = \lambda \left( \frac{\partial A}{\partial C} \right)$ . ■

**Demostración.** La derivada  $\frac{\partial (A + B)}{\partial C}$  según la definición 3.17, es la matriz de orden  $mp \times nq$

$$\frac{\partial (A + B)}{\partial C} = \left( \frac{\partial (a_{ij} + b_{ij})}{\partial C} \right)_{\substack{i=1,\dots,m \\ j=1,\dots,q}}$$

Ahora bien, para cada  $i = 1, \dots, m$  ;  $j = 1, \dots, q$ , a partir de la definición 3.13 se obtiene

$$\frac{\partial (a_{ij} + b_{ij})}{\partial C} = \frac{\partial a_{ij}}{\partial C} + \frac{\partial b_{ij}}{\partial C}$$

y sumando por bloques se puede concluir

$$\frac{\partial (A + B)}{\partial C} = \frac{\partial A}{\partial C} + \frac{\partial B}{\partial C} \quad \blacksquare$$

### **Demostración.**

a) Si  $p = n = 1$ , entonces

$$\frac{\partial AB}{\partial C} = \frac{\partial A}{\partial C} B + A \frac{\partial B}{\partial C},$$

ya que para cada  $i = 1, \dots, m$  ;  $j = 1, \dots, r$  la derivada del elemento  $d_{ij}$  de la matriz  $D = AB$  es

$$\begin{aligned} \frac{\partial d_{ij}(C)}{\partial C} &= \frac{\partial}{\partial C} \left[ \sum_{k=1}^q a_{ik}(C) b_{kj}(C) \right] \\ \frac{\partial d_{ij}(C)}{\partial C} &= \sum_{k=1}^q \frac{\partial a_{ik}(C)}{\partial C} b_{kj}(C) + \sum_{k=1}^q a_{ik}(C) \frac{\partial b_{kj}(C)}{\partial C} \end{aligned}$$

b) En el caso en que  $n = q = r = 1$ , según la definición 3.13 se tiene

$$\begin{aligned} \frac{\partial AB}{\partial C} &= \begin{pmatrix} \frac{\partial AB}{\partial c_{11}} & \dots & \frac{\partial AB}{\partial c_{1n}} \\ \vdots & & \vdots \\ \frac{\partial AB}{\partial c_{p1}} & \dots & \frac{\partial AB}{\partial c_{pn}} \end{pmatrix} \\ \frac{\partial AB}{\partial C} &= \begin{pmatrix} \frac{\partial A}{\partial c_{11}} & \dots & \frac{\partial A}{\partial c_{1n}} \\ \vdots & & \vdots \\ \frac{\partial A}{\partial c_{p1}} & \dots & \frac{\partial A}{\partial c_{pn}} \end{pmatrix} B + A \begin{pmatrix} \frac{\partial B}{\partial c_{11}} & \dots & \frac{\partial B}{\partial c_{1n}} \\ \vdots & & \vdots \\ \frac{\partial B}{\partial c_{p1}} & \dots & \frac{\partial B}{\partial c_{pn}} \end{pmatrix}, \end{aligned}$$

dado que para cada  $i = 1, \dots, p$  ;  $j = 1, \dots, n$

$$\frac{\partial (AB)}{\partial c_{ij}} = \frac{\partial A}{\partial c_{ij}} B + A \frac{\partial B}{\partial c_{ij}},$$

y por tanto,

$$\frac{\partial (AB)}{\partial C} = \frac{\partial A}{\partial C} B + A \frac{\partial B}{\partial C}$$

c) Ahora, si  $p \neq 1$ ,  $n \neq 1$  y  $D = AB = (d_{ij})$   $i = 1, \dots, m$  ;  $j = 1, \dots, r$ , se tiene que

$$\frac{\partial d_{ij}(C)}{\partial C} = \sum_{k=1}^q \frac{a_{ik}(C)}{\partial C} b_{kj}(C) + \sum_{k=1}^q a_{ik}(C) \frac{\partial b_{kj}(C)}{\partial C},$$

de donde

$$\begin{aligned} \frac{\partial d_{ij}(C)}{\partial C} &= \left( \frac{\partial a_{i1}}{\partial C} \dots \frac{\partial a_{iq}}{\partial C} \right) b_{ij}(C) I_n \begin{pmatrix} b_{1j}(C) I_n \\ b_{2j}(C) I_n \\ \vdots \\ b_{qj}(C) I_n \end{pmatrix} \\ &+ (a_{i1}(C) I_p \dots a_{in}(C) I_p) \begin{pmatrix} \frac{\partial b_{1j}(C)}{\partial C} \\ \vdots \\ \frac{\partial b_{qj}(C)}{\partial C} \end{pmatrix} \end{aligned}$$

y por la definición de producto Kronecker y la definición 3.12, resulta

$$\frac{\partial d_{ij}(C)}{\partial C} = \frac{\partial a_{i\bullet}}{\partial C} (b_{\bullet j}(C) \otimes I_n) + (a_{i\bullet}(C) \otimes I_p) \frac{\partial b_{\bullet j}(C)}{\partial C},$$

y, por consiguiente,

$$\frac{\partial D}{\partial C} = \frac{\partial (AB)}{\partial C} = \frac{\partial A}{\partial C} (B \otimes I_n) + (A \otimes I_p) \frac{\partial B}{\partial C}.$$

Si  $p = 1$  como  $I_p = 1$ , la expresión anterior se transforma en

$$\frac{\partial (AB)}{\partial C} = \frac{\partial A}{\partial C} (B \otimes I_n) + A \frac{\partial B}{\partial C} \blacksquare$$

**Demostración.** Atendiendo a la definición de producto Kronecker

$$\frac{\partial (A \otimes B)}{\partial C} = \frac{\partial a_{ij} B}{\partial C}, \substack{i=1, \dots, m, \\ j=1, \dots, q},$$

y por la definición 3.17, para cada  $i = 1, \dots, m$  ;  $j = 1, \dots, q$  se tiene que

$$\frac{\partial a_{ij} B}{\partial C} = \frac{\partial (a_{ij} b_{k\ell})}{\partial C} \substack{k=1, \dots, t, \\ \ell=1, \dots, r}.$$

Ahora bien, en virtud del apartado anterior,

$$\frac{\partial a_{ij} b_{k\ell}}{\partial C} = \frac{\partial a_{ij}}{\partial C} b_{k\ell} + a_{ij} \frac{\partial b_{k\ell}}{\partial C}.$$

Así pues,

$$\frac{\partial a_{ij} B}{\partial C} = \begin{pmatrix} \frac{\partial a_{ij}}{\partial C} b_{11} + a_{ij} \frac{\partial b_{11}}{\partial C} & \cdots & \frac{\partial a_{ij}}{\partial C} b_{1r} + a_{ij} \frac{\partial b_{1r}}{\partial C} \\ \vdots & & \vdots \\ \frac{\partial a_{ij}}{\partial C} b_{t1} + a_{ij} \frac{\partial b_{t1}}{\partial C} & \cdots & \frac{\partial a_{ij}}{\partial C} b_{tr} + a_{ij} \frac{\partial b_{tr}}{\partial C} \end{pmatrix}$$

o lo que es equivalente

$$\frac{\partial a_{ij} B}{\partial C} = \left( B \otimes \frac{\partial a_{ij}}{\partial C} \right) + a_{ij} \frac{\partial B}{\partial C}.$$

Por tanto,

$$\frac{\partial (A \otimes B)}{\partial C} = \begin{pmatrix} \left( B \otimes \frac{\partial a_{11}}{\partial C} \right) + \partial a_{11} \frac{\partial B}{\partial C} & \cdots & \left( B \otimes \frac{\partial a_{1q}}{\partial C} \right) + \partial a_{1q} \frac{\partial B}{\partial C} \\ \vdots & & \vdots \\ \left( B \otimes \frac{\partial a_{m1}}{\partial C} \right) + \partial a_{m1} \frac{\partial B}{\partial C} & \cdots & \left( B \otimes \frac{\partial a_{mq}}{\partial C} \right) + \partial a_{mq} \frac{\partial B}{\partial C} \end{pmatrix}$$

$$\frac{\partial (A \otimes B)}{\partial C} = \left( A \otimes \frac{\partial B}{\partial C} \right) + \begin{pmatrix} B \otimes \frac{\partial a_{11}}{\partial C} & \cdots & B \otimes \frac{\partial a_{1q}}{\partial C} \\ \vdots & & \vdots \\ B \otimes \frac{\partial a_{m1}}{\partial C} & \cdots & B \otimes \frac{\partial a_{mq}}{\partial C} \end{pmatrix}$$

$$\frac{\partial (A \otimes B)}{\partial C}$$

$$\textbf{Solución 3.1} = \left( A \otimes \frac{\partial B}{\partial C} \right) + (I_m \otimes B \otimes I_q) \frac{\partial (A \otimes I_r)}{\partial C}$$



Ahora bien, a partir de la propiedad 5 del producto de Kronecker, se tiene

$$A \otimes I_r = P_{r,m} (I_r \otimes A) P_{q,r}$$

y aplicando dos veces la regla de derivación del producto indicada en la proposición 3.4,3 se obtiene

$$\frac{\partial (A \otimes I_r)}{\partial C} = (P_{r,m} \otimes I_p) \frac{\partial (I_r \otimes A)}{\partial C} (P_{q,r} \otimes I_n),$$

lo que de acuerdo con la solución 3.1, tomando  $A = I_r$  y  $B = A$  conduce a

$$\frac{\partial (A \otimes I_r)}{\partial C} = (P_{r,m} \otimes I_p) \left( I_r \otimes \frac{\partial A}{\partial C} \right) (P_{q,r} \otimes I_n). \blacksquare$$

**Demostración.** Teniendo en cuenta la definición 3.13,

$$\frac{\partial y(A)}{\partial C} = \frac{\partial y(A)}{\partial c_{ij}}, \begin{matrix} i=1,\dots,p \\ j=1,\dots,n \end{matrix}$$

y por la regla de la cadena para el caso unidimensional

$$\frac{\partial y(A)}{\partial c_{ij}} = \sum_{k=1}^t \sum_{l=1}^r \frac{\partial y(A)}{\partial a_{kl}} \frac{\partial a_{kl}(C)}{\partial c_{ij}}.$$

Por tanto,

$$\frac{\partial y(A)}{\partial C} = \begin{pmatrix} \frac{\partial y}{\partial a_{11}} & \dots & \frac{\partial y}{\partial a_{1n}} \\ \vdots & & \vdots \\ \frac{\partial y}{\partial a_{p1}} & \dots & \frac{\partial y}{\partial a_{pn}} \end{pmatrix} = \sum_{k=1}^t \sum_{l=1}^r \left[ \frac{\partial y(A)}{\partial a_{kl}} \begin{pmatrix} \frac{\partial a_{kl}(C)}{\partial c_{11}} & \dots & \frac{\partial a_{kl}(C)}{\partial c_{1n}} \\ \vdots & & \vdots \\ \frac{\partial a_{kl}(C)}{\partial c_{p1}} & \dots & \frac{\partial a_{kl}(C)}{\partial c_{pn}} \end{pmatrix} \right]$$

de donde

$$\frac{\partial y(A)}{\partial C} = \sum_{k=1}^t \sum_{l=1}^r \frac{\partial y(A)}{\partial a_{kl}} \cdot \frac{\partial a_{kl}(C)}{\partial C}$$

ahora bien, como

$$\left( \frac{\partial y(A)}{\partial \text{vec}(A)} \right)' = \left( \frac{\partial y}{\partial a_{11}}, \dots, \frac{\partial y}{\partial a_{t1}}, \dots, \frac{\partial y}{\partial a_{1r}}, \dots, \frac{\partial y}{\partial a_{tr}} \right)$$

y

$$\frac{\partial \text{vec}(A)}{\partial C} = \begin{pmatrix} \frac{\partial a_{11}}{\partial C} \\ \vdots \\ \frac{\partial a_{t1}}{\partial C} \\ \vdots \\ \frac{\partial a_{1r}}{\partial C} \\ \vdots \\ \frac{\partial a_{tr}}{\partial C} \end{pmatrix}$$

$$\text{resulta } \frac{\partial y(A)}{\partial C} = \left[ \left( \frac{\partial y(A)}{\partial \text{vec}(A)} \right)' \otimes I_p \right] \cdot \frac{\partial \text{vec}(A)}{\partial C}$$

$$\textbf{Solución 3.2} = \left( \frac{\partial y(A)}{\partial \text{vec}(A)} \otimes I_p \right)' \cdot \frac{\partial \text{vec}(A)}{\partial C}$$

Para obtener la expresión alternativa de la regla de la cadena que se indica en el enunciado, hasta tener en cuenta, que de acuerdo con la propiedad 3 de la vectorización

$$\frac{\partial y(A)}{\partial \text{vec}(A)} = \text{vec} \left( \frac{\partial y(A)}{\partial A} \right) = \left[ \left( \frac{\partial y(A)}{\partial C} \right)' \otimes I_t \right] \text{vec}(I_t),$$

lo que sustituido en la solución 3.2 permite obtener

$$\frac{\partial y(A)}{\partial C} = \left( \text{vec} \left( \frac{\partial y(A)}{\partial A} \right) \otimes I_p \right)' \left( \frac{\partial \text{vec}(A)}{\partial C} \right)$$

$$\frac{\partial y(A)}{\partial C} = \left( \left\{ \left[ \left( \frac{\partial y(A)}{\partial A} \right)' \otimes I_t \right] \text{vec}(I_t) \right\} \otimes I_p \right)' \cdot \frac{\partial \text{vec}(A)}{\partial C}$$

$$\frac{\partial y(A)}{\partial C} = \left( \left[ \left( \frac{\partial y(A)}{\partial A} \right)' \otimes I_t \right] \text{vec}(I_t) \right)' \otimes I_p \cdot \frac{\partial \text{vec}(A)}{\partial C}$$

$$\frac{\partial y(A)}{\partial C} = \left( \left[ (\text{vec}(I_t))' \left( \frac{\partial y(A)}{\partial A} \otimes I_t \right) \right] \otimes I_p \right) \cdot \frac{\partial \text{vec}(A)}{\partial C},$$

y tomando  $B_2 = I_p$  en la propiedad 4 del producto de Kronecker se tiene

$$\frac{\partial y(A)}{\partial C} = ((\text{vec}(I_t))' \otimes I_p) \left( \frac{\partial y(A)}{\partial A} \otimes I_{tp} \right) \cdot \frac{\partial \text{vec}(A)}{\partial C}.$$

Finalmente de acuerdo con las formas alternativas, como

$$\frac{\partial \text{vec}(A)}{\partial C} = \left( I_r \otimes \frac{\partial A}{\partial C} \right) (\text{vec}(I_r) \otimes I_n)$$

sustituyendo en  $\frac{\partial \text{vec}(A)}{\partial C}$  y haciendo uso de nuevo de la propiedad 4 del producto de Kronecker se tiene

$$\frac{\partial \text{vec}(A)}{\partial C} = ((\text{vec}(I_t))' \otimes I_p) \left( \frac{\partial y(A)}{\partial A} \otimes \frac{\partial A}{\partial C} \right) (\text{vec}(I_r) \otimes I_n) \quad \blacksquare$$

**Ejemplo 3.15** Dadas las matrices

$$A = \begin{pmatrix} x^2 & y & xy \\ x+1 & y^3 & e^x \end{pmatrix}, B = \begin{pmatrix} (y+1)^2 & 2y & x+y \\ e^{x+y} & 0 & 1 \end{pmatrix} y$$

$$C = \begin{pmatrix} x & z \\ w & y \end{pmatrix}$$

$$\text{como } \frac{\partial A}{\partial C} = \begin{pmatrix} \frac{\partial x^2}{\partial x} & \frac{\partial x^2}{\partial z} & \frac{\partial y}{\partial x} & \frac{\partial y}{\partial z} & \frac{\partial xy}{\partial x} & \frac{\partial xy}{\partial z} \\ \frac{\partial x^2}{\partial w} & \frac{\partial x^2}{\partial y} & \frac{\partial y}{\partial w} & \frac{\partial y}{\partial y} & \frac{\partial xy}{\partial w} & \frac{\partial xy}{\partial y} \\ \frac{\partial(x+1)}{\partial x} & \frac{\partial(x+1)}{\partial z} & \frac{\partial y^3}{\partial x} & \frac{\partial y^3}{\partial z} & \frac{\partial e^z}{\partial x} & \frac{\partial e^z}{\partial z} \\ \frac{\partial(x+1)}{\partial w} & \frac{\partial(x+1)}{\partial y} & \frac{\partial y^3}{\partial w} & \frac{\partial y^3}{\partial y} & \frac{\partial e^z}{\partial w} & \frac{\partial e^z}{\partial y} \end{pmatrix}$$

$$= \begin{pmatrix} 2x & 0 & 0 & 0 & y & 0 \\ 0 & 0 & 0 & 1 & 0 & x \\ 1 & 0 & 0 & 0 & e^x & 0 \\ 0 & 0 & 0 & 3y^2 & 0 & 0 \end{pmatrix}$$

$$y \frac{\partial B}{\partial C} = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 2(y+1) & 0 & 2 & 0 & 1 \\ e^{x+y} & 0 & 0 & 0 & 0 & 0 \\ 0 & e^{x+y} & 0 & 0 & 0 & 0 \end{pmatrix}$$

se tiene

$$\frac{\partial(A+B)}{\partial C} = \frac{\partial A}{\partial C} + \frac{\partial B}{\partial C} = \begin{pmatrix} 2x & 0 & 0 & 0 & 1+y & 0 \\ 0 & 2(y+1) & 0 & 3 & 0 & 1+x \\ 1+e^{x+y} & 0 & 0 & 0 & e^x & 0 \\ 0 & e^{x+y} & 0 & 3y^2 & 0 & 0 \end{pmatrix}$$

Asi mismo se verifica que

$$\frac{\partial(3A)}{\partial C} = 3 \left( \frac{\partial A}{\partial C} \right) = \begin{pmatrix} 6x & 0 & 0 & 0 & 3y & 0 \\ 0 & 0 & 0 & 3 & 0 & 3x \\ 3 & 0 & 0 & 0 & 3e^x & 0 \\ 0 & 0 & 0 & 9y^2 & 0 & 0 \end{pmatrix}$$

**Ejemplo 3.16** Dadas las matrices  $A$  y  $C$  del ejemplo anterior y el vector

$$b = \begin{pmatrix} x+y \\ 2x \\ yz \end{pmatrix}$$

$$\text{teniendo en cuenta } \frac{\partial Ab}{\partial C} = \frac{\partial A}{\partial C} (b \otimes I_2) + (A \otimes I_2) \frac{\partial b}{\partial C}$$

$$\text{como ya se conoce } \frac{\partial A}{\partial C} \text{ y, } \frac{\partial B}{\partial C} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \\ 0 & 0 \\ 0 & y \\ 0 & z \end{pmatrix}$$

sustituyendo en la expresión inicial resulta

$$\begin{aligned} \frac{\partial Ab}{\partial C} &= \begin{pmatrix} 2x & 0 & 0 & 0 & y & 0 \\ 0 & 0 & 0 & 1 & 0 & x \\ 1 & 0 & 0 & 0 & e^x & 0 \\ 0 & 0 & 0 & 3y^2 & 0 & 0 \end{pmatrix} \begin{pmatrix} x+y & 0 \\ 0 & x+y \\ 2x & 0 \\ 0 & 2x \\ yz & 0 \\ 0 & yz \end{pmatrix} \\ &+ \begin{pmatrix} x^2 & 0 & y & 0 & xy & 0 \\ 0 & x^2 & 0 & y & 0 & xy \\ x+1 & 0 & y^3 & 0 & e^x & 0 \\ 0 & x+1 & 0 & y^3 & 0 & e^x \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 2 & 0 \\ 0 & 0 \\ 0 & y \\ 0 & z \end{pmatrix} \\ &= \begin{pmatrix} 3x^2 + 2xy + y^2z + 2y & xy^2 \\ 0 & x^2 + 2xyz + 2x \\ 2x + y + 2y^3 + yze^x + 1 & ye^x \\ 0 & 6y^2x + x + ze^x + 1 \end{pmatrix} \end{aligned}$$

Este resultado coincide con:

$$\frac{\partial Ab}{\partial C} = \frac{\partial}{\partial C} \begin{pmatrix} x^3 + x^2y + 2xy + xy^2z \\ x^2 + xy + x + y + 2xy^3 + yze^x \end{pmatrix}$$

**Ejemplo 3.17** Dadas las matrices  $A$ ,  $B$  y  $C$  definidas por

$$A = \begin{pmatrix} x+y & 2z \\ xy & x+z \\ e^x & y^3 \end{pmatrix}, B = \begin{pmatrix} 2y & z^2 \\ 1 & 3z \end{pmatrix}, C = \begin{pmatrix} x \\ y \\ z \end{pmatrix}$$

se verifica, de acuerdo con el apartado cuarto de la proposición anterior

$$\frac{\partial(A \otimes B)}{\partial C} = \left( A \otimes \frac{\partial B}{\partial C} \right) + (I_3 \otimes B \otimes I_3) \frac{\partial(A \otimes I_2)}{\partial C}$$

siendo en este caso

$$\frac{\partial B}{\partial C} = \begin{pmatrix} 0 & 0 \\ 2 & 0 \\ 0 & 2z \\ 0 & 0 \\ 0 & 0 \\ 0 & 3 \end{pmatrix}$$

$$\frac{\partial(A \otimes I_2)}{\partial C} = (P_{2,3} \otimes I_3) \left( I_2 \otimes \frac{\partial A}{\partial C} \right) P_{2,2}$$

donde

$$P_{2,3} = \left( \begin{array}{ccc|ccc} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ \hline 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ \hline 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{array} \right), P_{2,2} = \left( \begin{array}{cc|cc} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ \hline 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right)$$

$$\frac{\partial A}{\partial C} = \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 2 \\ y & 1 \\ x & 0 \\ 0 & 1 \\ e^x & 1 \\ 0 & 3y^2 \\ 0 & 0 \end{pmatrix}$$

Así pues

$$\frac{\partial(A \otimes B)}{\partial C} = \begin{pmatrix} 2y & z^2 & 0 & 0 \\ 2x+4y & z^2 & 4z & 0 \\ 0 & 2xz+2yz & 4y & 6z^2 \\ 1 & 3z & 0 & 0 \\ 1 & 3z & 0 & 0 \\ 0 & 3x+3y & 2 & 12z \\ 2y^2 & yz^2 & 2y & z^2 \\ 4xy & xz^2 & 2x+2z & 0 \\ 0 & 2xyz & 2y & 2xz+3z^2 \\ y & 3yz & 1 & 3z \\ x & 3xz & 0 & 0 \\ 0 & 3xy & 1 & 3x+6z \\ 2ye^x & z^2e^x & 0 & 0 \\ 2e^x & 0 & 8y^3 & 3y^2z^2 \\ 0 & 2ze^x & 0 & 2y^3z \\ e^x & 3ze^x & 0 & 0 \\ 0 & 0 & 3y^2 & 9y^2z \\ 0 & 3e^x & 0 & 3y^3 \end{pmatrix}$$

**Ejemplo 3.18** Si  $y = y(A) = |A| = xy - zw$  donde  $A = \begin{pmatrix} x & w \\ z & y \end{pmatrix}$  y cada

elemento de  $A$  es función del vector  $c = (t, u)$  en la forma  $x = t + u$ ,  $y = tu$ ,  $z = t^2 + u$ ,  $w = 2u$ , se verifica que

$\frac{\partial y(A)}{\partial c} = (u^2 - 2ut, 2ut - t^2 - 4u)$ , pues por la regla de la cadena, y las proposiciones 3.4 y 5

$$\begin{aligned} \frac{\partial y(A)}{\partial c} &= \left( \frac{\partial y(A)}{\partial \text{vec}(A)} \otimes I_1 \right)' \left( \frac{\partial \text{vec}(A)}{\partial c} \right) = \left( \frac{\partial y(A)}{\partial \text{vec}(A)} \right)' \left( \frac{\partial \text{vec}(A)}{\partial c} \right) \\ &= \begin{pmatrix} y & -z & -w & x \end{pmatrix} \begin{pmatrix} 1 & 1 \\ 0 & 2 \\ 2t & 1 \\ u & t \end{pmatrix} = \begin{pmatrix} y - 2tw + xu & y - 2x - w + xt \end{pmatrix} \end{aligned}$$

### 3.6. Propiedades de Derivación de Funciones Matriciales.

En esta sección se recogen algunas de las que aparecen con mayor frecuencia. Los resultados que se indican se obtienen utilizando las definiciones propuestas en este trabajo, las reglas de derivación y las diferentes propiedades tanto del producto de Kronecker como de la vectorización de matrices.

**Proposición 3.5** Sea  $A$  una matriz  $m \times n$  y sean  $x$  y  $z$  vectores de  $\mathbb{R}^n$  y  $\mathbb{R}^m$ , respectivamente. Entonces

1. Si  $y(x) = Ax$  se verifica que  $\frac{\partial y(x)}{\partial x} = A$ .

2. Si  $y(x, z) = z^T Ax$  se verifica que

$$\frac{\partial y}{\partial x} = z^T A \quad \frac{\partial y}{\partial z} = x^T A^T.$$

3. Si  $m = n$  e  $y(x) = x^T Ax$ , se verifica que

$$\frac{\partial y(x)}{\partial x} = x^T (A + A^T), \quad \frac{\partial^2 y(x)}{\partial x \partial x} = (A + A^T),$$

y si además  $A$  es una matriz simétrica se tiene que

$$\frac{\partial y(x)}{\partial x} = 2x^T A, \quad \frac{\partial^2 y(x)}{\partial x \partial x} = 2A.$$

4. Si  $m = n$ ,  $A$  es simétrica,  $X$  es una matriz de orden  $n \times p$  e  $Y(X) = X^T AX$ , se verifica que

$$\frac{\partial Y(X)}{\partial X} = P_{n,p}(AX \otimes I_p) + (\text{vec}(AX))(\text{vec} I_p)^T.$$

**Demostración.** Si  $y(x) = Ax$  entonces

$$y(x) = Ax = \begin{pmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{pmatrix}$$

y teniendo en cuenta la definición 3.12

$$\frac{\partial y(x)}{\partial x} = \begin{pmatrix} \frac{\partial}{\partial x} \left( \sum_{j=1}^n a_{1j} x_j \right) \\ \frac{\partial}{\partial x} \left( \sum_{j=1}^n a_{2j} x_j \right) \\ \vdots \\ \frac{\partial}{\partial x} \left( \sum_{j=1}^n a_{mj} x_j \right) \end{pmatrix} = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{pmatrix} = A \quad \blacksquare$$

**Demostración.** Dado que  $y(x, z) = z \lrcorner A x = (z \lrcorner A) x$ , por la proposición 3.5, 1 resulta

$$\frac{\partial y}{\partial x} = \frac{\partial (z \lrcorner A) x}{\partial x} = z \lrcorner A$$

Por otra parte, como  $y = z \lrcorner A x = y \lrcorner = (x \lrcorner A') z$ , razonando de forma análoga se tiene que

$$\frac{\partial y}{\partial x} = x \lrcorner A'. \quad \blacksquare$$

**Demostración.** Si  $y(x) = x \lrcorner A x$  dado que  $y = \sum_{j=1}^n \sum_{i=1}^n x_i x_j a_{ij}$ , entonces, como para cada  $k = 1, \dots, n$ , se tiene que

$$\frac{\partial y}{\partial x_k} = \sum_{j=1}^n a_{kj} x_j + \sum_{i=1}^n a_{ik} x_i = x \lrcorner a'_{k\bullet} + x \lrcorner a_{\bullet k}$$

aplicando la definición 3.10 resulta

$$\frac{\partial y(x)}{\partial x} = \left( \frac{\partial y}{\partial x_1} \cdots \frac{\partial y}{\partial x_n} \right) = x \lrcorner A' + x \lrcorner A = x \lrcorner (A' + A)$$

teniendo en cuenta esta expresión y la proposición 3.5, 1, se deduce que

$$\frac{\partial^2 y(x)}{\partial x \partial x} = \frac{\partial}{\partial x} \left( \frac{\partial y(x)}{\partial x} \right) = \frac{\partial}{\partial x} [x \lrcorner (A' + A)] = (A' + A) \lrcorner = A + A'$$

lo que coincide con

$$\frac{\partial^2 y(x)}{\partial x \partial x'} = \frac{\partial}{\partial x'} \left( \frac{\partial y(x)}{\partial x} \right) = \frac{\partial}{\partial x} \left( \left( \frac{\partial y(x)}{\partial x} \right) \lrcorner \right) = \frac{\partial}{\partial x} [(A + A') x] = A + A'$$

Si  $A$  es una matriz simétrica, los resultados son inmediatos.  $\blacksquare$



**Demostración.** Si  $Y(X) = X'AX = X'(AX)$  aplicando la regla de derivación de un producto 3, se tiene

$$\frac{\partial X'(AX)}{\partial X} = \frac{\partial X'}{\partial X} (AX \otimes I_p) + (X' \otimes I_n) \frac{\partial AX}{\partial X}$$

y aplicándola nuevamente al producto  $AX$  resulta

$$\frac{\partial X'(AX)}{\partial X} = P_{n,p} (AX \otimes I_p) + (X' \otimes I_n) + (A \otimes I_n) \frac{\partial X}{\partial X'}$$

ya que  $\frac{\partial X}{\partial X'} = P_{n,p}$  y  $\frac{\partial A}{\partial X}$  es la matriz nula.

Finalmente a partir de las propiedades 4 del producto de Kronecker y la propiedad 2 de la vectorización, se concluye

$$\frac{\partial X'(AX)}{\partial X} = P_{n,p} (AX \otimes I_p) + (X' \otimes I_n) (vec I_n) (vec I_p)'$$

$$\frac{\partial X'(AX)}{\partial X} = P_{n,p} (AX \otimes I_p) + (vec(AX)) (vec I_p)',$$

ya que  $vec(A'X) = vec(AX)$  por ser  $A$  una matriz simétrica. ■

**Proposición 3.6** Sea  $A$  una matriz regular de orden  $m$ . Entonces se verifica

$$1. \frac{\partial A^{-1}}{\partial A} = -vec(A^{-1}) vec(A^{-1})'.$$

$$2. \frac{\partial vec(A^{-1})}{\partial (vec A)'} = -((A^{-1})' \otimes A^{-1}).$$

Si además se supone que los elementos de la matriz  $A$  son funciones del vector  $x \in \mathbb{R}^n$ , entonces

$$3. \frac{\partial A^{-1}(x)}{\partial x} = \left[ (I_m \otimes (vec I_m)') \left( \frac{\partial A^{-1}}{\partial A} \otimes I_m \right) \right] \otimes \frac{\partial (vec A)}{\partial x}$$

$$4. \frac{\partial vec(A^{-1}(x))}{\partial x} = -((A')^{-1} \otimes A^{-1}) \frac{\partial (vec A)}{\partial x}$$

**Demostración.** Si  $A$  es una matriz regular, entonces existe  $A^{-1}$  verificando

$$AA^{-1} = I_m,$$

y aplicando en esta expresión la regla de derivación de un producto resulta

$$0_{m^2} = \frac{\partial AA^{-1}}{\partial A} = \frac{\partial A}{\partial A} (A^{-1} \otimes I_m) + (A \otimes I_m) \frac{\partial A^{-1}}{\partial A},$$

de donde

$$\frac{\partial A^{-1}}{\partial A} = -(A \otimes I_m)^{-1} \frac{\partial A}{\partial A} (A^{-1} \otimes I_m).$$

Ahora bien como

$$\frac{\partial A}{\partial A} = (\text{vec} I_m) (\text{vec} I_m)'$$

a partir de las propiedades 6, 7 del producto de Kronecker y la propiedad 3 de la vectorización se obtiene

$$\frac{\partial A^{-1}}{\partial A} = -(A^{-1} \otimes I_m) (\text{vec} I_m) (\text{vec} I_m)' (A^{-1} \otimes I_m)$$

$$\frac{\partial A^{-1}}{\partial A} = -(A^{-1} \otimes I_m) (\text{vec} I_m) [((A^{-1})' \otimes I_m) \text{vec} I_m]'$$

$$\frac{\partial A^{-1}}{\partial A} = -\text{vec} (A^{-1})' [\text{vec} (A^{-1})]' \blacksquare$$

**Demostración.** Dado que  $AA^{-1} = I_m$ , aplicando el resultado 3a de la Proposición 3.4 para cada  $i, j = 1, \dots, m$  se tiene que

$$\frac{\partial (AA^{-1})}{\partial a_{ij}} = 0_m = \frac{\partial A}{\partial a_{ij}} A^{-1} + A \frac{\partial A^{-1}}{\partial a_{ij}}$$

y, por tanto,

$$\frac{\partial A^{-1}}{\partial a_{ij}} = -A^{-1} \frac{\partial A}{\partial a_{ij}} A^{-1}.$$

Como

$$\begin{matrix} j \\ \downarrow \end{matrix}$$

$$\frac{\partial A^{-1}}{\partial a_{ij}} = \begin{pmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \leftarrow i$$

si se denota por  $\bar{a}_{i\bullet}$  y  $\bar{a}_{\bullet j}$  la fila  $i$ -ésima y la columna  $j$ -ésima de  $A^{-1}$ , respectivamente, se obtiene

$$\begin{aligned} \frac{\partial A^{-1}}{\partial a_{ij}} &= - \begin{pmatrix} - & \bar{a}_{1\bullet} & - \\ - & \bar{a}_{2\bullet} & - \\ & \vdots & \\ - & \bar{a}_{m\bullet} & - \end{pmatrix} \begin{pmatrix} 0 & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 1 & \cdots & 0 \\ \vdots & \vdots & & \vdots & & \vdots \\ 0 & 0 & \cdots & 0 & \cdots & 0 \end{pmatrix} \begin{pmatrix} | & & | \\ \bar{a}_{\bullet 1} & \cdots & \bar{a}_{\bullet m} \\ | & & | \end{pmatrix} = \\ &= -\bar{a}_{\bullet i} \bar{a}_{j\bullet}. \end{aligned}$$

que de acuerdo con la segunda expresión alternativa y la propiedad 2 de la vectorización de matrices, resulta

$$\frac{\partial \text{vec}(A^{-1})}{\partial a_{ij}} = \text{vec}(-\bar{a}_{\bullet i} \bar{a}_{j\bullet}) = -(I_m \otimes \bar{a}_{\bullet i}) \text{vec} \bar{a}_{j\bullet} = -(\bar{a}_{j\bullet} \otimes \bar{a}_{\bullet i})$$

de donde

$$\begin{aligned} \frac{\partial \text{vec}(A^{-1})}{\partial \mathbf{a}_{j\bullet}} &= \left( \frac{\partial \text{vec}(A^{-1})}{\partial a_{1j}}, \frac{\partial \text{vec}(A^{-1})}{\partial a_{2j}}, \dots, \frac{\partial \text{vec}(A^{-1})}{\partial a_{mj}} \right) \\ &= -[\bar{a}_{j\bullet} \otimes \bar{a}_{\bullet 1}, \bar{a}_{j\bullet} \otimes \bar{a}_{\bullet 2}, \dots, \bar{a}_{j\bullet} \otimes \bar{a}_{\bullet m}] = (\bar{a}_{j\bullet} \otimes A^{-1}), \end{aligned}$$

y por tanto

$$\begin{aligned} \frac{\partial \text{vec}(A^{-1})}{\partial \text{vec}(A)'} &= -[\bar{a}_{1\bullet} \otimes A^{-1}, \bar{a}_{2\bullet} \otimes A^{-1}, \dots, \bar{a}_{m\bullet} \otimes A^{-1}] = \\ &= -[(A^{-1})' \otimes A^{-1}] \quad \blacksquare \end{aligned}$$

**Demostración.** Si los elementos de la matriz  $A$  son función del vector  $x \in \mathbb{R}^n$ , entonces, para cada  $i, j = 1, \dots, m$ , la derivada del elemento  $\bar{a}_{ij}$  de  $A^{-1}$  respecto de  $x$ , en virtud de la regla de la cadena (Proposición 3.4, 5) es

$$\frac{\partial \bar{a}_{ij}}{\partial x} = \left( \frac{\partial \bar{a}_{ij}}{\partial \text{vec}(A)} \right)' \frac{\partial \text{vec}(A)}{\partial x} = \frac{\partial \bar{a}_{ij}}{\partial (\text{vec} A)'} \frac{\partial \text{vec}(A)}{\partial x},$$

y por tanto

$$\frac{\partial A^{-1}}{\partial x} = \left( \frac{\partial A^{-1}}{\partial (vec A)'} \otimes \frac{\partial vec(A)}{\partial x} \right).$$

Ahora bien, como

$$\frac{\partial A^{-1}}{\partial (vec A)'} = \left( \frac{\partial (A^{-1})'}{\partial vec(A)} \right)',$$

y

$$\frac{\partial (A^{-1})'}{\partial A'} = \left( \frac{\partial A^{-1}}{\partial A} \right)',$$

de acuerdo con las formas alternativas

$$\frac{\partial (A^{-1})'}{\partial vec(A)} = \left( \frac{\partial (A^{-1})'}{\partial A'} \otimes I_m \right) (I_m \otimes vec(I_m)) =$$

$$\left[ \left( \frac{\partial A^{-1}}{\partial A} \right)' \otimes I_m \right] (I_m \otimes vec(I_m)),$$

de donde

$$\frac{\partial A^{-1}}{\partial x} = \left\{ \left[ \left( \frac{\partial A^{-1}}{\partial A} \right)' \otimes I_m \right] (I_m \otimes vec(I_m)) \right\}' \otimes \frac{\partial vec(A)}{\partial x},$$

o lo que es igual

$$\frac{\partial A^{-1}}{\partial x} = \left[ (I_m \otimes vec(I_m))' \left( \frac{\partial A^{-1}}{\partial A} \otimes I_m \right) \right]' \otimes \frac{\partial vec(A)}{\partial x}.$$

Pero como de acuerdo con las expresiones alternativas

$$\frac{\partial vec(A)}{\partial x} = \left( I_m \otimes \frac{\partial A}{\partial x} \right) (vec(I_m) \otimes I_n)$$

se obtiene finalmente

$$\frac{\partial A^{-1}(x)}{\partial x} = \left[ (I_m \otimes (vec I_m)') \left( \frac{\partial A^{-1}}{\partial A} \otimes I_m \right) \right]' \otimes \left[ \left( I_m \otimes \frac{\partial A}{\partial x} \right) (vec(I_m) \otimes I_n) \right].$$

■

**Demostración.** Como ya se indicó antes

$$\frac{\partial \bar{a}_{ij}}{\partial x} = \left( \frac{\partial \bar{a}_{ij}}{\partial (vec A)'} \right) \left( \frac{\partial vec(A)}{\partial x} \right),$$

así pues, de acuerdo con la proposición 3.6, 2 resulta

$$\frac{\partial \text{vec}(A^{-1})}{\partial x} = \frac{\partial \text{vec}(A^{-1})}{\partial (\text{vec} A)'} \frac{\partial \text{vec}(A)}{\partial x} = -[(A^{-1})' \otimes A^{-1}] \frac{\partial \text{vec}(A)}{\partial x}. \blacksquare$$

**Proposición 3.7** sean  $A, B$  y  $X$  matrices de ordenes  $m \times m$ ,  $m \times n$ , y  $n \times m$  respectivamente. Entonces

$$1. \text{ Si } y(A) = \text{tr}(A), \text{ se tiene que } \frac{\partial y(A)}{\partial A} = \frac{\partial \text{tr}(A)}{\partial A} = I_m$$

$$2. \text{ Si } y(X) = \text{tr}(BX), \text{ se tiene que}$$

$$\frac{\partial y(X)}{\partial X} = \frac{\partial \text{tr}(BX)}{\partial X} = B'$$

$$3. \text{ Si } y(X) = \text{tr}(BXA), \text{ se tiene que}$$

$$\frac{\partial y(X)}{\partial X} = \frac{\partial \text{tr}(BXA)}{\partial X} = B'A'$$

$$4. \text{ Si } C \text{ es una matriz de orden } n \text{ e } y(X) = \text{tr}(AX'CX), \text{ se tiene que}$$

$$\frac{\partial y(X)}{\partial X} = \frac{\partial \text{tr}(AX'CX)}{\partial X} = C'XA' + CXA$$

$$\text{y si además } C \text{ es una matriz simétrica y } A = I'_m, \text{ entonces } \frac{\partial y(X)}{\partial X} = 2CX$$

**Demostración.** Dado que  $\text{tr}(A) = \sum_{i=1}^m a_{ii}$ , de acuerdo con la Definición 3.13, es trivial obtener

$$\frac{\partial y(A)}{\partial A} = \frac{\partial \text{tr}(A)}{\partial A} = I_m \blacksquare$$

**Demostración.** Si  $y(X) = \text{tr}(BX)$ , entonces  $y(X) = \sum_{i=1}^m \sum_{k=1}^n b_{ik}x_{ki}$  y por tanto,

$$\frac{\partial \text{tr}(BX)}{\partial X} = \frac{\partial}{\partial X} \left( \sum_{i=1}^m \sum_{k=1}^n b_{ik}x_{ki} \right) = \begin{pmatrix} \frac{\partial}{\partial x_1} \left( \sum_{i=1}^m \sum_{k=1}^n b_{ik}x_{ki} \right) \\ \vdots \\ \frac{\partial}{\partial x_n} \left( \sum_{i=1}^m \sum_{k=1}^n b_{ik}x_{ki} \right) \end{pmatrix} =$$

$$\begin{pmatrix} b'_{\cdot 1} \\ \vdots \\ b'_{\cdot n} \end{pmatrix} = B' \blacksquare$$

**Demostración.** Teniendo en cuenta que  $\text{tr}(BXA) = \text{tr}(ABX)$ , aplicando la proposición 3.7, 2, resulta

$$\frac{\partial y(X)}{\partial X} = \frac{\partial \text{tr}(ABX)}{\partial X} = (AB)' = B'A' \blacksquare$$

**Demostración.** De acuerdo con las propiedades 8 y 2 de la vectorización de una matriz y la propiedad 4 del producto de Kronecker se tiene

$$\begin{aligned} \text{tr}(AX'CX) &= [\text{vec}(AX')] \iota \text{vec}(CX) = [\text{vec}(XA')] \iota \text{vec}(CX) \\ &= [(A \otimes I_n) \text{vec}(X)] \iota (I_m \otimes C) \text{vec}(X) \\ &= [\text{vec}(X)] \iota (A \otimes C) \text{vec}(X) \end{aligned}$$

La derivada de esta expresión respecto a  $\text{vec}(X)$  es, en virtud de la Proposición 3.5,3

$$\begin{aligned} \frac{\partial \text{tr}(AX'CX)}{\partial \text{vec}(X)} &= \left( \frac{\partial \text{tr}(AX'CX)}{\partial [\text{vec}(X)]} \right)' \iota = ((\text{vec}(X))' \iota [(A' \otimes C) + (A \otimes C')])' \iota \\ &= [(A \otimes C') \text{vec}(X)] + [(A' \otimes C) \text{vec}(X)] \\ &= \text{vec}(C'XA') + \text{vec}(CXA) = \text{vec}(C'XA' + CXA) \end{aligned}$$

Por ultimo, teniendo en cuenta la primera expresión alternativa, como

$$\text{vec} \left( \frac{\partial \text{tr}(AX'CX)}{\partial X} \right) = \frac{\partial \text{tr}(AX'CX)}{\partial \text{vec}(X)} = \text{vec}(C'XA' + CXA)$$

se concluye que

$$\frac{\partial \text{tr}(AX'CX)}{\partial X} = C'XA' + CXA$$

y, en caso de ser  $C$  simétrica y  $A = I_m$ , es inmediato que

$$\frac{\partial \text{tr}(XCX)}{\partial X} = 2CX \blacksquare$$

**Proposición 3.8** Sea  $A$  una matriz cuadrada de orden  $m$  y  $X$  una matriz de orden  $m \times p$ . Entonces

1. Si  $y(A) = |A|$ , se tiene que  $\frac{\partial y(A)}{\partial A} = \frac{\partial |A|}{\partial A} = \text{Adj}(A)'$

2. Si  $A$  es simétrica e  $y(X) = |X'AX|$ , se tiene que

$$\frac{\partial y(X)}{\partial X} = 2AX \text{Adj}(X'AX)$$

3. Si  $A$  es simétrica e  $y(X) = \ln |X'AX|$  con  $X'AX$  no singular, se tiene que

$$\frac{\partial y(X)}{\partial X} = \frac{2}{|X'AX|} AX \text{Adj}(X'AX) = 2AX (X'AX)^{-1}$$

Si se supone que los elementos de la matriz  $A$  son funciones de  $x \in \mathbb{R}^n$ , entonces

4. Si  $y(x) = |A(x)|$ , se tiene que

$$\frac{\partial y(x)}{\partial x} = [\text{vec}(\text{Adj}(A)')] \left( I_m \otimes \frac{\partial A(x)}{\partial x} \right) (\text{vec}(I_m) \otimes I_n)$$

verificandose tambien que si  $A(x)$  es no singular

$$\frac{\partial \ln |A(x)|}{\partial x} = \left[ \text{tr} \left( (A^{-1})' \frac{\partial A(x)}{\partial x_i} \right) \right]_{i=1, \dots, n}$$

**Demostración.** Dada  $A$  matriz cuadrada de orden  $m$ , si para cada elemento  $a_{ij}$   $i, j = 1, \dots, n$  se denota por  $A_{ij}$  su cofactor correspondiente, se tiene que

$$|A| = \sum_{k=1}^n a_{ik} A_{ik}$$

Entonces

$$\frac{\partial |A|}{\partial a_{ij}} = \frac{\partial}{\partial a_{ij}} \left[ \sum_{k=1}^n a_{ik} A_{ik} \right] = A_{ij}$$

Teniendo en cuenta la Definición 3.12 resulta

$$\frac{\partial |A|}{\partial a_{ij}} = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1n} \\ \vdots & \vdots & & \vdots \\ A_{n1} & A_{n2} & \cdots & A_{nn} \end{pmatrix} = \text{Adj}(A)' \blacksquare$$

**Demostración.** Si se denota por  $b_{ij}$   $i, j = 1, \dots, p$  a los elementos de la matriz  $B = X \wr AX$ , aplicando la regla de la cadena, resulta para cada  $k = 1, \dots, m$ ,  $\ell = 1, \dots, p$

$$\frac{\partial y(X)}{\partial x_{k\ell}} = \sum_{i=1}^p \sum_{j=1}^p \frac{\partial |B|}{\partial b_{ij}} \frac{\partial b_{ij}}{\partial x_{k\ell}}$$

Como para cada  $i, j = 1, \dots, p$   $\frac{\partial |B|}{\partial b_{ij}}$  es el cofactor correspondiente  $B_{ij}$ , se tiene

$$\frac{\partial y(X)}{\partial x_{k\ell}} = \frac{\partial |X \wr AX|}{\partial x_{k\ell}} = \sum_{i=1}^p \sum_{j=1}^p B_{ij} \frac{\partial b_{ij}}{\partial x_{k\ell}} =$$

$$[vec(Adj(X \wr AX))] \wr ec \left( \frac{\partial X \wr AX}{\partial x_{k\ell}} \right)$$

que de acuerdo con la propiedad 8 de la vectorización se puede reducir a

$$\frac{\partial |X \wr AX|}{\partial x_{k\ell}} = tr \left( Adj(X \wr AX) \frac{\partial X \wr AX}{\partial x_{k\ell}} \right)$$

Teniendo en cuenta la derivada de un producto (Proposición 3, 3a), para cada  $k = 1, \dots, m$ ,  $\ell = 1, \dots, p$ , se verifica

$$\frac{\partial X \wr AX}{\partial x_{k\ell}} = \frac{\partial X \wr}{\partial x_{k\ell}} AX + X \wr A \frac{\partial X}{\partial x_{k\ell}} \text{ donde}$$

$$\frac{\partial X \wr}{\partial x_{k\ell}} = e_\ell^m e_k^{p'} \quad \text{y} \quad \frac{\partial X}{\partial x_{k\ell}} = e_k^m e_\ell^{p'}$$

además,  $e_k^m$  y  $e_k^{p'}$  representan dos vectores columna de  $m$  y  $p$  elementos todos ellos nulos salvo el  $k$ -ésimo y el  $\ell$ -ésimo, respectivamente, que son iguales a la unidad. En virtud de esto y teniendo en cuenta las propiedades de la traza,

$$\begin{aligned} \frac{\partial |X \wr AX|}{\partial x_{k\ell}} &= tr [Adj(X \wr AX) (e_\ell^m e_k^{p'} AX + X \wr A e_k^m e_\ell^{p'})] \\ &= (Adj(X \wr AX) e_\ell^m e_k^{p'} AX) + tr (X \wr A e_k^m e_\ell^{p'} Adj(X \wr AX)) \\ &= tr (b_{\ell}^* a_k \cdot X) + tr (X \wr a_{\cdot k} b_{\ell}^*) \\ &= tr (a_k \cdot X b_{\ell}^*) + tr (a'_{\cdot k} X b_{\ell}^{*'}) \\ &= a_k \cdot X b_{\ell}^* + a'_{\cdot k} X b_{\ell}^{*'} \end{aligned}$$



siendo  $b_{\ell}^*$  y  $b_{\ell}^*$  la  $\ell$ -ésima columna y fila, respectivamente, de la matriz  $Adj(B) = Adj(X \wr AX)$  y  $a_{\cdot k}$  y  $a_k$  la  $k$ -ésima columna y fila, respectivamente, de la matriz  $A$

De acuerdo con la Definición 3.13,

$$\begin{aligned} \frac{\partial x(X)}{\partial X} &= \frac{\partial |X \wr AX|}{\partial X} = AX \ Adj(X \wr AX) + A \wr X (Adj(X \wr AX)) \wr = \\ &= 2AX Adj(X \wr AX) \end{aligned}$$

ya que tanto  $A$  como  $X \wr AX$  y, por consiguiente,  $Adj(X \wr AX)$ , son matrices simétricas ■

**Demostración.** Aplicando la regla de la cadena y teniendo en cuenta el apartado anterior, resulta

$$\begin{aligned} \frac{\partial x(X)}{\partial X} &= \frac{\partial In |X \wr AX|}{\partial |X \wr AX|} \frac{\partial |X \wr AX|}{\partial X} \\ &= \frac{1}{|X \wr AX|} 2AX Adj(X \wr AX) \\ &= 2AX (X \wr AX)^{-1} \end{aligned}$$

pues por hipótesis,  $X \wr AX$  es simétrica y no singular ■

**Demostración.** Si  $y(x) = |A(x)|$  con  $x \in \mathbb{R}^n$ , por la regla de la cadena,

$$\frac{\partial |A(x)|}{\partial x} = \left( \frac{\partial |A(x)|}{\partial vec A(x)} \right) \wr \frac{\partial vec A(x)}{\partial x} = \left[ vec \left( \frac{\partial |A(x)|}{\partial A(x)} \right) \right] \wr \frac{\partial vec A(x)}{\partial x}$$

de donde, en virtud de la demostración 3.6 se obtiene

$$\frac{\partial |A(x)|}{\partial x} = [vec(Adj(A(x))) \wr] \wr \left( I_m \otimes \frac{\partial A(x)}{\partial x} \right) (vec I_m \otimes I_n)$$

considerando ahora que  $A(x)$  es una matriz no singular para cada  $x \in \mathbb{R}^n$ , por definición se tiene que

$$\frac{\partial In |A(x)|}{\partial x} = \left( \frac{\partial In |A(x)|}{\partial x_1}, \dots, \frac{\partial In |A(x)|}{\partial x_n} \right)$$

Para cada  $i = 1, \dots, n$

$$\frac{\partial In |A(x)|}{\partial x_i} = \frac{\partial In |A(x)|}{\partial |A(x)|} \frac{\partial |A(x)|}{\partial x_i} = \frac{1}{|A(x)|} \frac{\partial |A(x)|}{\partial x_i}$$

y según la regla de la cadena

$$\frac{\partial |A(x)|}{\partial x_i} = \left( \frac{\partial |A(x)|}{\partial \text{vec}(A(x))} \right)' \frac{\partial \text{vec}(A(x))}{\partial x_i} = [\text{vec}(\text{Adj}(A(x)))]' \frac{\partial \text{vec}(A(x))}{\partial x_i}$$

se tiene

$$\begin{aligned} \frac{\partial \ln |A(x)|}{\partial x_i} &= \left[ \text{vec} \left( \frac{1}{|A(x)|} \text{Adj}(A(x))' \right) \right]' \frac{\partial \text{vec}(A(x))}{\partial x_i} \\ &= (\text{vec}(A^{-1}))' \text{vec} \left( \frac{\partial A(x)}{\partial x_i} \right) \end{aligned}$$

y utilizando la propiedad 8 de la vectorización, se obtiene

$$\frac{\partial \ln |A(x)|}{\partial x_i} = \text{tr} \left( \frac{\partial A(x)}{\partial x_i} (A^{-1})' \right) = \text{tr} \left( (A^{-1})' \frac{\partial A(x)}{\partial x_i} \right)$$

Por tanto

$$\frac{\partial \ln |A(x)|}{\partial x} = \left[ \text{tr} \left( (A^{-1})' \frac{\partial A(x)}{\partial x_1} \right), \dots, \text{tr} \left( (A^{-1})' \frac{\partial A(x)}{\partial x_n} \right) \right] \blacksquare$$

**Ejemplo 3.19** Sea la función  $y(x) = x'Ax$  con  $A$  matriz simétrica y no singular de orden  $n$ . Supóngase que se desea conocer si  $y(x)$  tiene máximo o mínimo global en algún punto  $x \in \mathbb{R}^n$ . para ello habrá que estudiar si  $y(x)$  tiene puntos críticos, lo que presupone resolver el sistema de ecuaciones

$$\frac{\partial y(x)}{\partial x} = 0$$

es decir, de acuerdo con la Proposición 3.5, 3

$$\frac{\partial y(x)}{\partial x} = 2x'A = 0$$

Cuya solución es  $x^* = 0$  por ser  $A$  no singular.

Si  $y(x)$  tiene máximo, mínimo o punto de silla lo alcanzará en  $x^* = 0$ . Para poder decidir el comportamiento de  $y(x)$  en  $x^*$  bastará utilizar condiciones de segundo orden, y como en este caso

$$\frac{\partial^2 y(x)}{\partial x \partial x} = 2A$$

para todo  $x \in \mathbb{R}^n$  se verificará que

$x^* = 0$  es mínimo global si  $A$  es definida o semidefinida positiva

$x^* = 0$  es máximo global si  $A$  es definida o semidefinida negativa

$x^* = 0$  es punto de silla si  $A$  es indefinida

**Ejemplo 3.20** Sea la matriz

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

cuyos elementos dependen de  $x = (x_1, x_2, x_3)$  de forma que

$$a_{11} = a_{22} = x_1 + x_3$$

$$a_{12} = x_1 + x_2$$

$$a_{21} = x_1 - x_2$$

De acuerdo con las proposiciones 3.6 y 3, se verifica que para todo  $x \in \mathbb{R}_{++}^3$

$$\frac{\partial A^{-1}(x)}{\partial x} = \left[ (I_2 \otimes (\text{vec} I_2)) \iota \left( \frac{\partial A^{-1}}{\partial A} \otimes I_2 \right) \right] \otimes \frac{\partial \text{vec}[A(x)]}{\partial x}$$

dado que

$$A^{-1} = \frac{1}{|A|} \begin{pmatrix} a_{22} & -a_{12} \\ -a_{21} & a_{11} \end{pmatrix} \quad \text{resulta}$$

$$\frac{\partial A^{-1}}{\partial A} = -\text{vec}((A^{-1}) \iota) (\text{vec} A^{-1}) \iota =$$

$$- \left( \frac{1}{|A|} \right)^2 \begin{pmatrix} a_{22} \\ -a_{12} \\ -a_{21} \\ a_{11} \end{pmatrix} \begin{pmatrix} a_{22} & -a_{21} & a_{12} & a_{11} \end{pmatrix}$$

$$= -\frac{1}{|A|} \begin{pmatrix} a_{22}^2 & -a_{21}a_{22} & -a_{12}a_{22} & a_{22}a_{11} \\ -a_{12}a_{22} & a_{12}a_{21} & a_{12}^2 & -a_{12}a_{11} \\ -a_{21}a_{22} & a_{21}^2 & a_{12}a_{21} & -a_{21}a_{11} \\ a_{11}a_{22} & -a_{11}a_{21} & -a_{12}a_{11} & a_{11}^2 \end{pmatrix}$$

Por otra parte se obtiene que

$$\frac{\partial \text{vec}[A(x)]}{\partial x} \begin{pmatrix} \frac{\partial a_{11}(x)}{\partial x} \\ \frac{\partial a_{21}(x)}{\partial x} \\ \frac{\partial a_{12}(x)}{\partial x} \\ \frac{\partial a_{22}(x)}{\partial x} \end{pmatrix} = \begin{pmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}$$

Por tanto, efectuando los correspondientes productos y teniendo en cuenta la propiedad 2 del producto de Kronecker, se tiene que

$$\frac{\partial A^{-1}(x)}{\partial x} = -\frac{\text{tr}(A)}{|A|^2}$$

$$\left[ \begin{pmatrix} a_{22} & 0 & -a_{21} & 0 & -a_{12} & 0 & a_{11} & 0 \\ 0 & a_{22} & 0 & -a_{21} & 0 & -a_{12} & 0 & a_{11} \end{pmatrix} \otimes \begin{pmatrix} 1 & 0 & 1 \\ 1 & -1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix} \right]$$

**Ejemplo 3.21** Sea la función  $y(x) = \text{tr}(AX)$  donde

$$A = \begin{pmatrix} 1 & 3 \\ 2 & 0 \\ -1 & 1 \end{pmatrix} \quad y \quad X = \begin{pmatrix} x_{11} & x_{12} & x_{13} \\ x_{21} & x_{22} & x_{23} \end{pmatrix}$$

como  $\text{tr}(AX) = x_{11} + 3x_{21} + 2x_{12} - x_{13} + x_{23}$  entonces

$$\frac{\partial \text{tr}(AX)}{\partial X} = \begin{pmatrix} 1 & 2 & -1 \\ 3 & 0 & 1 \end{pmatrix} = A'$$

tal como se indica en la Proposición 3.7, 2

**Ejemplo 3.22** Sea  $A$  una matriz no singular de orden  $n$ , entonces aplicando la regla de la cadena

$$\frac{\partial \ln |A|}{\partial A} = \frac{1}{|A|} \frac{\partial |A|}{\partial A}$$

y, de acuerdo con la Proposición 3.8, 1, se tiene

$$\frac{\partial \ln |A|}{\partial A} = \frac{1}{|A|} \text{Adj}(A)' = A^{-1}$$

es decir, en virtud de la Definición 3.13, para cada  $i, j = 1, \dots, n$

$$\frac{\partial \ln |A|}{\partial a_{ij}} = \frac{1}{|A|} A_{ji}$$

**Ejemplo 3.23** Dada  $A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$

cuyos elementos dependen de  $x = (x_1, x_2)$  de forma que  $a_{ij} = ix_1 + jx_2$  con  $x_1 \succ 0$  y  $x_2 \prec 0$ ,  $i, j = 1, 2$ , se verifica que

$$\begin{aligned} \frac{\partial \ln |A(x)|}{\partial x} &= \frac{\partial \ln (a_{11}(x)a_{22}(x) - a_{21}(x)a_{12}(x))}{\partial x} \\ &= \frac{1}{|A|} \left[ \frac{\partial (a_{11}(x)a_{22}(x) - a_{21}(x)a_{12}(x))}{\partial x_1}, \frac{\partial (a_{11}(x)a_{22}(x) - a_{21}(x)a_{12}(x))}{\partial x_2} \right] \end{aligned}$$

$$= \left( \frac{1}{x_1}, \frac{1}{x_2} \right)$$

que coincide con lo que se señala en la Proposición 3.8, 4

$$\frac{\partial \ln |A(x)|}{\partial x} = \left[ \text{tr} \left( (A^{-1})' \frac{\partial A(x)}{\partial x_1} \right), \text{tr} \left( (A^{-1})' \frac{\partial A(x)}{\partial x_2} \right) \right] \text{ pues como}$$

$$A(x) = \begin{pmatrix} x_1 + x_2 & x_1 + 2x_2 \\ 2x_1 + x_2 & 2x_1 + 2x_2 \end{pmatrix}$$

entonces

$$A^{-1} = \frac{1}{x_1 + x_2} \begin{pmatrix} -2(x_1 + x_2) & x_1 + 2x_2 \\ 2x_1 + x_2 & -(x_1 + x_2) \end{pmatrix}$$

y

$$\frac{\partial A(x)}{\partial x_1} = \begin{pmatrix} 1 & 1 \\ 2 & 2 \end{pmatrix}, \quad \frac{\partial A(x)}{\partial x_2} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$$

y, por tanto

$$\text{tr} \left( (A^{-1})' \frac{\partial A(x)}{\partial x_1} \right) = \frac{1}{x_1 x_2} (2x_2 - x_2) = \frac{1}{x_1}$$

$$\text{tr} \left( (A^{-1})' \frac{\partial A(x)}{\partial x_2} \right) = \frac{1}{x_1 x_2} (2x_1 - x_1) = \frac{1}{x_2}$$

Cuando la matriz respecto de la que se deriva es simétrica, debe de tenerse en cuenta esta característica pues, en caso contrario, pueden obtenerse resultados aparentemente contradictorios [Bar98]. Por ejemplo si

$$y(A) = a_{11}^2 + a_{22}^2 + a_{12}^2, \text{ donde}$$

$$A = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}$$

y cada elemento de la matriz  $A$  es función del vector  $x = (x_1, x_2)$  en la forma:

$$a_{11} = x_1 + 2x_2, \quad a_{22} = x_1, \quad a_{12} = x_1 x_2$$

al aplicar la regla de la cadena (Proposiciones 3.4, 4) resulta

$$\frac{\partial y(A)}{\partial x} = \left( \frac{\partial y(A)}{\partial \text{vec}(A)} \right)' \frac{\partial \text{vec}(A(x))}{\partial x}$$

$$= \begin{pmatrix} 2a_{11} & 2a_{12} & 2a_{12} & 2a_{22} \end{pmatrix} \begin{pmatrix} 1 & 2 \\ x_2 & x_1 \\ x_2 & x_1 \\ 1 & 0 \end{pmatrix}$$

$$= (4x_1 + 4x_2 + 4x_1x_2^2 \quad 4x_1 + 8x_2 + 4x_1^2x_2)$$

sustituyendo directamente y considerando que

$$y(A(x)) = (x_1 + 2x_2)^2 + x_1^2 + (x_1x_2)^2 = 2x_1^2 + 4x_2^2 + 4x_1x_2 + x_1^2x_2^2$$

al calcular el gradiente se obtiene un resultado diferente, debido a que

$$\nabla y = (4x_1 + 4x_2 + 2x_1x_2^2, 8x_2 + 4x_1 + 2x_1^2x_2)$$

Hay una aparente contradicción, que se debe a que tras sustituir se ha considerado  $y(A)$  como una función de tres variables  $a_{11}, a_{12}, a_{22}$  cuando en realidad es una función de  $\mathbb{R}^4 \rightarrow \mathbb{R}$ , la cual está definida en puntos cuyas componentes segunda y tercera son iguales. Al tener esto en cuenta, los cálculos de la derivada coinciden. Para manejar esta aparente disparidad en los resultados al calcular la derivada respecto de una matriz simétrica  $A$ , usualmente se deriva únicamente respecto de los elementos distintos de  $A$ , utilizando la  $\widetilde{vec}(A)$  cuya descripción se encuentra en la definición 3.9.

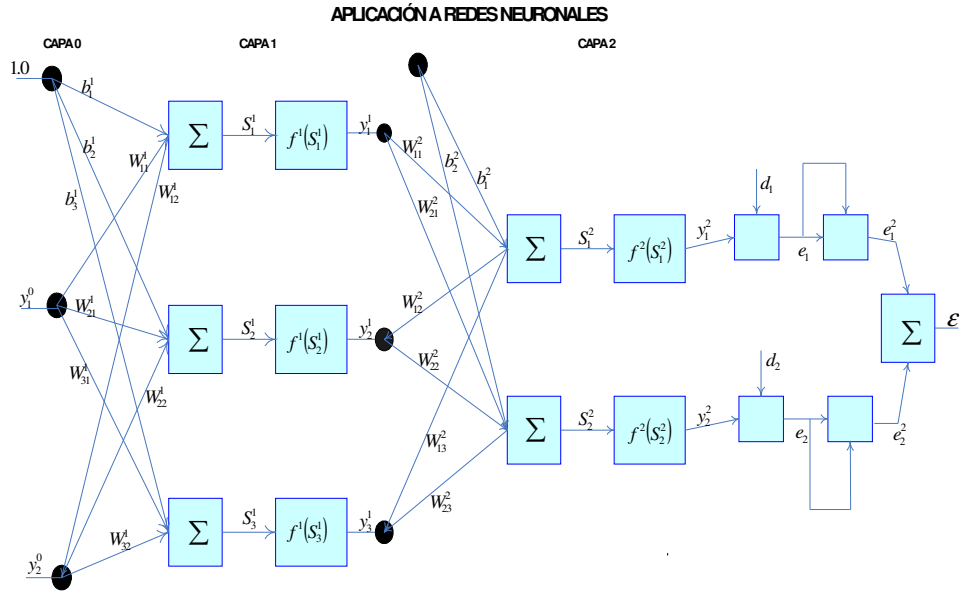


Figura 2:

## 4. Aplicaciones

### 4.1. Redes Neuronales

A continuación se expone el desarrollo completo de una aplicación típica de la derivación de funciones vectoriales y matriciales, que es el caso de la solución de problemas mediante las redes neuronales artificiales. Esta aplicación es un trabajo realizado por [Per07]

Utilizando notación matricial, es fácil comprobar que una red multicapa cuyas funciones de activación son lineales se pueden reducir a una red de una capa lineal; por ello es preciso utilizar funciones de activación no lineales. El algoritmo Backpropagation es un algoritmo iterativo que permite entrenar redes multicapa. La figura ?? muestra una red neuronal típica, donde  $y$  son las entradas,  $S$  las salidas,  $W$  los pesos,  $e$  el error y  $f(s)$  la salida deseada

#### 4.1.1. Características del Algoritmo

1. Busca el mínimo de la función error a partir de un conjunto de patrones de entrenamiento.
2. Precisa que la función de activación sea diferenciable (fácilmente).
3. Entrenar consiste en modificar los pesos de la red.

4. Los pesos se modifican hacia la direcció descendente de la función error.
5. La red entrenada es capaz de generalizar, clasificando correctamente patrones ruidosos o incompletos.
6.  $\alpha$  indica cuánto cambian los pesos.
7. Si  $\alpha$  es muy pequeño, la velocidad de aprendizaje es muy lenta. Si  $\alpha$  es muy grande aparecen efectos oscilatorios.

$$b^1 = \begin{bmatrix} b_1^1 \\ b_2^1 \\ b_3^1 \end{bmatrix}_{nn(1)x1}, b^2 = \begin{bmatrix} b_1^2 \\ b_2^2 \end{bmatrix}_{nn(2)x1}, y^1 = \begin{bmatrix} y_1^1 \\ y_2^1 \\ y_3^1 \end{bmatrix}_{nn(1)x1}, e = \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

$$S^1 = \begin{bmatrix} S_1^1 \\ S_2^1 \\ S_3^1 \end{bmatrix}_{nn(1)x1}, S^2 = \begin{bmatrix} S_1^2 \\ S_2^2 \end{bmatrix}_{nn(2)x1}, y^2 = \begin{bmatrix} y_1^2 \\ y_2^2 \end{bmatrix}_{nn(2)x1}$$

$$W^1 = \begin{bmatrix} W_{11}^1 & W_{12}^1 \\ W_{21}^1 & W_{22}^1 \\ W_{31}^1 & W_{32}^1 \end{bmatrix}_{nn(1) \times nn(0)}, W^2 = \begin{bmatrix} W_{11}^2 & W_{12}^2 \\ W_{21}^2 & W_{22}^2 \\ W_{31}^2 & W_{32}^2 \end{bmatrix}_{nn(2) \times nn(1)}$$

$nn(k)$  = número de neuronas de la capa  $k$ ,  $nc$  = índice de la última capa

$$\text{En general } [S^k]_{nn(k)x1}, [y^k]_{nn(k)x1}, [b^k]_{nn(k)x1}, [W^k]_{nn(k) \times nn(k-1)}, [e]_{nn(nc)x1}$$

#### 4.1.2. Propagación hacia adelante.

Cálculo

$$S_1^1 = W_{11}^1 y_1^0 + W_{12}^1 y_2^0 + b_1^1$$

$$S_2^1 = W_{21}^1 y_1^0 + W_{22}^1 y_2^0 + b_2^1$$

$$S_3^1 = W_{31}^1 y_1^0 + W_{32}^1 y_2^0 + b_3^1$$

$$\text{Matricialmente } S^1 = \begin{bmatrix} W_{11}^1 & W_{12}^1 \\ W_{21}^1 & W_{22}^1 \\ W_{31}^1 & W_{32}^1 \end{bmatrix} \begin{bmatrix} y_1^0 \\ y_2^0 \end{bmatrix} + \begin{bmatrix} b_1^1 \\ b_2^1 \\ b_3^1 \end{bmatrix}$$

Por lo tanto,  $S^1 = W^1 y^0 + b^1$ , y en general,  $S^k = W^k y^{k-1} + b^k$   
 $y_1^1 = f_1^1(S_1^1), y_2^1 = f_2^1(S_2^1), y_3^1 = f_3^1(S_3^1)$



$$\begin{bmatrix} y_1^1 \\ y_2^1 \\ y_3^1 \end{bmatrix} = \begin{bmatrix} f_1^1(S_1^1) \\ f_2^1(S_2^1) \\ f_3^1(S_3^1) \end{bmatrix}, \text{ haciendo } F^1(S^1) = \begin{bmatrix} f_1^1(S_1^1) \\ f_2^1(S_2^1) \\ f_3^1(S_3^1) \end{bmatrix}, \text{ queda } y^1 = F^1(S^1)$$

En general:

$$y^k = F^k(S^k)$$

$$e_1 = d_1 - y_1^2$$

$$e_2 = d_2 - y_2^2$$

$$\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} d_1 \\ d_2 \end{bmatrix} - \begin{bmatrix} y_1^2 \\ y_2^2 \end{bmatrix} \longrightarrow e = d - y^2, \text{ en general } e = d - y^{nc}$$

donde  $nc$  = índice de la última capa

$$\text{El error cuadrático } \epsilon = e_1^2 + e_2^2 = \begin{bmatrix} e_1 & e_2 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = e^T e$$

#### 4.1.3. Propagación hacia atrás.

Escalarmente

$$W_{ij}^2(n+1) = W_{ij}^2(n) - \alpha \frac{\partial \epsilon(n)}{\partial W_{ij}^2(n)}$$

$$b_i^2(n+1) = b_i^2(n) - \alpha \frac{\partial \epsilon(n)}{\partial b_i^2(n)}$$

donde  $n$  indica el número de iteración

$$\begin{bmatrix} W_{11}^2(n+1) & W_{12}^2(n+1) & W_{13}^2(n+1) \\ W_{21}^2(n+1) & W_{22}^2(n+1) & W_{23}^2(n+1) \end{bmatrix} =$$

$$\begin{bmatrix} W_{11}^2(n) & W_{12}^2(n) & W_{13}^2(n) \\ W_{21}^2(n) & W_{22}^2(n) & W_{23}^2(n) \end{bmatrix} - \alpha \begin{bmatrix} \frac{\partial \epsilon(n)}{\partial W_{11}^2(n)} & \frac{\partial \epsilon(n)}{\partial W_{12}^2(n)} & \frac{\partial \epsilon(n)}{\partial W_{13}^2(n)} \\ \frac{\partial \epsilon(n)}{\partial W_{21}^2(n)} & \frac{\partial \epsilon(n)}{\partial W_{22}^2(n)} & \frac{\partial \epsilon(n)}{\partial W_{23}^2(n)} \end{bmatrix}$$

$$\text{Es decir, } W^2(n+1) = W^2(n) - \alpha \frac{\partial \epsilon(n)}{\partial W^2(n)}$$

y escalarmente:

$$\frac{\partial \epsilon}{\partial W_{ij}^2} = \frac{\partial \epsilon}{\partial e_i} \frac{\partial e_i}{\partial y_i^2} \frac{\partial y_i^2}{\partial S_i^2} \frac{\partial S_i^2}{\partial W_{ij}^2} = \frac{\partial \epsilon}{\partial S_i^2} \frac{\partial S_i^2}{\partial W_{ij}^2}$$

Matricialmente

$$\frac{\partial \epsilon}{\partial W^2} = \frac{\partial \epsilon}{\partial S^2} \frac{\partial S^2}{\partial W^2}$$

$\frac{\partial \epsilon}{\partial S^2}$  = derivada de una función escalar con respecto a un vector

$\frac{\partial S^2}{\partial W^2}$  = derivada de una función vectorial de argumento matricial con respecto a una matriz

Escalarmente

$$\frac{\partial \epsilon}{\partial S_i^2} = \frac{\partial \epsilon}{\partial e_i} \frac{\partial e_i}{\partial y_i^2} \frac{\partial y_i^2}{\partial S_i^2}$$

Matricialmente se debe tener en cuenta que si  $z$  es una función de  $p$ , la cual es una función de  $y$ , que a su vez es una función del vector  $x$ , entonces

$$\frac{dz}{dx} = \frac{dy}{dx} \frac{\partial p}{\partial y} \frac{\partial z}{\partial p}$$

Teniendo en cuenta lo anterior

$$\frac{\partial \epsilon}{\partial S^2} = \frac{\partial y^2}{\partial S^2} \frac{\partial e}{\partial y^2} \frac{\partial \epsilon}{\partial e} \longrightarrow \frac{\partial \epsilon}{\partial W^2} = \frac{\partial y^2}{\partial S^2} \frac{\partial e}{\partial y^2} \frac{\partial \epsilon}{\partial e} * \frac{\partial S^2}{\partial W^2}$$

Teniendo en cuenta que

$$\frac{dAx}{dA} = x^T \longrightarrow \frac{\partial S^2}{\partial W^2} = \frac{\partial}{\partial W^2} (W^2 y^1 + b^2) = (y^1)^T$$

En general,

$$\frac{\partial S^k}{\partial W^k} = (y^{k-1})^T$$

$$\frac{\partial y^2}{\partial S^2} = \begin{bmatrix} \frac{\partial y_1^2}{\partial S_1^2} & \frac{\partial y_2^2}{\partial S_1^2} \\ \frac{\partial y_1^2}{\partial S_2^2} & \frac{\partial y_2^2}{\partial S_2^2} \end{bmatrix}, \text{ ya que } y_i^2 = f_i^2(S_i^2), \text{ entonces, } \frac{\partial y_i^2}{\partial S_j^2} = 0, \text{ para}$$

$$\forall i \neq j \text{ y } \frac{\partial y_i^2}{\partial S_j^2} = f_i^{2'}(S_i^2)$$

$$\frac{\partial y^2}{\partial S^2} = \begin{bmatrix} \frac{\partial y_1^2}{\partial S_1^2} & 0 \\ 0 & \frac{\partial y_2^2}{\partial S_2^2} \end{bmatrix}, \text{ como } y_i^2 = f_i^2(S_i^2) \longrightarrow \frac{\partial y_i^2}{\partial S_i^2} = f_i^{2'}(S_i^2)$$

$$y^2 = F^2(S^2) \longrightarrow \text{se puede hacer } F'^2(S^2) = \frac{\partial y^2}{\partial S^2}$$

$$\frac{\partial e}{\partial y^2} = \begin{bmatrix} \frac{\partial e_1}{\partial y_1^2} & \frac{\partial e_2}{\partial y_1^2} \\ \frac{\partial e_1}{\partial y_2^2} & \frac{\partial e_2}{\partial y_2^2} \end{bmatrix}, \text{ ya que } e_i = d_i - y_i^2, \text{ entonces } \frac{\partial e_i}{\partial y_j^2} = 0, \text{ para } \forall i \neq j$$

y  $\frac{\partial e_i}{\partial y_i^2} = -1$

$$\frac{\partial e}{\partial y^2} = \begin{bmatrix} \frac{\partial e_1}{\partial y_1^2} & 0 \\ 0 & \frac{\partial e_2}{\partial y_2^2} \end{bmatrix} = \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix}$$

$$\frac{\partial \varepsilon}{\partial e} = \frac{\partial}{\partial e} (e^T e) = 2e = 2 \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Acoplando todos estos resultados

$$\frac{\partial \varepsilon}{\partial W^2} = \begin{bmatrix} \frac{\partial y_1^2}{\partial S_1^2} & 0 \\ 0 & \frac{\partial y_2^2}{\partial S_2^2} \end{bmatrix} \begin{bmatrix} \frac{\partial e_1}{\partial y_1^2} & 0 \\ 0 & \frac{\partial e_2}{\partial y_2^2} \end{bmatrix} * 2 \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} * \begin{bmatrix} y_1^1 & y_2^1 & y_3^1 \end{bmatrix}$$

$$\frac{\partial \varepsilon}{\partial W^2} = 2 \begin{bmatrix} \frac{\partial y_1^2}{\partial S_1^2} & 0 \\ 0 & \frac{\partial y_2^2}{\partial S_2^2} \end{bmatrix} \begin{bmatrix} -1 & 0 \\ 0 & -1 \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \begin{bmatrix} y_1^1 & y_2^1 & y_3^1 \end{bmatrix}$$

$$\frac{\partial \varepsilon}{\partial W^2} = 2 \begin{bmatrix} -\frac{\partial y_1^2}{\partial S_1^2} & 0 \\ 0 & -\frac{\partial y_2^2}{\partial S_2^2} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \begin{bmatrix} y_1^1 & y_2^1 & y_3^1 \end{bmatrix}$$

$$\frac{\partial \varepsilon}{\partial W^2} = 2 \begin{bmatrix} -\frac{\partial y_1^2}{\partial S_1^2} & e_1 \\ -\frac{\partial y_2^2}{\partial S_2^2} & e_2 \end{bmatrix} \begin{bmatrix} y_1^1 & y_2^1 & y_3^1 \end{bmatrix} =$$

$$\begin{bmatrix} -2e_1 \frac{\partial y_1^2}{\partial S_1^2} y_1^1 & -2e_1 \frac{\partial y_1^2}{\partial S_1^2} y_2^1 & -2e_1 \frac{\partial y_1^2}{\partial S_1^2} y_3^1 \\ -2e_2 \frac{\partial y_2^2}{\partial S_2^2} y_1^1 & -2e_2 \frac{\partial y_2^2}{\partial S_2^2} y_2^1 & -2e_2 \frac{\partial y_2^2}{\partial S_2^2} y_3^1 \end{bmatrix}$$

Si se hace

$$\delta^2 = -\frac{\partial \varepsilon}{\partial S^2} = 2 \begin{bmatrix} -\frac{\partial y_1^2}{\partial S_1^2} & 0 \\ 0 & -\frac{\partial y_2^2}{\partial S_2^2} \end{bmatrix} \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} = \begin{bmatrix} 2 \frac{\partial y_1^2}{\partial S_1^2} e_1 \\ 2 \frac{\partial y_2^2}{\partial S_2^2} e_2 \end{bmatrix} = \begin{bmatrix} 2f_1'^2(S_1^2) e_1 \\ 2f_2'^2(S_2^2) e_2 \end{bmatrix} = \begin{bmatrix} \delta_1^2 \\ \delta_2^2 \end{bmatrix}$$

Matricialmente

$$\delta^2 = 2F^2{}' (S^2) e \longrightarrow \text{ donde } F^2{}' (S^2) = \begin{bmatrix} f_1^2{}' (S_1^2) & 0 \\ 0 & f_2^2{}' (S_2^2) \end{bmatrix}$$

Entonces,

$$\frac{\partial \varepsilon}{\partial W^2} = -\delta^2 (y^1)^T$$

recordando que

$$W^2 (n+1) = W^2 (n) - \alpha \frac{\partial \varepsilon (n)}{\partial W^2 (n)} \longrightarrow \text{ donde } n \text{ es el contador de iteraciones}$$

$$W^2 (n+1) = W^2 (n) + \alpha \delta^2 (n) \left( y_{(n)}^1 \right)^T$$

Ahora

$$b^2 (n+1) = b^2 (n) - \alpha \frac{\partial \varepsilon (n)}{\partial b^2 (n)}$$

$$\frac{\partial \varepsilon}{\partial b^2} = \frac{\partial S^2}{\partial b^2} \frac{\partial \varepsilon}{\partial S^2} = -\frac{\partial S^2}{\partial b^2} \delta^2$$

pero  $S^2 = W^2 y^1 + b^2$  recuérdese que  $y^{k+1} = W^k S^K + b^k$

$$\frac{\partial S^2}{\partial b^2} = \begin{bmatrix} \frac{\partial S_1^2}{\partial b_1^2} & \frac{\partial S_2^2}{\partial b_2^2} \\ \frac{\partial S_1^2}{\partial b_2^2} & \frac{\partial S_2^2}{\partial b_1^2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$\frac{\partial \varepsilon}{\partial b^2} = -\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \delta_1^2 \\ \delta_2^2 \end{bmatrix} = -\begin{bmatrix} \delta_1^2 \\ \delta_2^2 \end{bmatrix} = -\delta^2$$

Por lo tanto

$$b^2 (n+1) = b^2 (n) + \alpha \delta^2$$

En general, para los pesos y el *BIAS* de la última capa se tiene  $\delta^{nc} = 2F^{nc}{}' (S^{nc}) e$

$$\frac{\partial \varepsilon}{\partial W^{nc}} = -\delta^{nc} (y^{nc-1})^T$$

$$W^{nc} (n+1) = W^{nc} (n) + \alpha \delta^2 (n) \left( y_{(n)}^{nc-1} \right)^T$$

$\frac{\partial \varepsilon}{\partial b^{nc}} = -\delta^{nc}$  y  $b^{nc}(n+1) = b^{nc}(n) + \delta_{(n)}^{nc}$  donde  $nc$  es el número de capas de la red

Ahora es necesario hallar la fórmula para la actualización de  $W^1$  y  $b^1$

$$W^1(n+1) = W^1(n) - \alpha \frac{\partial \varepsilon(n)}{\partial W^1(n)}$$

$$\frac{\partial \varepsilon}{\partial W^1} = \frac{\partial \varepsilon}{\partial S^1} \frac{\partial S^1}{\partial W^1} \text{ y } \frac{\partial \varepsilon}{\partial S^1} = \frac{\partial y^1}{\partial S^1} \frac{\partial S^2}{\partial y^1} \frac{\partial y^2}{\partial S^2} \frac{\partial \varepsilon}{\partial y^2} \frac{\partial \varepsilon}{\partial e} = \frac{\partial y^1}{\partial S^1} \frac{\partial S^2}{\partial y^1} \frac{\partial \varepsilon}{\partial S^2} = \frac{\partial S^2}{\partial S^1} \frac{\partial \varepsilon}{\partial S^2}$$

$$\frac{\partial y^1}{\partial S^1} = \begin{bmatrix} \frac{\partial y_1^1}{\partial S_1^1} & \frac{\partial y_2^1}{\partial S_1^1} & \frac{\partial y_3^1}{\partial S_1^1} \\ \frac{\partial y_1^1}{\partial S_2^1} & \frac{\partial y_2^1}{\partial S_2^1} & \frac{\partial y_3^1}{\partial S_2^1} \\ \frac{\partial y_1^1}{\partial S_3^1} & \frac{\partial y_2^1}{\partial S_3^1} & \frac{\partial y_3^1}{\partial S_3^1} \end{bmatrix}$$

Recordando que  $y_i^1 = f_i^1(S_i^1) \longrightarrow \frac{\partial y_i^1}{\partial S_j^1} = 0, \forall i \neq j \longrightarrow \frac{\partial y_i^1}{\partial S_i^1} = f_i^{1'}(S_i^1)$

$$\frac{\partial y^1}{\partial S^1} = \begin{bmatrix} \frac{\partial y_1^1}{\partial S_1^1} & 0 & 0 \\ 0 & \frac{\partial y_2^1}{\partial S_2^1} & 0 \\ 0 & 0 & \frac{\partial y_3^1}{\partial S_3^1} \end{bmatrix} = \begin{bmatrix} f_1^{1'}(S_1^1) & 0 & 0 \\ 0 & f_2^{1'}(S_2^1) & 0 \\ 0 & 0 & f_3^{1'}(S_3^1) \end{bmatrix} = F^{1'}(S^1)$$

$$\frac{\partial S^2}{\partial y^1} = \frac{\partial}{\partial y^1} (W^2 y^1 + b^2) = (W^2)^T$$

Es decir

$$\frac{\partial S^2}{\partial y^1} = \begin{bmatrix} \frac{\partial S_1^2}{\partial y_1^1} & \frac{\partial S_2^2}{\partial y_1^1} \\ \frac{\partial S_1^2}{\partial y_2^1} & \frac{\partial S_2^2}{\partial y_2^1} \\ \frac{\partial S_1^2}{\partial y_3^1} & \frac{\partial S_2^2}{\partial y_3^1} \end{bmatrix} \text{ recordando que}$$

$$S_1^2 = W_{i1}^2 y_1^1 + W_{i2}^2 y_2^1 + W_{i3}^2 y_3^1 + b_i^2 \longrightarrow \frac{\partial S_i^2}{\partial y_j^1} = W_{ij}^2$$

$$\frac{\partial S^2}{\partial y^1} = \begin{bmatrix} W_{11}^2 & W_{21}^2 \\ W_{12}^2 & W_{22}^2 \\ W_{13}^2 & W_{23}^2 \end{bmatrix} = (W^2)^T$$

Recordando que  $\frac{\partial \varepsilon}{\partial S^2} = -\delta^2$

$$\frac{\partial \epsilon}{\partial S^1} = - \begin{bmatrix} \frac{\partial y_1^1}{\partial S_1^1} & 0 & 0 \\ 0 & \frac{\partial y_2^1}{\partial S_2^1} & 0 \\ 0 & 0 & \frac{\partial y_3^1}{\partial S_3^1} \end{bmatrix} \begin{bmatrix} W_{11}^2 & W_{21}^2 \\ W_{12}^2 & W_{22}^2 \\ W_{13}^2 & W_{23}^2 \end{bmatrix} \begin{bmatrix} \delta_1^2 \\ \delta_2^2 \end{bmatrix}$$

$$\frac{\partial \epsilon}{\partial S^1} = -F^1 \iota (S^1) (W^2)^T \delta^2$$

$$\text{haciendo } \delta^1 = F^1 \iota (S^1) (W^2)^T \delta^2$$

$$\frac{\partial \epsilon}{\partial S^1} = -\delta^1$$

$$\frac{\partial S^1}{\partial W^1} = \frac{\partial (W^1 y^0 + b^1)}{\partial W^1} = (y^0)^T$$

$$\text{en general } \frac{\partial S^k}{\partial W^k} = (y^{k-1})^T$$

$$\frac{\partial \epsilon}{\partial W^1} = \frac{\partial \epsilon}{\partial S^1} \frac{\partial S^1}{\partial W^1} = -\delta^1 (y^0)^T$$

En general

$$\frac{\partial \epsilon}{\partial W^k} = \frac{\partial \epsilon}{\partial S^k} \frac{\partial S^k}{\partial W^k}$$

$$\frac{\partial \epsilon}{\partial S^k} = \frac{\partial S^{k+1}}{\partial S^k} \frac{\partial \epsilon}{\partial S^{k+1}}$$

$$\frac{\partial S^{k+1}}{\partial S^k} = \frac{\partial y^k}{\partial S^k} \frac{\partial S^{k+1}}{\partial y^k} \longrightarrow \frac{\partial y^k}{\partial S^k} = F^k \iota (S^k)$$

$$\frac{\partial S^{k+1}}{\partial y^k} = \frac{\partial (W^{k+1} y^k + b^{k+1})}{\partial y^k} \longrightarrow \frac{\partial S^{k+1}}{\partial y^k} = (W^{k+1})^T$$

$$\frac{\partial S^{k+1}}{\partial y^k} = F^k \iota (S^k) (W^{k+1})^T$$

$$\frac{\partial \epsilon}{\partial S^{k+1}} = -\delta^{k+1} \longrightarrow \frac{\partial \epsilon}{\partial S^k} = -F^k \iota (S^k) (W^{k+1})^T \delta^{k+1}$$

Si se generaliza

$$\delta^k = -\frac{\partial \epsilon}{\partial S^k} \longrightarrow \delta^k = F^k \iota (S^k) (W^{k+1})^T \delta^{k+1} \longrightarrow \delta^k = (W^{k+1} F^k \iota (S^k))^T \delta^{k+1} \text{ ya que } F^k \iota (S^k) \text{ es simétrica}$$

$$\frac{\partial \varepsilon}{\partial W^k} = \frac{\partial \varepsilon}{\partial S^k} \frac{\partial S^k}{\partial W^k} = \delta^k (y^{k-1})^T$$

$$W^1(n+1) = W^1(n) + \alpha \delta_{(n)}^1 (y_{(n)}^0)^T$$

y, en general

$$W^k(n+1) = W^k(n) + \alpha \delta_{(n)}^k (y_{(n)}^{k-1})^T$$

Ahora se determina la fórmula de actualización para  $b^1$

$$b^1(n+1) = b^1(n) - \alpha \frac{\partial \varepsilon(n)}{\partial b^1(n)}$$

$$\frac{\partial \varepsilon}{\partial b^1} = \frac{\partial S^1}{\partial b^1} \frac{\partial \varepsilon}{\partial S^1} \quad \frac{\partial \varepsilon}{\partial S^1} = -\delta^1$$

$$\frac{\partial S^1}{\partial b^1} = \begin{bmatrix} \frac{\partial S_1^1}{\partial b_1^1} & \frac{\partial S_2^1}{\partial b_1^1} & \frac{\partial S_3^1}{\partial b_1^1} \\ \frac{\partial S_1^1}{\partial b_2^1} & \frac{\partial S_2^1}{\partial b_2^1} & \frac{\partial S_3^1}{\partial b_2^1} \\ \frac{\partial S_1^1}{\partial b_3^1} & \frac{\partial S_2^1}{\partial b_3^1} & \frac{\partial S_3^1}{\partial b_3^1} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} = [I]_{nn(1)}$$

En general,

$$\frac{\partial S^k}{\partial b^k} = [I]_{nn(k)}$$

$$\frac{\partial \varepsilon}{\partial b^1} = - \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \delta^1 = -\delta^1$$

En general,

$$\frac{\partial \varepsilon}{\partial b^k} = - [I]_{nn(k)} \delta^k = -\delta^k$$

$$b^1(n+1) = b^1(n) + \alpha \delta^1(n)$$

$$b^k(n+1) = b^k(n) + \alpha \delta^k(n)$$

#### 4.1.4. Resumen (Algoritmo)

##### Datos de Entrada

- Patrones de entrenamiento  $X$  con su salida deseada.
- $[X]_{nn(0)Xnp}$  donde  $np$  = número de patrones.
- $[d]_{nn(c)Xnp}$  donde  $nc$  = número de capas de la red (recuerde que la capa de entrada no se considera como una capa de la red).

- Número máximo de iteraciones:  $NUMIT$

- Valor de la tasa de aprendizaje:  $\alpha$

- Valor de la tolerancia:  $TOL$

- Inicializar:

$$W_{(1)}^1, W_{(1)}^2, \dots, W_{(1)}^{nc}$$
$$b_{(1)}^1, b_{(1)}^2, \dots, b_{(1)}^{nc}$$

- con valores aleatorios pequeños,

$$[W_{ij}^k]_{nn(k)Xnn(k-1)} \in [-0,5, 0,5]$$

$$[b_i^k]_{nn(k)} \in [-0,5, 0,5]$$

$$n = 0, \quad \varepsilon = 0$$

- Mientras que  $n \leq NUMIT$  ó  $\varepsilon < TOL$

$$y_{(n)}^0 = X_{(n)}, \text{ donde } X_{(n)} \text{ es un vector de entrada escogido aleatoriamente}$$



#### 4.1.5. Propagación hacia adelante.

- Para  $k = 1$  hasta  $nc$
- $S_{(n)}^k = W_{(n)}^k \bullet y_{(n)}^{k-1} + b_{(n)}^k$
- $y_{(n)}^k = F^k \left( S_{(n)}^k \right)$
- Fin para  $k$ .
- $e(n) = d(n) - y^{nc}(n)$
- $\varepsilon(n) = e(n)^T * e(n)$

#### 4.1.6. Propagación hacia atrás.

- $\delta^{nc}(n) = 2F^{nc\prime}(S^{nc}(n)) e(n)$
- $W^{nc}(n+1) = W^{nc}(n) + \alpha \delta^{nc}(n) (y^{nc-1}(n))^T$
- $b^{nc}(n+1) = b^{nc}(n) + \alpha \delta^{nc}(n)$
- Para  $k = nc - 1$  hasta  $1$
- $\delta^k(n) = F^{k\prime}(S^k(n)) (W^{k+1}(n))^T \delta^{k+1}(n)$
- $W^k(n+1) = W^k(n) + \alpha \delta^k(n) (y^{k-1}(n))^T$
- $b^k(n+1) = b^k(n) + \alpha \delta^k(n)$
- Fin para  $k$

# APLICACIÓN A REDES NEURONALES

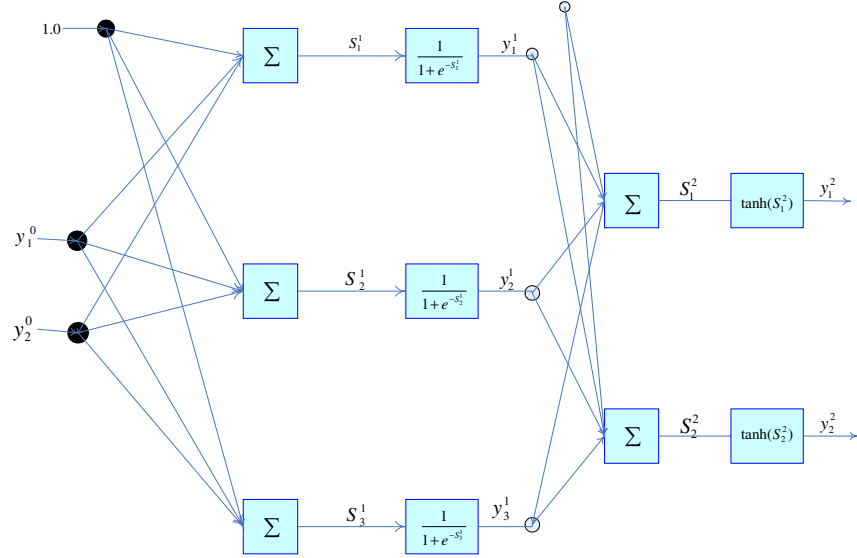


Figura 3:

**Ejemplo 4.1** Entrenar una red neuronal para aprender los siguientes patrones

| n | $x_1$ | $x_2$ | $d_1$ | $d_2$ |
|---|-------|-------|-------|-------|
| 1 | -1    | -1    | -1    | -1    |
| 2 | -1    | 1     | -1    | 1     |
| 3 | 1     | -1    | -1    | 1     |
| 4 | 1     | 1     | 1     | -1    |

Utilice una red 2-3-2 (2 entradas, 3 neuronas en la capa oculta y dos neuronas en la capa de salida).

Utilice función logística como función de activación para todas las neuronas de la capa oculta (capa 1) y tanh para todas las neuronas de la capa de salida (capa 2).

Se va a trabajar con  $\alpha = 0,5$

La figura ?? muestra la red propuesta.

Primer Patrón  $n = 0$

Inicialización:

$$W^1(1) = \begin{bmatrix} 0,7826 & 0,609 \\ -1,4372 & -1,6513 \\ -2,713 & -2,8013 \end{bmatrix}, W^2(1) = \begin{bmatrix} 0,0636 & -0,6596 & -2,0639 \\ 0,582 & -1,8127 & 2,2377 \end{bmatrix}$$

$$b^1(1) = \begin{bmatrix} 0,437 \\ -1,298 \\ 2,2137 \end{bmatrix}, b^2(1) = \begin{bmatrix} -0,7279 \\ -1,9579 \end{bmatrix}$$

$$y^0(1) = X_{(1)} = \begin{bmatrix} x_1(1) \\ x_2(1) \end{bmatrix} = \begin{bmatrix} -1 \\ -1 \end{bmatrix}$$

$$S^1(1) = W^1(1) \cdot y^0(1) + b^1(1)$$

$$S^1(1) = \begin{bmatrix} 0,7826 & 0,609 \\ -1,4372 & -1,6513 \\ -2,713 & -2,8013 \end{bmatrix} \begin{bmatrix} -1 \\ -1 \end{bmatrix} + \begin{bmatrix} 0,437 \\ -1,298 \\ 2,2137 \end{bmatrix}$$

$$S^1(1) = \begin{bmatrix} -0,9576 \\ 1,7905 \\ 7,728 \end{bmatrix}$$

$$y^1(1) = \begin{bmatrix} f_1^1(S_1^1(1)) \\ f_2^1(S_2^1(1)) \\ f_3^1(S_3^1(1)) \end{bmatrix} = \begin{bmatrix} \frac{1}{1+e^{-S_1^1}} \\ \frac{1}{1+e^{-S_2^1}} \\ \frac{1}{1+e^{-S_3^1}} \end{bmatrix} = \begin{bmatrix} 0,27796067 \\ 0,85698857 \\ 0,99955987 \end{bmatrix}$$

$$S^2(1) = W^2(1) \cdot y^1(1) + b^2(1)$$

$$S^2(1) = \begin{bmatrix} 0,0636 & -0,6596 & -2,0639 \\ 0,582 & -1,8127 & 2,2377 \end{bmatrix} \begin{bmatrix} 0,27796067 \\ 0,85698857 \\ 0,99955987 \end{bmatrix} + \begin{bmatrix} -0,7279 \\ -1,9579 \end{bmatrix}$$

$$S^2(1) = \begin{bmatrix} -3,33848298 \\ -1,11287495 \end{bmatrix}$$

$$y^2(1) = \begin{bmatrix} f_1^2(S_1^2(1)) \\ f_2^2(S_2^2(1)) \end{bmatrix} = \begin{bmatrix} \tanh(S_1^2(1)) \\ \tanh(S_2^2(1)) \end{bmatrix} = \begin{bmatrix} \tanh(-3,33848298) \\ \tanh(-1,11287495) \end{bmatrix}$$

$$y^2(1) = \begin{bmatrix} -0,99748398 \\ -0,80507629 \end{bmatrix}$$

$$e(1) = d(1) - y^2(1)$$

$$e(1) = \begin{bmatrix} -1 \\ -1 \end{bmatrix} - \begin{bmatrix} -0,99748398 \\ -0,80507629 \end{bmatrix} = \begin{bmatrix} -0,00251602 \\ -0,19492371 \end{bmatrix}$$

#### 4.1.7. Propagación hacia atrás.

**Ejemplo 4.2** 1.  $\delta^2(1) = 2F^2(S^2(1))e(1)$

Recuerde que:

$$f_1^2(S_1^2(1)) = \tanh(S_2^2(1)) = y_1^2(1)$$

$$f_1^2(S_1^2(1)) = 1 - (y_1^2(1))^2$$

$$f_2^2(S_2^2(1)) = \tanh(S_2^2(1)) = y_2^2(1)$$

$$f_2^2(S_2^2(1)) = 1 - (y_2^2(1))^2$$

$$2F^2(S^2(1)) = \begin{bmatrix} f_1^2(S_1^2(1)) & 0 \\ 0 & f_2^2(S_2^2(1)) \end{bmatrix}$$

$$= \begin{bmatrix} 1 - (y_1^2(1))^2 & 0 \\ 0 & 1 - (y_2^2(1))^2 \end{bmatrix}$$

$$\delta^2(1) = 2 \begin{bmatrix} 1 - (-0,99748398) & 0 \\ 0 & 1 - (-0,80507629) \end{bmatrix} \begin{bmatrix} -0,00251602 \\ -0,19492371 \end{bmatrix}$$

$$\delta^2(1) = \begin{bmatrix} -2,52896 \times 10^{-5} \\ -0,13716866 \end{bmatrix}$$

$$W^2(2) = W^2(1) + \alpha \delta^2(1) \cdot (y^1(1))^T$$

$$W^2(2) = \begin{bmatrix} 0,0636 & -0,6596 & -2,0639 \\ 0,582 & -1,8127 & 2,2377 \end{bmatrix} +$$

$$0,5 * \begin{bmatrix} -2,52896 \times 10^{-5} \\ -0,13716866 \end{bmatrix} \begin{bmatrix} 0,27796067 & 0,85698857 & 0,99955987 \end{bmatrix}$$

$$W^2(2) = \begin{bmatrix} 0,0636 & -0,6596 & -2,0639 \\ 0,582 & -1,8127 & 2,2377 \end{bmatrix} +$$

$$\begin{bmatrix} -3,514775 \times 10^{-6} & -1,08364 \times 10^{-5} & -1,2639 \times 10^{-5} \\ -0,019063746 & -0,0587755987 & -0,06855414 \end{bmatrix}$$

$$W^2(2) = \begin{bmatrix} 0,063596485 & -0,659610836 & -2,063912639 \\ 0,562936254 & -1,871475987 & 2,169145856 \end{bmatrix}$$

$$b^2(2) = b^2(1) + \alpha \delta^2(1)$$

$$b^2(2) = \begin{bmatrix} -0,7279 \\ -1,9579 \end{bmatrix} + 0,5 *$$

$$\delta^1(1) = F^1(S^1(1)) (W^2(1))^T \delta^2(1)$$

$$\delta^1(1) = \begin{bmatrix} y_1^1(1)(1 - y_1^1(1)) & 0 & 0 \\ 0 & y_2^1(1)(1 - y_2^1(1)) & 0 \\ 0 & 0 & y_3^1(1)(1 - y_3^1(1)) \end{bmatrix} *$$

$$\begin{bmatrix} 0,0636 & 0,582 \\ -0,6596 & -1,8127 \\ -2,0639 & 2,2377 \end{bmatrix} \begin{bmatrix} -2,52896 \times 10^{-5} \\ -0,13716866 \end{bmatrix}$$

$$\delta^1(1) = \begin{bmatrix} -0,01602252 \\ 0,03047584 \\ -0,00013501 \end{bmatrix}$$

$$W^1(2) = W^1(1) + \alpha \delta^1(1) \cdot (y^0(1))^T$$

$$W^1(2) = \begin{bmatrix} 0,7826 & 0,609 \\ -1,4372 & -1,6513 \\ -2,713 & -2,8013 \end{bmatrix} + 0,5 * \begin{bmatrix} -0,01602252 \\ 0,03047584 \\ -0,00013501 \end{bmatrix} \begin{bmatrix} -1 & -1 \end{bmatrix}$$

$$W^1(2) = \begin{bmatrix} 0,7826 & 0,609 \\ -1,4372 & -1,6513 \\ -2,713 & -2,8013 \end{bmatrix} + \begin{bmatrix} 0,00801126 & 0,00801126 \\ -0,01523792 & -0,01523792 \\ 6,7506 \times 10^{-5} & 6,7506 \times 10^{-5} \end{bmatrix}$$

$$W^1(2) = \begin{bmatrix} 0,79061126 & 0,61701126 \\ -1,45243792 & -1,66653792 \\ -2,71293249 & -2,80123249 \end{bmatrix}$$

$$b^1(2) = b^1(1) + \alpha \delta^1(1)$$

$$b^1(2) = \begin{bmatrix} 0,437 \\ -1,298 \\ 2,2137 \end{bmatrix} + 0,5 * \begin{bmatrix} -0,01602252 \\ 0,03047584 \\ -0,00013501 \end{bmatrix}$$

$$b^1(2) = \begin{bmatrix} 0,42898874 \\ -1,28276208 \\ 2,21363249 \end{bmatrix}$$

## 4.2. Derivadas Matriciales en Matlab.

Aprovechando las funciones básicas y las operaciones que con respecto a matrices se pueden utilizar en Matlab, se diseñaron unas pequeñas funciones que permiten a través de la utilización de las propiedades de la derivación matricial, empleando la vectorización y el producto de Kronecker, resolver derivadas matriciales. Se adjuntan los códigos fuente de las mismas, las cuales son solo un ejemplo, teniendo en cuenta que las funciones `f1`, `f2` y demás que se presentan, pueden ser modificadas de acuerdo con la necesidad de solución.

Un exámen previo de las figuras 4 hasta la 8 puede facilitar la lectura de esta sección, ya que el lenguaje Matlab es bastante intuitivo. Inicialmente se han creado dos funciones `vec()` que realiza la vectorización de una matriz por columnas, tal como se observa en la figura 4 y `vecf()` que realiza la vectorización por filas de una matriz, tal como se observa en la figura 5.

En primera instancia, se debe tener claro que es posible definir funciones nuevas y archivos de secuencia de comandos (*script*) en Matlab. De hecho, las

```
function [vec] = vec(mat)
    %vectorizacion (columna)

    %CAPTURA TAMAÑO
    [m,n] = size(mat);
    %LEE ELEMENTOS Y REUBICA
    for I = 1:m,
        for J = 1:n,
            pos=n*(I-1)+J;
            vecf(pos) = mat(I,J);
        end
    end
    %TRANSPONE
    vec=vecf.';
end
```

Figura 4: Vectorización de una matriz.

```
function [vecf] = vecf(mat)
    %CAPTURA TAMAÑO
    [m,n] = size(mat);
    %LEE ELEMENTOS Y REUBICA
    for I = 1:m,
        for J = 1:n,
            pos=n*(I-1)+J;
            vecf(pos) = mat(I,J);
        end
    end
end
```

Figura 5: Vectorización por filas de una matriz.

```
% CREA VARIABLES
syms v
syms w
syms x
syms y
syms z

%CREA MATRICES IDENTICAS DE ORDEN 1 HASTA 5
I1=eye(1);
I2=eye(2);
I3=eye(3);
I4=eye(4);
I5=eye(5);

return
```

Figura 6: Inicialización, crea variables simbólicas y matrices idénticas.

```

% CASO:
% DERIVACIÓN DE f(x,y,z,...) RESPECTO DE CADA VARIABLE INDEPENDIENT

% BORRA VARIABLES DE LA MEMORIA
clear
% INICIALIZA VARIABLES Y CREA MATRICES UTILITARIAS
ini

% USUARIO: SE DEFINEN FUNCIONES
% EJEMPLO:
f1=x^2+y^3;
f2=2*x*y;
%Fn=...

% USUARIO: SE DEFINEN VECTORES O MATRICES
% EJEMPLO:
Y=[f1 ; f2];
X=[x y];

%%% PROCESANDO... %%%

% EXTRAER EL TAMAÑO DE LOS VECTORES O MATRICES:
[P,N]=size(X);
[M,Q]=size(Y);

% EXPANSION:
for I = 1:M,
    for J = 1:Q,
        %DERIVAR CADA ELEMENTO DE Y RESPECTO A MATRIZ X:
        for K=1:P,
            for L=1:N,
                %DERIVACIÓN INDIVIDUAL:
                R(P*(I-1)+K,N*(J-1)+L)=diff(Y(I,J),X(K,L));
            end
        end
    end
end
R

```

Figura 7: Código completo para realizar una derivación matricial simple.

figuras referenciadas muestran código de cada uno de ellos. Tanto las funciones como los scripts pueden utilizar funciones de control de flujo para detallar la forma como se desea ejecutar una secuencia de procesos. Por ejemplo, en Matlab se dispone de los comandos: *for*, *if*, *else*, *elseif*, *while*, *return*, *break*, con un uso bastante similar al de otros programas más conocidos, aunque con una sintaxis ligeramente diferente.

De interés particular para este trabajo, se encuentran las funciones para la generación, manipulación de matrices y operaciones con matrices. Algunas de ellas son:

- **trace(M)**: calcula la traza de una matriz.  $M$
- **kron(M, N)** : desarrolla el producto de kronecker de las matrices  $M$  y  $N$ .
- **eye(M, N)**: devuelve una matriz idéntica de tamaño  $M \times N$ .
- **ones(M, N)**: devuelve una matriz de unos de tamaño  $M \times N$ .

```

%BORRA VARIABLES DE LA MEMORIA
clear
%INICIALIZA VARIABLES Y CREA IDENTICAS
ini

%SE DEFINEN FUNCIONES
f1=2*x+3*y;
f2=x^2+2*y;
f3=exp(x);
f4=2*x*y;
f5=1/x;
f6=4*y-x^2;

%SE DEFINEN MATRICES
X=[f1 f2 ; f3 f4; f5 f6];

[filas,col]=size(X);

dXrX=(vec(eye(filas))*vecf(eye(col)))
R=kron(eye(col),dXrX)*kron(vec(eye(col)),eye(col))

```

Figura 8: Desarrollo de  $\frac{\delta \text{vec}(X)}{\delta X}$  utilizando las propiedades.

- **zeros(n)**: devuelve una matriz idéntica de tamaño  $M \times N$ .
- **:** (dos puntos) crea un vector espaciado regularmente.
- **linspace**: crea un vector espaciado linealmente.
- **logspace**: crea un vector espaciado logarítmicamente.
- **diag**: crea o extrae diagonales.
- **reshape**: cambia la dimensión de la matriz.
- **tril**: extrae la matriz triangular inferior.
- **triu**: extrae la matriz triangular superior.
- **det**: calcula el determinante de una matriz.
- **eig**: calcula los autovalores y autovectores.
- **expm**: calcula la exponencial de una matriz.
- **logm**: calcula el logaritmo de una matriz.

Se pueden consultar funciones adicionales mediante el menú de ayuda del programa o mediante el comando *help* de Matlab.

El código que muestra la figura 6 tiene por objeto crear variables simbólicas, con el fin de poder utilizarlas dentro de expresiones más complejas. Además, crea un conjunto de matrices idénticas, para utilizar cuando se necesiten.



En la figura 7, se expone el código necesario para derivar una función matricial de variable matricial, aunque se ha desarrollado como primer ejemplo el caso particular en que tanto la función como la variable son vectoriales.

En el ejemplo de la figura 8 se muestra como resolver el problema de  $\frac{\delta \text{vec}(X)}{\delta X}$ .

Para ello, después de observar las propiedades que cumple este desarrollo, se ha planteado la solución como lo indica el código allí expuesto. Al utilizar las propiedades de la derivación, el proceso se hace más simple y menos costoso computacionalmente.

Los resultados de la ejecución de los ejercicios expuestos coinciden plenamente con los desarrollos mediante otros paquetes o técnicas, siempre y cuando se definan correctamente los tipos de datos (entero, real, etc).

Mediante la aplicación iterativa de las fórmulas presentadas, se pueden obtener las derivadas de orden superior o la aplicación de la regla de la cadena. En primera instancia, hay una gran cantidad de opciones para desarrollar en este caso, cada una de las cuales requiere que se definan correctamente las variables dependientes e independientes, la relación funcional, y las matrices o vectores que intervienen.

Matlab es un lenguaje de programación que facilita el trabajo con matrices. Dado que la representación interna de Matlab es en forma matricial, los programas finales son compactos y relativamente sencillos.

## 5. Conclusiones

Al finalizar el trabajo, es necesario presentar algunos puntos importantes, considerados como vitales en la comprensión y realización del mismo.

Al iniciar el estudio de la derivación matricial es fundamental tener claridad en conceptos y definiciones de la teoría básica del álgebra lineal y del cálculo diferencial.

Antes de introducirse directamente en el tema, es necesario comprender las operaciones que complementan el álgebra lineal y que son el soporte de la derivación matricial, tales como la permutación de matrices por bloques, el producto de Kronecker y la vectorización de matrices por fila y por columna.

Al abordar la derivación matricial es fundamental analizar las propiedades del producto de Kronecker, de la vectorización de matrices, las reglas de derivación matricial y las formas alternativas de derivación con el fin de minimizar la solución de problemas, teniendo en cuenta que en algunos casos la expansión de las matrices es bastante amplia y utilizando adecuadamente la o las reglas enumeradas anteriormente, el procedimiento de solución es más ágil, más eficiente y es posible que la expansión de las matrices sea más reducida.

A nivel personal puedo concluir que adquirí el conocimiento de nuevas operaciones del álgebra lineal y el del tema central: el cálculo matricial, lo que significa un avance personal muy interesante, que aunque mínimo si se tiene en cuenta la amplitud del campo matemático, si muy importante en el desarrollo y el conocimiento de nuevos temas relacionados con la matemática.

A nivel profesional, es como haber dado el primer paso en la profundización de conceptos y conocimientos, los cuales se logran solo a través del tiempo y con la dedicación necesaria que permiten realmente formar al verdadero matemático.

Por último, espero que el trabajo haya respondido a las expectativas de las directivas y docentes de la Fundación y que sirva como un aporte interesante no solo para los matemáticos, sino para los profesionales que de una u otra forma en sus aplicaciones requieren de la derivación matricial.

## Referencias

- [Her86] I. Herstein, *Álgebra Abstracta*. México: Grupo Editorial Iberoamericano (1985).
- [Bar98] R. Barbolla, P. Sans, *Álgebra Lineal y Teoría de Matrices*. Madrid: Prentice-Hall (1998).
- [San94] R. Sánchez, A. Velasco, *Curso Básico de Álgebra Lineal*. Bogotá: Ed Trillas (1994).
- [Mey07] C. Meyer, *Matrix Analysis and Applied Linear Algebra*. <http://www.matrixanalysis.com/> (2007)
- [Rad99] C. Radhakrishna, H. Toutenburg, *Linear Models: Least Squares and Alternatives*. New York: Springer (1999).
- [Per07] P Rengifo. *Borradores de Trabajo y Notas de Clase Sistemas Basados en el Conocimiento*. FUKL. (2007)
- [Mag06] J. Magnus, *Matrix Calculus and Econometrics*. Tilburg University (2006)