

eman ta zabal zazu



Universidad  
del País Vasco

Euskal Herriko  
Unibertsitatea

# ANÁLISIS MATRICIAL APLICADO Y AMPLIACIÓN DE MÉTODOS NUMÉRICOS

Ion Zaballa

Departamento de Matemática Aplicada  
y Estadística e Investigación Operativa

Euskal Herriko Unibertsitatea



# Capítulo 1

## Vectores y Matrices

### 1.1. Matrices y Aplicaciones Lineales

El objetivo de este capítulo es fijar algunas notaciones y convenciones, y también dar un repaso rápido a algunos conceptos básicos de álgebra lineal pero desde un punto de vista que, quizás, no es el habitual.

Comenzamos recordando que una matriz no es más que una familia de elementos expuestos en un determinado orden de forma que su apariencia externa es la de un rectángulo de elementos ordenados:

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

Esta parece ser la forma en que Sylvester (1885) introdujo originalmente las matrices. De forma abreviada escribiremos  $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$  para indicar la matriz que tiene

en la posición o entrada  $(i, j)$  el elemento  $a_{ij}$ . También diremos que  $a_{ij}$  es el elemento  $(i, j)$  de  $A$ . Si  $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$  entonces diremos que  $A$  tiene  $m$  filas y  $n$  columnas y que  $A$  es de orden o tamaño  $m \times n$  o simplemente que es una matriz  $m \times n$ . Las matrices con igual número de filas y columnas se llaman cuadradas. La  $i$ -ésima fila y  $j$ -ésima columna de  $A$  son las matrices

$$a^i = [a_{i1} \ a_{i2} \ \dots \ a_{in}] \quad \text{y} \quad a_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix},$$

respectivamente.

Los elementos de una matriz pueden ser de muy diferente naturaleza: números, polinomios, funciones continuas, etc. Por lo general son elementos de algún conjunto con estructura de anillo (normalmente conmutativo y con elemento 1) de forma tal que si  $\mathfrak{R}$  es uno de tales anillos entonces podemos definir formalmente una matriz  $A$  de orden  $m \times n$  como una aplicación:

$$\mu : I \times J \rightarrow \mathfrak{R}$$

donde  $I = \{1, 2, \dots, m\}$ ,  $J = \{1, 2, \dots, n\}$  y  $\mu(i, j) = a_{ij}$ . De esta forma  $A$  es la aplicación  $\mu$  y dos matrices son iguales si lo son las aplicaciones que las definen. Es decir,  $A = [a_{ij}]_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}}$  y  $B = [b_{ij}]_{\substack{1 \leq i \leq p \\ 1 \leq j \leq q}}$  son iguales si y sólo si  $n = q$ ,  $m = p$  y  $a_{ij} = b_{ij} \ \forall i, j$ . Denotaremos con  $\mathfrak{R}^{m \times n}$  al conjunto de las matrices de tamaño  $m \times n$  con elementos en  $\mathfrak{R}$ . Podemos entonces definir la suma y producto de matrices: Si  $A = [a_{ij}] \in \mathfrak{R}^{m \times n}$  y  $B = [b_{ij}] \in \mathfrak{R}^{m \times n}$  son matrices del mismo tamaño entonces

$$A + B = [a_{ij} + b_{ij}] \in \mathfrak{R}^{m \times n};$$

y si  $A = [a_{ij}] \in \mathfrak{R}^{m \times n}$  y  $B = [b_{ij}] \in \mathfrak{R}^{n \times p}$ , entonces

$$AB = \left[ \sum_{k=1}^n a_{ik} b_{kj} \right] \in \mathfrak{R}^{m \times p}.$$

En particular, si  $n = m = p$  entonces cualesquiera dos matrices de  $\mathfrak{R}^{n \times n}$  se pueden sumar y multiplicar y  $(\mathfrak{R}^{n \times n}, +, \cdot)$  es un anillo **no conmutativo** con elemento

identidad. Este elemento es la matriz

$$I_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = [\delta_{ij}],$$

llamada *matriz identidad*. Las unidades de un anillo son aquellos elementos que admiten un inverso. En nuestro caso, y dado que  $\mathfrak{R}^{n \times n}$  es un anillo no conmutativo, diremos que una matriz  $A \in \mathfrak{R}^{n \times n}$  es una unidad, que admite inversa o, más comúnmente, que es invertible si existe una única  $B \in \mathfrak{R}^{n \times n}$  tal que  $AB = BA = I_n$ . Si  $A$  es invertible, a la única matriz  $B$  tal que  $AB = BA = I_n$  se le llama inversa de  $A$  y se representa por  $A^{-1}$ . Se puede demostrar que si  $\mathfrak{R} = \mathbb{F}$  es un cuerpo y  $AB = I_n$  entonces  $A^{-1} = B$ . Esto es útil porque nos dice que basta comprobar que  $A$  es invertible por la derecha (o izquierda) para demostrar que lo es por los dos lados.

Debe observarse que si  $A_1, A_2 \in \mathfrak{R}^{n \times n}$  son invertibles entonces el producto  $A_1 A_2$  lo es y  $(A_1 A_2)^{-1} = A_2^{-1} A_1^{-1}$ . As pues, el conjunto de las matrices invertibles  $n \times n$  es un grupo, no abeliano, llamado *grupo general lineal* de orden  $n$  sobre  $\mathfrak{R}$  y que denotaremos por  $\text{Gl}_n(\mathfrak{R})$ .

Las matrices de  $\mathfrak{R}^{m \times n}$  se pueden multiplicar por elementos de  $\mathfrak{R}$  (escalares): Si  $x \in \mathfrak{R}$  y  $A = [a_{ij}] \in \mathfrak{R}^{m \times n}$  entonces

$$xA = [xa_{ij}] \in \mathfrak{R}^{m \times n}.$$

De esta forma el conjunto  $\mathfrak{R}^{m \times n}$  tiene estructura de *módulo libre* sobre  $\mathfrak{R}$ . Una base de este módulo es el conjunto  $\{E_{ij} : 1 \leq i \leq m, 1 \leq j \leq n\}$  siendo  $E_{ij}$  la matriz cuyos elementos son todos cero excepto el de la posición  $(i, j)$  que es 1. Dado que  $\mathfrak{R}$  es conmutativo con elemento identidad, todas las bases de  $\mathfrak{R}^{m \times n}$  tienen el mismo número de elementos (a este número se le llama *dimensión* o *rango* del módulo), en este caso  $mn$ . En particular, si  $\mathfrak{R} = \mathbb{F}$  es un cuerpo, entonces  $\mathbb{F}^{m \times n}$  es un espacio vectorial y  $\mathbb{F}^{n \times n}$  un álgebra, no conmutativa, de dimensión  $n^2$ .

### 1.1.1. Algunos tipos de matrices

#### 1. Transpuesta y transpuesta conjugada

Dada una matriz  $A \in \mathfrak{R}^{m \times n}$  a la matriz que se obtiene al intercambiar las filas por las columnas se le llama matriz *transpuesta* de  $A$  y se denota por  $A^T$ . Así, si  $A = [a_{ij}] \in \mathfrak{R}^{m \times n}$  entonces  $A^T = [a_{ji}] \in \mathfrak{R}^{n \times m}$ . Por su parte, si  $\mathfrak{R} = \mathbb{C}$ , el cuerpo de los números complejos y la unidad imaginaria la representamos con  $i$  ( $= \sqrt{-1}$ ), entonces cada elemento  $a_{kj} = x_{kj} + iy_{kj}$  de una matriz  $A$  admite un conjugado  $\bar{a}_{kj} = x_{kj} - iy_{kj}$ . La matriz  $\bar{A} = [\bar{a}_{ij}]$  se llama matriz *conjugada* de  $A$ . La transpuesta conjugada de  $A \in \mathbb{C}^{m \times n}$  la denotaremos por  $A^* \in \mathbb{C}^{n \times m}$ .

Algunas propiedades elementales de la transposición son las siguientes:

- $(A^T)^T = A$ ,  $(A^*)^* = A$
- Si  $\lambda$  es una escalar entonces  $(\lambda A)^T = \lambda A^T$  y  $(\lambda A)^* = \bar{\lambda} A^*$
- $(A + B)^T = A^T + B^T$ ,  $(A + B)^* = A^* + B^*$
- $(AB)^T = BA$ ,  $(AB)^* = B^* A^*$
- $(A^T)^{-1} = (A^{-1})^T$ ,  $(A^*)^{-1} = (A^{-1})^*$ .

## 2. Matrices Simétricas y Hermíticas

Las matrices que tienen la propiedad de ser iguales a sus transpuestas,  $A = A^T$ ; i. e. matrices  $A = [a_{ij}]$  tales que  $a_{ij} = a_{ji}$  para todos  $i$  y  $j$ , reciben el nombre de matrices *simétricas*. El nombre es debido, posiblemente, a que sus elementos son simétricos respecto de la diagonal principal. Es claro que estas matrices deben ser cuadradas.

Las matrices *skew-simétricas* son las que cumple que  $A^T = -A$ .

Por su parte, si  $\mathfrak{R} = \mathbb{C}$  y  $A = A^*$  entonces  $A$  se dice que es *hermítica*. Y si  $A = -A^*$  entonces  $A$  es *skew-hermítica*.

Algunos ejemplos de estos tipos de matrices son los siguientes:

$$\begin{array}{ll}
 \begin{bmatrix} i & 2i \\ 2i & 4 \end{bmatrix} \text{ simétrica} & \begin{bmatrix} 1 & 2i \\ -2i & 4 \end{bmatrix} \text{ hermítica} \\
 \begin{bmatrix} 0 & 2i \\ -2i & 0 \end{bmatrix} \text{ skew-simétrica} & \begin{bmatrix} i & 2i \\ 2i & 4i \end{bmatrix} \text{ skew-hermítica} \\
 \begin{bmatrix} 0 & i \\ i & 0 \end{bmatrix} \text{ simétrica y skew-hermítica} & \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix} \text{ hermítica y skew-simétrica}
 \end{array}$$

Cualquiera que sea  $A \in \mathfrak{R}^{n \times m}$ , las matrices  $AA^T$ ,  $A^T A$  y  $A + A^T$  son simétricas; mientras que si  $A \in \mathbb{C}^{m \times n}$  entonces  $AA^*$ ,  $A^* A$  y  $A + A^*$  son hermíticas.

### 3. Matrices Diagonales

Una matriz  $n \times n$  se dice que es diagonal si todos sus elementos son cero excepto los de las posiciones  $(i, i)$ ,  $1 \leq i \leq n$ . Si  $D \in \mathfrak{R}^{n \times n}$  y  $d_i$  es elemento que ocupa la posición  $(i, i)$ , escribiremos:

$$D = \text{Diag}(d_1, \dots, d_n).$$

En general si  $A \in \mathfrak{R}^{n \times n}$  los elementos  $a_{ii}$ ,  $1 \leq i \leq n$ , se dice que son o forman la diagonal principal de  $A$ . Las matrices diagonales se pueden considerar como un caso especial de matrices *diagonales por bloques*. Éstas son las que tienen la forma:

$$\begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_p \end{bmatrix}$$

siendo  $A_i$  una matriz cuadrada de tamaño  $n_i \times n_i$ ,  $1 \leq i \leq p$ .

### 4. Matrices Triangulares

Una matriz  $A = [a_{ij}] \in \mathfrak{R}^{n \times n}$  se dice que es *triangular superior* si  $a_{ij} = 0$  para  $i > j$ . Si  $a_{ij} = 0$  para  $i < j$  entonces  $A$  es triangular inferior. De la misma forma que con matrices diagonales, las triangulares son un caso particular de las matrices *triangulares por bloques*. Una matriz triangular superior por bloques tiene la siguiente forma:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1p} \\ 0 & A_{22} & \dots & A_{2p} \\ \vdots & \vdots & \dots & \vdots \\ 0 & 0 & \dots & A_{pp} \end{bmatrix}.$$

con  $A_{ij} \in \mathfrak{R}^{n_i \times n_j}$ .

Las inversas de matrices triangulares superiores (inferiores) son triangulares superiores (inferiores, respectivamente).

### 5. Matrices de Permutación

Una matriz  $P \in \mathfrak{R}^{n \times n}$  se dice que es de permutación si en cada fila y columna hay un elemento igual a 1 y todos los demás son cero. La multiplicación por

este tipo de matrices efectúa una permutación de las filas o columnas del objeto multiplicado. Así por ejemplo,

$$P = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{y} \quad P^T = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{bmatrix}$$

son matrices de permutación y

$$P \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 1 \\ 2 \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} 1 & 2 & 3 \end{bmatrix} P^T = \begin{bmatrix} 3 & 1 & 2 \end{bmatrix}.$$

Es decir,  $P$  pone la fila 1 en la 2, la 2 en la 3 y la 3 en la 1. Y  $P^T$  hace lo mismo pero en las columnas.

En general, si  $S_n$  es el grupo simétrico de orden  $n$ ; i. e. el grupo de permutaciones de  $n$  elementos, y  $\sigma = (i_1, i_2, \dots, i_n) \in S_n$ ; i. e.  $\sigma(k) = i_k$ , denotamos con  $P_\sigma = [\delta_{i\sigma(j)}] = [\delta_{ii_k}] \in \mathfrak{R}^{n \times n}$ . Es decir,  $P_\sigma$  es la matriz que tiene exactamente un 1 en la posición  $(\sigma(i), i)$ ,  $1 \leq i \leq n$  y todos los demás elementos son cero. Así, en el ejemplo de arriba,  $P$  es la matriz de la permutación  $\sigma = (2, 3, 1)$ .

Notemos que  $P^T = P_{\sigma^{-1}} = P^{-1}$  y si  $A = [a_{ij}]$  entonces

$$P_\sigma A = [a_{\sigma(i)j}] \quad \text{y} \quad AP^T = [a_{i\sigma(j)}].$$

Es decir,  $P_\sigma A$  se obtiene de  $A$  poniendo la fila  $i$  en la  $\sigma(i)$ , y  $AP^T$  se obtiene de  $A$  poniendo la columna  $j$  en la  $\sigma(j)$ .

Hay otros tipos de matrices que iremos definiendo a medida que los vayamos necesitando.

### 1.1.2. Aplicaciones lineales

Si  $\mathfrak{R} = \mathbb{F}$  es un cuerpo (sobreentendemos que es conmutativo) hay una relación muy estrecha entre las matrices con elementos en  $\mathbb{F}$  y las aplicaciones lineales definidas entre espacios vectoriales sobre  $\mathbb{F}$ . En efecto, sean  $\mathcal{V}_1$  y  $\mathcal{V}_2$  espacios vectoriales de dimensiones  $n$  y  $m$ , respectivamente, sobre  $\mathbb{F}$ , y fijemos bases  $\mathcal{B}_1$  y  $\mathcal{B}_2$  de  $\mathcal{V}_1$  y  $\mathcal{V}_2$ . Así a cada aplicación lineal  $f : \mathcal{V}_1 \rightarrow \mathcal{V}_2$  le corresponde una matriz  $A = [a_{ij}] \in \mathbb{F}^{m \times n}$



de la siguiente forma: los elementos de la  $j$ -ésima columna de  $A$  son las componentes de la imagen del  $j$ -ésimo vector de  $\mathcal{B}_1$  respecto de la base  $\mathcal{B}_2$ . Es decir, si  $\mathcal{B}_1 = \{v_1, \dots, v_n\}$ ,  $\mathcal{B}_2 = \{u_1, \dots, u_m\}$ , y

$$f(v_j) = \sum_{i=1}^m a_{ij} u_i, \quad j = 1, \dots, n \quad (1.1)$$

entonces la matriz  $A = [a_{ij}] \in \mathbb{F}^{m \times n}$  se llama la matriz de  $f$  respecto de las bases  $\mathcal{B}_1$  y  $\mathcal{B}_2$ . Recíprocamente, dada la matriz  $A = [a_{ij}] \in \mathbb{F}^{m \times n}$  y fijadas las bases  $\mathcal{B}_1$  y  $\mathcal{B}_2$  de  $\mathcal{V}_1$  y  $\mathcal{V}_2$ , respectivamente, la relación (1.1) define una única aplicación lineal entre estos espacios vectoriales. Dicho formalmente, si  $\text{Hom}(\mathcal{V}_1, \mathcal{V}_2)$  es el espacio vectorial de las aplicaciones lineales de  $\mathcal{V}_1$  en  $\mathcal{V}_2$ , y fijadas bases en estos dos espacios vectoriales, los espacios vectoriales  $\mathbb{F}^{m \times n}$  y  $\text{Hom}(\mathcal{V}_1, \mathcal{V}_2)$  son isomorfos y el isomorfismo está definido a partir de la relación (1.1). Conviene tener siempre presente este isomorfismo porque nos permite traducir propiedades de las matrices en propiedades de las aplicaciones lineales entre espacios vectoriales y al revés. En particular, toda matriz  $A \in \mathbb{F}^{m \times n}$  puede verse como una aplicación lineal entre los espacios vectoriales  $\mathbb{F}^m$  y  $\mathbb{F}^n$ . Para ello identificamos  $\mathbb{F}^n$  con  $\mathbb{F}^{n \times 1}$ ; es decir, escribimos y hablamos indistintamente del vector  $x = (x_1, x_2, \dots, x_n) \in \mathbb{F}^n$  y de la matriz columna (a la que llamaremos con mucha frecuencia vector columna):

$$x = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix}^T \in \mathbb{F}^{n \times 1}.$$

De esta forma, dada  $A \in \mathbb{F}^{m \times n}$ , queda definida la aplicación lineal (que seguimos denotando con la letra  $A$ ):

$$\begin{aligned} A &: \mathbb{F}^n \rightarrow \mathbb{F}^m \\ x &\rightsquigarrow Ax \end{aligned}$$

donde debemos entender que  $Ax$  es el vector columna que resulta de multiplicar  $A$  y el vector columna  $x$ .

### 1.1.3. Dos interpretaciones del producto de matrices

Todo lo dicho hasta aquí concuerda con la interpretación habitual del Álgebra lineal. Desde este punto de vista la ecuación  $Ax = b$ , que es un sistema no homogéneo

de ecuaciones lineales, se debe interpretar de la siguiente forma: *b es la imagen por A del vector x*. Por supuesto, esto es correcto, pero hay otra forma de mirar esta ecuación que es mucho más útil para este curso y que pasamos a analizar ahora.

En primer lugar,  $b = Ax$  significa que el vector  $b$  es el resultado de multiplicar la matriz  $A$  y el vector  $x$ :

$$b_i = \sum_{j=1}^n a_{ij}x_j, \quad i = 1, \dots, m.$$

Poniendo todas las componentes de  $b$  juntas:

$$\begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix} = \begin{bmatrix} a_{11} \\ a_{21} \\ \vdots \\ a_{m1} \end{bmatrix} x_1 + \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} x_2 + \dots + \begin{bmatrix} a_{1n} \\ a_{2n} \\ \vdots \\ a_{mn} \end{bmatrix} x_n.$$

y escribiendo  $A$  en función de sus columnas:

$$A = [a_1 \quad a_2 \quad \dots \quad a_n],$$

tenemos que

$$b = x_1 a_1 + x_2 a_2 + \dots + x_n a_n.$$

Es decir *b es una combinación lineal de las columnas de A cuyos coeficientes son las componentes del vector x*. Así pues, no es  $A$  quien actúa sobre  $x$  para producir  $b$ , sino que es  $x$  quien actúa sobre  $A$  para producir  $b$ . Este es el punto de vista habitual que adoptaremos en este curso.

De la misma forma, si  $A \in \mathfrak{R}^{m \times n}$  y  $B \in \mathfrak{R}^{n \times p}$  tenemos una doble interpretación para la matriz  $C = AB \in \mathfrak{R}^{m \times p}$ . La tradicional será que  $C$  es la matriz de la composición de las aplicaciones

$$\mathfrak{R}^p \xrightarrow{B} \mathfrak{R}^n \xrightarrow{A} \mathfrak{R}^m$$

en las bases canónicas. La que más nos interesa en este curso es la siguiente: Si  $c_{ij}$  es el elemento en la posición  $(i, j)$  de  $C$  entonces

$$c_{ij} = \sum_{k=1}^n a_{ik} b_{kj}, \quad 1 \leq i \leq m, \quad 1 \leq j \leq p.$$

Por consiguiente

$$c_{ij} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}, \quad (1.2)$$

y si  $c_j$  y  $b_j$  son las  $j$ -ésimas columnas de  $C$  y  $B$ :

$$c_j = \begin{bmatrix} c_{1j} \\ c_{2j} \\ \vdots \\ c_{mj} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix} \begin{bmatrix} b_{1j} \\ b_{2j} \\ \vdots \\ b_{nj} \end{bmatrix} = Ab_j, \quad j = 1, \dots, p.$$

Así

$$\begin{bmatrix} c_1 & c_2 & \cdots & c_p \end{bmatrix} = \begin{bmatrix} Ab_1 & Ab_2 & \cdots & Ab_p \end{bmatrix},$$

de modo que la  $j$ -ésima columna de  $C$  es una combinación lineal de las columnas de  $A$  cuyos coeficientes son las componentes de la  $j$ -ésima columna de  $B$ .

De forma similar, de (1.2) sacamos que si  $c^i$  y  $a^i$  son las  $i$ -ésimas filas de  $C$  y  $A$ , entonces

$$c^i = a^i B.$$

Y esto significa que la  $i$ -ésima fila de  $C$  es una combinación lineal de las filas de  $B$  cuyos coeficientes son las componentes de la  $i$ -ésima fila de  $A$ .

En resumen, en el producto de matrices  $AB$ :

- $A$  acta sobre  $B$  provocando combinaciones lineales en las filas de  $B$ , y
- $B$  acta sobre  $A$  provocando combinaciones lineales en las columnas de  $A$ .

### Ejemplo 1.1 : Producto interno y producto externo de vectores.

Sean  $u, v \in \mathfrak{R}^n$ . Se define le **producto interno** de estos dos vectores como  $u^T v$ , mientras que el **producto externo** es  $uv^T$ .

Hay una gran diferencia entre estos dos productos. Para empezar  $u^T v \in \mathfrak{R}^{1 \times 1}$ ; es decir, es un escalar mientras que  $uv^T \in \mathfrak{R}^{n \times n}$  es una matriz. Además,  $u^T v$  es una combinación lineal de las columnas de  $u^T$  con los coeficientes de  $v$  o una combinación lineal de las filas de  $v$  con los coeficientes de  $u$ . Tanto las columnas de  $u^T$  como las

filas de  $v$  son de tamaño  $1 \times 1$ , de modo que el resultado es un escalar:

$$u^T v = \begin{bmatrix} u_1 & u_2 & \cdots & u_n \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} = \sum_{i=1}^n u_i v_i \in \mathfrak{R}.$$

Por su parte  $uv^T$  se puede ver como una combinación de las filas de  $v^T$  (sólo hay una) cuyos coeficientes son las componentes de  $u$ ; o, como una combinación lineal de las columnas de  $u$  (sólo hay una) cuyos coeficientes son las componentes de  $v^T$ . Es decir,

$$uv^T = \begin{bmatrix} v_1 u & v_2 u & \cdots & v_n u \end{bmatrix} = \begin{bmatrix} u_1 v^T \\ u_2 v^T \\ \vdots \\ u_n v^T \end{bmatrix} = \begin{bmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \cdots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n v_1 & u_n v_2 & \cdots & u_n v_n \end{bmatrix}$$

En conclusión, el producto exterior de  $u$  y  $v$  es una matriz  $n \times n$  cuyas columnas son "múltiplos" del vector  $u$ , y cuyas filas son "múltiplos" del vector  $v^T$ .

### Otras dos interpretaciones del producto de matrices

Se pueden dar dos nuevas interpretaciones del producto de matrices a partir de los productos interno y externo de vectores. Sean  $A \in \mathfrak{R}^{m \times n}$  y  $B \in \mathfrak{R}^{n \times p}$ . Entonces

- Por una parte, el elemento  $(i, j)$  de  $AB$  es  $(AB)_{ij} = \sum_{k=1}^n a_{ik} b_{kj}$ . Es decir, es el producto interno de la  $i$ -ésima fila de  $A$  y la  $j$ -ésima columna de  $B$ . Siendo completamente formales:

$$(AB)_{ij} = (a'_i)^T b_j,$$

donde  $a'_i$  es la  $i$ -ésima columna de  $A^T$  y  $b_j$  la  $j$ -ésima columna de  $B$ .

- Por otra, si  $a_i$  representa la columna  $i$ -ésima de  $A$  y  $b^j$  la  $j$ -ésima fila de  $B$ , entonces el elemento  $(i, j)$  del producto externo  $a_k b^k$  es precisamente  $a_{ik} b_{kj}$ . Por

lo tanto, si  $b'_k$  es la  $k$ -ésima columna de  $B^T$  tenemos que  $(a_k(b'_k)^T)_{ij} = a_{ik}b_{kj}$ . Es decir,

$$AB = a_1(b'_1)^T + a_2(b'_2)^T + \cdots + a_n(b'_n)^T,$$

es la suma de  $n$  productos externos. Cada sumando es el producto externo de la  $i$ -ésima columna de  $A$  y la  $i$ -ésima fila de  $B$  (formalmente la  $i$ -ésima columna de  $B^T$ ).

### Ejemplo 1.2 :Matrices elementales

Dada una matriz  $A \in \mathfrak{R}^{m \times n}$ , hay tres tipos de **transformaciones elementales** que se pueden realizar sobre  $A$ :

- ( $t_1$ ) Sumar a una fila (o columna) de  $A$  otra fila (o columna) multiplicada por un escalar.
- ( $t_2$ ) Multiplicar una fila (o columna) de  $A$  por un escalar distinto de cero.
- ( $t_3$ ) Permutar dos filas (o columnas) de  $A$ .

Si las transformaciones se hacen en las filas se habla de **transformaciones elementales por filas** y si se hacen en las columnas de **transformaciones elementales por columnas**.

Puesto que las transformaciones elementales por filas en  $A$  producen combinaciones lineales de sus filas, se tienen que poder realizar multiplicando una matriz apropiada por  $A$ . Y lo mismo por columnas: las transformaciones elementales por columnas se pueden realizar multiplicando  $A$  por una matriz apropiada. Supongamos, por ejemplo, que se quiere sumar a la  $i$ -ésima fila la  $j$ -ésima multiplicada por  $\alpha \in \mathfrak{R}$ . Y supongamos, por fijar ideas, que  $i < j$ . Si escribimos  $A$  en función de sus filas

$$A = \begin{bmatrix} a^1 \\ a^2 \\ \vdots \\ a^m \end{bmatrix}$$

queremos obtener la matriz

$$B = \begin{bmatrix} a^1 \\ \vdots \\ a^i + \alpha a^j \\ \vdots \\ a^j \\ \vdots \\ a^m \end{bmatrix}$$

Puesto que cada fila de  $B$  es una combinación lineal de las filas de  $A$ , debe existir una matriz  $E_1(\alpha)$  tal que  $B = E_1(\alpha)A$ . Ahora bien, todas las filas de  $A$  y  $B$  coinciden, excepto la  $i$ -ésima que es  $a^i + \alpha a^j$ . Por lo tanto

$$E_1(\alpha) = \begin{matrix} & & i & & j & & \\ i & \begin{bmatrix} 1 & & & & \\ & \ddots & & & \\ & & 1 & \cdots & \alpha \\ & & & \ddots & \vdots \\ & & & & 1 \\ & & & & & \ddots \\ & & & & & & 1 \end{bmatrix} & & \end{matrix}$$

Es importante observar que  $E_1(\alpha) = I_n + \alpha e_j^T$  donde  $e_j$  es el  $j$ -ésimo vector (columna) canónico (todos los elementos cero excepto el de la posición  $j$  que es 1) y  $a$  es el vector (columna) cuyos elementos son todos cero excepto el de la posición  $i$  que es  $\alpha$ . Así pues  $E_1(\alpha)$  es una modificación de rango 1 de la matriz identidad.

Es claro que  $B = E_1(\alpha)A$ . De la misma forma

$$AE_1(\alpha) = \begin{bmatrix} a_1 & \cdots & a_i & \cdots & a_j + \alpha a_i & \cdots & a_n \end{bmatrix}.$$

Nótese que en este caso, la columna que cambia es la  $j$ -ésima.

Si lo que se quiere es multiplicar la  $i$ -ésima fila de  $A$  por  $\alpha \neq 0$  entonces la

matriz  $E_2(\alpha)$  que lo produce es

$$E_2(\alpha) = i \begin{bmatrix} & & & & & & i \\ & & & & & & \\ & & 1 & & & & \\ & & \ddots & \ddots & & & \\ & & & 1 & & & \\ & & & & \alpha & & \\ & & & & & 1 & \\ & & & & & & \ddots & \ddots \\ & & & & & & & 1 \end{bmatrix}$$

En efecto

$$E_2(\alpha)A = \begin{bmatrix} a^1 \\ \vdots \\ \alpha a^i \\ \vdots \\ a^m \end{bmatrix} \quad \text{y} \quad AE_2(\alpha) = \begin{bmatrix} a_1 & \cdots & \alpha a_i & \cdots & a_n \end{bmatrix}.$$

Finalmente, si lo que se quiere es permutar las filas o columnas  $i$  y  $j$  debemos utilizar una matriz de permutación, en realidad una transposición:

$$P = \begin{bmatrix} & & & & i & & j & & \\ & & & & & & & & \\ & 1 & & & & & & & \\ & & \ddots & & & & & & \\ & & & 1 & & & & & \\ i & & & & 0 & \dots & 1 & & \\ & & & & \vdots & \ddots & \vdots & & \\ j & & & & 1 & \dots & 0 & & \\ & & & & & & & 1 & \\ & & & & & & & & \ddots & \\ & & & & & & & & & 1 \end{bmatrix}$$

Así

$$PA = \begin{bmatrix} a^1 \\ \vdots \\ a^j \\ \vdots \\ a^i \\ \vdots \\ a^m \end{bmatrix} \quad \text{y} \quad AP = \begin{bmatrix} a_1 & \cdots & a_j & \cdots & a_i & \cdots & a_n \end{bmatrix}.$$

Las matrices de la forma  $E_1(\alpha)$ ,  $E_2(\alpha)$  y  $P$  se llaman **matrices elementales**. Debe notarse que estas matrices y las matrices de transposición se obtienen de la matriz identidad realizando sobre ésta la misma transformación elemental que se quiere conseguir sobre la matriz que van a actuar. Así, la matriz  $E_1(\alpha)$  de más arriba se obtiene de la matriz  $I_n$  sumando a la fila  $i$  la  $j$ -ésima multiplicada por  $\alpha$ . Y también se obtiene sumando a la columna  $j$  la  $i$ -ésima multiplicada por  $\alpha$ .

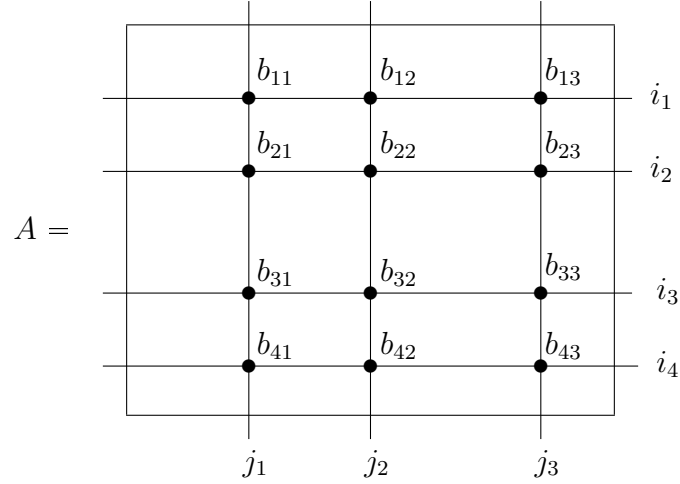
#### 1.1.4. Submatrices

Dada una matriz  $A = [a_{ij}] \in \mathfrak{R}^{m \times n}$ , diremos que la matriz  $B = [b_{ij}] \in \mathfrak{R}^{p \times q}$  es una submatriz de  $A$  si existen índices  $(i_1, \dots, i_p)$  y  $(j_1, \dots, j_q)$  tales que  $1 \leq i_1 < \dots < i_p \leq m$ ,  $1 \leq j_1 < \dots < j_q \leq n$  y

$$a_{i_r j_s} = b_{rs}, \quad 1 \leq r \leq p, 1 \leq s \leq q.$$

Esto significa, simplemente, que los elementos que componen  $B$  están en la intersección de algunas filas y columnas de  $A$ :





Trabajaremos a menudo con submatrices para lo que vamos a introducir una notación más compacta y útil. Denotamos con

$$Q_{r,n} = \{(i_1, \dots, i_r) : 1 \leq i_1 < \dots < i_r \leq n\}$$

el conjunto de todas las secuencias de  $r$  números enteros positivos entre 1 y  $n$ .

Si  $\alpha, \beta \in Q_{r,n}$ , diremos que  $\alpha$  precede a  $\beta$  en el *orden lexicográfico*, y escribiremos

$$\alpha \leq \beta,$$

si la primera diferencia  $\beta_i - \alpha_i$  que no sea cero es positiva. Por ejemplo, si  $\alpha = (1, 3, 4, 7)$  y  $\beta = (1, 3, 5, 6)$  entonces  $\alpha - \beta = (0, 0, -1, 1)$  y  $\beta - \alpha = (0, 0, 1, -1)$ . por lo tanto  $\alpha$  precede a  $\beta$  en el orden lexicográfico:  $\beta \geq \alpha$ .

Dada una matriz  $A = [a_{ij}] \in \mathfrak{R}^{m \times n}$ , si  $\alpha \in Q_{r,m}$  y  $\beta \in Q_{s,n}$ , la matriz  $B = [b_{ij}] \in \mathfrak{R}^{r \times s}$  definida por

$$b_{ij} = a_{\alpha_i \beta_j}, \quad 1 \leq i \leq r, \quad 1 \leq j \leq s$$

es una submatriz de  $A$  y se denotar por

$$B = A[\alpha|\beta].$$

Esta notación es muy parecida a la que usa MATLAB para definir submatrices (veáse la Gua de MATLAB para este curso). MATLAB utiliza paréntesis en vez de

corchetes. Aquí los paréntesis se usarán para designar la submatriz complementaria. Es decir, si  $\alpha \in Q_{r,n}$  entonces la secuencia complementaria de  $\alpha$  es  $\alpha' \in Q_{n-r,n}$  que cumple la siguiente condición: Si  $\alpha = (\alpha_1, \dots, \alpha_r)$  y  $\alpha' = (\alpha'_1, \dots, \alpha'_{n-r})$  entonces

$$\{\alpha_1, \dots, \alpha_r, \alpha'_1, \dots, \alpha'_{n-r}\} = \{1, 2, \dots, n\}.$$

Si  $\alpha \in Q_{r,m}$  y  $\beta \in Q_{s,n}$  y  $\alpha', \beta'$  son las secuencias complementarias en  $\{1, \dots, m\}$  y  $\{1, \dots, n\}$ , entonces pondremos

$$A(\alpha|\beta) = A = [\alpha'|\beta'].$$

**Ejemplo 1.3** .- El siguiente ejemplo puede ayudar a clarificar la notación. Sea

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

$$A[1, 2|1, 3] = \begin{bmatrix} a_{11} & a_{13} \\ a_{21} & a_{23} \end{bmatrix} = \begin{bmatrix} 0 & 2 \\ -3 & -1 \end{bmatrix}$$

$$A(2, 3|3) = A[1, 4|1, 2, 4] = \begin{bmatrix} a_{11} & a_{12} & a_{14} \\ a_{41} & a_{42} & a_{44} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 3 \\ -7 & -6 & -4 \end{bmatrix}$$

## 1.2. Imagen y Núcleo de una matriz

Aunque en la sección anterior hemos procurado actuar con la máxima generalidad acerca del anillo en el que están definidas las matrices, a partir de ahora, nuestras matrices supondremos que tienen sus elementos en un cuerpo (siempre conmutativo),  $\mathbb{F}$ , y casi siempre podremos pensar que este cuerpo es el de los números reales o complejos. Comenzamos con dos conceptos básicos bien conocidos: la imagen y el núcleo de una matriz se definen como la imagen y núcleo de la correspondiente aplicación lineal asociada.

**Definición 1.4** .- Sea  $A \in \mathbb{F}^{m \times n}$ . Se define

$$\text{Im}(A) = \{y \in \mathbb{F}^m | y = Ax \text{ para alg } x \in \mathbb{F}^n\}.$$

$y$

$$\text{Ker}(A) = \{x \in \mathbb{F}^n | Ax = 0\}.$$

En estos dos conceptos aparece un sistema de ecuaciones lineales. Podemos utilizar la interpretación que hemos hecho en la Sección 1.1.3 acerca del significado de multiplicar matrices para conseguir una nueva caracterización de estos conjuntos. Recordemos que si  $a_1, \dots, a_n$  son vectores de  $\mathbb{F}^n$ , el subespacio de  $\mathbb{F}^n$  generado por estos vectores, y que denotaremos indistintamente por  $\langle a_1, \dots, a_n \rangle$  o por  $\text{Span}(a_1, \dots, a_n)$ , es el conjunto de todas las combinaciones lineales de  $a_1, \dots, a_n$ :

$$\langle a_1, \dots, a_n \rangle = \text{Span}(a_1, \dots, a_n) = \left\{ \sum_{i=1}^n x_i a_i \mid x_i \in \mathbb{F} \right\}.$$

**Proposición 1.5** .-  $\text{Im}(A)$  es subespacio de  $\mathbb{F}^m$  generado por las columnas de  $A$ .

**Demostración.**- Tenemos que demostrar que si  $A = [a_1 \ a_2 \ \dots \ a_n]$  entonces  $\text{Im } A = \langle a_1, a_2, \dots, a_n \rangle$ . Ahora bien

$$y \in \text{Im } A \Leftrightarrow y = Ax, \text{ para alg } x \in \mathbb{F}^n \Leftrightarrow y = \sum_{i=1}^n x_i a_i,$$

donde  $x_1, \dots, x_n$  son las componentes de  $x$ . La última equivalencia es, como ya hemos mencionado, debida a que  $y = Ax$  si y sólo si  $y$  es una combinación lineal de las columnas de  $A$  cuyos coeficientes son las componentes de  $x$ . ■

De forma similar, los vectores  $x \in \text{Ker } A$  son aquellos cuyas componentes son los coeficientes del vector 0 como combinación lineal de las columnas de  $A$ :  $0 = x_1 a_1 + x_2 a_2 + \dots + x_n a_n$ .

### 1.2.1. Rango y nulidad

El *rango por columnas* de una matriz es la dimensión del espacio que generan sus columnas. Y el *rango por filas* de una matriz es la dimensión del espacio que generan sus filas. Ambos números coinciden para cada matriz (hay varias pruebas posibles de este hecho, una de ellas se puede obtener como corolario del Teorema SVD que

discutiremos en Lecciones posteriores). Nos referiremos a este número común como el *rango* de la matriz y, para una matriz  $A$ , lo representaremos por  $\text{rang}(A)$ .

Por otra parte, la *nulidad* de  $A \in \mathbb{F}^{m \times n}$  se define como la dimensión de  $\text{Ker } A$  y la denotaremos con el símbolo  $\nu(A)$ . Los teoremas de isomorfa nos dan una relación estrecha entre nulidad y rango. En efecto, de acuerdo con la Proposición 1.5,  $\text{rang}(A) = \dim \text{Im } A$  y de acuerdo con los teoremas de isomorfa de aplicaciones lineales  $\dim \text{Im } A = n - \dim \text{Ker } A$ . As pues

$$\text{rang}(A) = n - \nu(A).$$

Diremos que una matriz  $m \times n$  tiene *rango completo* si su rango es el máximo posible; es decir, el más pequeño de los números  $m$  y  $n$ . As pues, si  $m \geq n$  y  $A \in \mathbb{F}^{m \times n}$  tiene rango completo, se tiene que las  $n$  columnas de  $A$  son linealmente independientes. Esto a su vez significa que el único vector  $x = (x_1, \dots, x_n)$  para el que  $x_1 a_1 + \dots + x_n a_n = 0$  es  $x = 0$ ; es decir,  $\text{Ker } A = \{0\}$ . Y esto equivale a que la aplicación lineal que define  $A$  es inyectiva. De la misma forma, si  $m \leq n$  y  $A$  tiene rango completo entonces  $\dim \text{Im } A = m$  y como  $\text{Im } A \subseteq \mathbb{F}^m$  debe ser  $\text{Im } A = \mathbb{F}^m$ ; i.e.,  $A$  es suprayectiva. Finalmente, si  $m = n$  y  $A$  tiene rango completo entonces la aplicación lineal que define es biyectiva. Es decir,  $A$  es invertible: hay una nica matriz  $B \in \mathbb{F}^{n \times n}$  tal que  $AB = BA = I_n$ . A esta nica matriz se le llama *inversa* de  $A$  y se representa por  $A^{-1}$ .

El siguiente Teorema recoge unas cuantas propiedades bien conocidas equivalentes a que una matriz sea invertible. Algunas de ellas ya han sido demostradas más arriba.

**Teorema 1.6** *Para  $A \in \mathbb{F}^{n \times n}$  las siguientes condiciones son equivalentes:*

1.  $A$  tiene inversa  $A^{-1}$ .
2.  $\text{rang}(A) = n$ .
3.  $\text{Im } A = \mathbb{F}^n$ .
4.  $\nu(A) = 0$ .
5.  $\text{Ker } A = \{0\}$ .

6. 0 no es un valor propio (autovalor) de  $A$ .

7.  $\det(A) \neq 0$ .

Recordemos que  $\lambda$ , escalar en, quizá, un cuerpo extensión de  $\mathbb{F}$ , es un valor propio si hay un vector no nulo  $x$  tal que  $Ax = \lambda x$ . Si 0 fuera valor propio de  $A$  entonces  $Ax = 0$  para algún  $x \neq 0$ . As  $\text{Ker } A \neq \{0\}$ . Y recíprocamente.

En cuanto a la propiedad  $\det(A) \neq 0$ , se trata de una consecuencia inmediata de las propiedades básicas de los determinantes para matrices sobre cuerpos. Y en cuanto a terminología, las matrices que cumplen la propiedad  $\det(A) \neq 0$  se llaman *no singulares* o *regulares*. Por lo tanto, no singularidad e invertibilidad son conceptos equivalentes para matrices sobre cuerpos.

Asociada al determinante recordemos que hay una caracterización del rango de una matriz que la mencionamos por completitud porque tanto ella como el propio concepto de determinante no son de utilidad desde un punto de vista numérico: el rango de  $A$  es el tamaño de la mayor submatriz con determinante no nulo de  $A$ . A los determinantes de las submatrices de una matriz también se les llama *menores* de la matriz.

Una observación final en la línea de lo expuesto en la Sección 1.1.3. Cuando escribimos

$$x = A^{-1}b$$

no debemos pensar en el vector  $x$  como el resultado de multiplicar la inversa de la matriz  $A$  por el vector  $b$ , sino que  $A^{-1}b$  es el vector cuyas componentes son los coeficientes de  $b$  escrito como combinación lineal de las columnas de  $A$ . En efecto, esto es consecuencia de la siguiente obvia equivalencia para las matrices invertibles:

$$Ax = b \Leftrightarrow x = A^{-1}b$$

O también, si  $A$  es invertible entonces  $\text{Im } A = \mathbb{F}^n$  y las columnas de  $A$  forman una base de  $\mathbb{F}^n$ . Las componentes del vector  $A^{-1}b$  son las coordenadas de  $b$  en esta base.

## 1.3. Factorización LU

La factorización  $LU$  se estudia en el curso de Métodos Numéricos bajo la denominación de Algoritmos de Doolittle y Crout. La repasamos ahora brevemente.

El objetivo de la factorización  $LU$  es la resolución de sistemas de ecuaciones lineales; y es la forma actual de representar la eliminación Gaussiana. La eliminación Gaussiana era conocida ya en la China antigua [15] y los estudiantes la emplean de manera intuitiva en la escolarización obligatoria o bachillerato como un método de eliminación de variables al resolver sistemas lineales. Gauss la usó en el contexto de problemas de mínimos cuadrados lineales [7],[10], [20]. En los cursos de álgebra lineal se suele enseñar en conjunción con la reducción de una matriz a *forma escalera* y, en este contexto, se trata como una herramienta para obtener todas las soluciones de un sistema lineal, para calcular el determinante y el rango de la matriz de los coeficientes. En esta sección estudiaremos su conexión con la factorización  $LU$  y justificaremos por qué ésta es la forma eficiente de implementar la eliminación Gaussiana.

Supongamos, en una primera aproximación, que se nos pide resolver el siguiente sistema de ecuaciones lineales:

$$\begin{bmatrix} 10 & -20 & 0 \\ -2 & 4 & 6 \\ 5 & -2 & 8 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 6 \\ 4 \\ 11 \end{bmatrix}$$

y pongamos

$$A = \begin{bmatrix} 10 & -20 & 0 \\ -2 & 4 & 6 \\ 5 & -2 & 8 \end{bmatrix}, x = \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}, b = \begin{bmatrix} 6 \\ 4 \\ 11 \end{bmatrix}$$

La eliminación Gaussiana consiste en operar sobre las filas del sistema para reducir a cero todos los elementos de la primera columna de la matriz de los coeficientes. El primer paso será sumar a la segunda fila la primera multiplicada por  $2/10$  y sumar a la tercera fila la primera multiplicada por  $-5/10$ . Se trata, en realidad, de hacer transformaciones elementales: si

$$M_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 2/10 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ y } M_{12} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -5/10 & 0 & 1 \end{bmatrix}$$

entonces

$$M_{12}M_{11}A = \begin{bmatrix} 10 & -20 & 0 \\ 0 & 0 & 6 \\ 0 & 8 & 8 \end{bmatrix} \text{ y } M_{12}M_{11}b = \begin{bmatrix} 6 \\ 5,2 \\ 8 \end{bmatrix}$$

Notemos que

$$M_1 = M_{12}M_{11} = \begin{bmatrix} 1 & 0 & 0 \\ 2/10 & 1 & 0 \\ -5/10 & 0 & 1 \end{bmatrix};$$

es decir, podemos condensar todas las operaciones elementales hechas en la primera columna en una única matriz que difiere de la identidad en que los elementos en la primera columna son los números por los que hay que multiplicar el elemento en la posición  $(1, 1)$  para anular los restantes elementos de la primera columna. En definitiva,

$$M_1 = I_3 + m_1 e_1^T, \quad m_1 = \begin{bmatrix} 0 \\ 2/10 \\ -5/10 \end{bmatrix}.$$

El siguiente paso es proceder con los elementos de la segunda columna que están por debajo de la diagonal de la misma forma. En nuestro ejemplo, sin embargo, tenemos un problema: el elemento en la posición  $(2, 2)$  es 0; no podemos usarlo para eliminar el elemento en la posición  $(3, 2)$ . El remedio es fácil: permutamos las filas 2 y 3. Para ello definimos

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

Esta es una matriz de permutación que al premultiplicar cualquier matriz por ella, permuta en aquella las filas 2 y 3. Así

$$PM_1 A = \begin{bmatrix} 10 & -20 & 0 \\ 0 & 8 & 8 \\ 0 & 0 & 6 \end{bmatrix} \quad \text{y} \quad PM_1 b = \begin{bmatrix} 6 \\ 8 \\ 5,2 \end{bmatrix}$$

La estrategia de permutar filas se debe usar, por motivos que por ahora no se pueden explicar con brevedad, no sólo cuando el elemento de la posición diagonal es cero, sino para colocar en dicha posición el elemento que, en valor absoluto, sea el mayor de la columna por debajo de la diagonal. Por ejemplo, si sustituímos el 4 en la posición  $(2, 2)$  de la matriz original  $A$  por 4,1 entonces

$$M_1 A = \begin{bmatrix} 10 & -20 & 0 \\ 0 & 0,1 & 6 \\ 0 & 8 & 8 \end{bmatrix} \quad \text{y} \quad PM_1 A = \begin{bmatrix} 10 & -20 & 0 \\ 0 & 8 & 8 \\ 0 & 0,1 & 6 \end{bmatrix}.$$

Siguiendo con esta matriz, sumamos a su tercera fila la segunda multiplicada por  $-1/80$ ; i.e, premultiplicamos por

$$M_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & -1/80 & 0 \end{bmatrix} = I_3 + m_2 e_2^T, \quad m_2 = \begin{bmatrix} 0 \\ 0 \\ -1/80 \end{bmatrix},$$

para obtener

$$M_2 P M_1 A = \begin{bmatrix} 10 & -20 & 0 \\ 0 & 8 & 8 \\ 0 & 0 & 5,9 \end{bmatrix} \text{ y } M_2 P M_1 b = \begin{bmatrix} 6 \\ 8 \\ 5,1 \end{bmatrix}$$

De aquí se obtiene fácilmente la solución:

$$x_3 = \frac{51}{59}, x_2 = \frac{1}{8} \left( 8 - 8 \frac{51}{59} \right) = \frac{8}{59}, x_1 = \frac{1}{10} \left( 6 + 20 \frac{8}{59} \right) = \frac{6}{10} + \frac{16}{59}.$$

Recordamos que a este proceso se le llama sustitución hacia atrás.

Observemos ahora que

$$P M_1 P = \begin{bmatrix} 1 & 0 & 0 \\ -5/10 & 1 & 0 \\ 2/10 & 0 & 1 \end{bmatrix} = M'_1$$

y que  $P^T = P$ . Es decir,  $P M_1 = M'_1 P$  siendo  $M'_1$  la matriz que se obtiene de  $M_1$  al intercambiar los elementos en las posiciones  $(2, 1)$  y  $(3, 1)$ . As pues, si  $M'_2 = M_2$

$$M'_2 M'_1 P A = U$$

siendo

$$U = \begin{bmatrix} 10 & -20 & 0 \\ 0 & 8 & 8 \\ 0 & 0 & 5,9 \end{bmatrix}$$

una matriz triangular superior. Finalmente, las inversas de  $M'_1$  y  $M'_2$  son muy fáciles de calcular: basta cambiar de signo los elementos que están fuera de la diagonal:

$$L_1 = M'^{-1}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 5/10 & 1 & 0 \\ -2/10 & 0 & 1 \end{bmatrix}, L_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 1/80 & 0 \end{bmatrix}$$

Poniendo  $L = L_1 L_2$  obtenemos

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 5/10 & 1 & 0 \\ -2/10 & 1/80 & 1 \end{bmatrix}$$



que es una matriz triangular inferior con elementos todos iguales a 1 en la diagonal y cuyos elementos por debajo de la diagonal  $i$  son los de la matriz  $L_i$ . Además, existe una matriz de permutación  $P$  tal que

$$PA = LU$$

A esta forma de escribir  $PA$  se le llama *descomposición* o *factorización LU* de  $A$  con pivoteo parcial por columnas.

Este ha sido un ejemplo de un resultado general: dada una matriz no singular  $A \in \mathbb{F}^{n \times n}$ , ( $\mathbb{F} = \mathbb{R}$  o  $\mathbb{C}$ ), realizando transformaciones elementales por filas se puede reducir  $A$  a una matriz  $U$  que es triangular superior. En concreto, existen matrices  $M_1, \dots, M_{n-1}$  con la forma

$$M_k = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & m_{k+1k} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & m_{nk} & 0 & \cdots & 1 \end{bmatrix} = I_n + m_k e_k^T, \quad m_k = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ m_{k+1k} \\ \vdots \\ m_{nk} \end{bmatrix}$$

y matrices de transposición  $P_1, \dots, P_{n-1}$  tales que

$$M_{n-1}P_{n-1} \cdot \dots \cdot M_2P_2M_1P_1A = U.$$

Explícitamente, si para  $k = 1, \dots, n-1$ ,  $A^{(k)} = P_k M_{k-1} P_{k-1} \cdot \dots \cdot P_2 M_1 P_1 A$ , entonces

$$A^{(k)} = \begin{matrix} & k & n-k \\ k & & \\ n-k & \begin{bmatrix} A_{11}^{(k)} & A_{12}^{(k)} \\ 0 & A_{22}^{(k)} \end{bmatrix} \end{matrix}$$

siendo  $A_{11}^{(k)}$  una matriz triangular y para  $i = k+1, \dots, n$  las cantidades

$$m_{ik} = \frac{a_{ik}^{(k)}}{a_{kk}^{(k)}}.$$

se llaman *multiplicadores* y a  $a_{kk}^{(k)}$  el *pivote* del paso  $k$ . Además,  $A^{(k+1)} = M_k A^{(k)}$  es una matriz con la misma forma que  $A^{(k)}$  pero con ceros en la columna  $k+1$  por

debajo de la diagonal. Los elementos de  $A_{22}^{(k+1)}$  se obtienen de los de  $A^{(k)}$  mediante las siguientes operaciones:

$$a_{ij}^{(k+1)} = a_{ij}^{(k)} - m_{ik}a_{kj}^{(k)}, \quad i, j = k+1, \dots, n.$$

Ahora, si  $L_i$  es la matriz que se obtiene cambiando los signos de los multiplicadores en la columna  $i$  de  $M_i$  e intercambiando los elementos de dicha columna correspondientes a la transposiciones  $P_{n-1}, \dots, P_{i+1}$ , entonces

$$P_{n-1} \cdot \dots \cdot P_1 A = L_1 \cdot \dots \cdot L_{n-1} U.$$

Poniendo  $P = P_{n-1} \cdot \dots \cdot P_1$  y  $L = L_1 \cdot \dots \cdot L_{n-1}$  tenemos que

$$PA = LU$$

es una descomposición  $LU$  de  $A$  con pivoteo por columnas.

Debe observarse la siguiente propiedad que es importante para realizar el algoritmo que produce la factorización  $LU$  de  $A$  con pivoteo parcial de columnas: Para  $i = 1, \dots, n-1$  denotemos

$$L_i = \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 \\ 0 & \cdots & \ell_{i+1i} & 1 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \ell_{ni} & 0 & \cdots & 1 \end{bmatrix} = I_n + \ell_i e_i^T, \quad \ell_i = i \begin{bmatrix} 0 \\ \vdots \\ 0 \\ \ell_{i+1i} \\ \vdots \\ \ell_{ni} \end{bmatrix}$$

y supongamos  $i < j$ . Entonces teniendo en cuenta que  $e_i^T \ell_j = 0$ , se concluye que

$$L_i L_j = I_n + \ell_i e_i^T + \ell_j e_j^T = \begin{matrix} & i & & j \\ i & \begin{bmatrix} 1 & \cdots & 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & 1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \ell_{i+1i} & 1 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & \ddots & \vdots & \vdots & \cdots & \vdots \\ 0 & \cdots & \ell_{ji} & 0 & \cdots & 1 & 0 & 0 & \cdots & 0 \\ 0 & \cdots & \ell_{j+1i} & 0 & \cdots & \ell_{j+1j} & 1 & 0 & \cdots & 0 \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \ell_{ni} & 0 & \cdots & \ell_{nj} & 0 & 0 & \cdots & 1 \end{bmatrix} \end{matrix}.$$

En definitiva,  $L = L_1 \cdot \dots \cdot L_{n-1} = I_n + \ell_i e_i^T + \dots + \ell_{n-1} e_{n-1}^T$ . Es decir, para  $j = 1, \dots, n-1$  la  $j$ -ésima columna de  $L$  es la de  $L_j$ . En otras palabras, los elementos de  $L$  en la columna  $j$  por debajo de la diagonal son los multiplicadores del paso  $j$ , de modo que  $L$  se obtiene asignado a la posición  $(i, j)$ ,  $i > j$ , el multiplicador  $m_{ij} = a_{ij}^{(j)} / a_{jj}^{(j)}$ .

Se presenta a continuación una función de MATLAB que teniendo en cuenta todos los comentarios anteriores devuelve las matrices  $L$ ,  $U$  y  $P$  tales que  $PA = LU$  es una factorización  $LU$  de  $A$  con pivoteo parcial por columnas. La matriz  $P$  puede no ser única (puede haber en cada paso varios elementos en la correspondiente columna con máximo valor absoluto) pero una vez fijada ésta, los factores  $L$  y  $U$  son únicos ([6, pag. 35 Vol. 1]). Se pone el símbolo % para explicar brevemente el significado de cada línea.

```
function [L,U,P] = lutxv1(A)

%LUTXV1 devuelve matrices L, triangular inferior con 1 en la diagonal,
%U, triangular superior y P e permutación tal que PA=LU. En vez
%trabajar con la matriz P se trabaja con un vector p que almacena las
%transposiciones y a partir del cual se construye P.

% Calculamos el tamaño de A
[n,n] = size(A);
%Iniciamos p=[1 2 ... n]^T
p = (1:n)';
% y L
L=eye(n);

%Transformaciones elementales en las columnas 1, 2,..., n.
for k = 1:n-1
    %Calculamos el máximo, r, (en valor absoluto) de los elementos
    % A(k,k), A(k,k+1),...,A(k,n), y la posición, m, donde se encuentra
    [r,m] = max(abs(A(k:n,k)));
    % Calculamos la posición del máximo en toda la columna k-ésima de A
    m = m+k-1;
    %Si el máximo fuera cero es porque toda la columna es cero y A no sería
    %invertible
    if (A(m,k) ~= 0)
        %Si el máximo no está en la diagonal permutamos
```

```

    if (m ~= k)
    %permutamos las filas k y m en A
        A([k m],k:n) = A([m k],k:n);
    %en L
        L([k m],1:k-1) = A([m k],1:k-1);
    % y las posiciones k y m de p
        p([k m]) = p([m k]);
    end
    %L está formada por los multiplicadores
    i = k+1:n;
    L(i,k) = A(i,k)/A(k,k);
    % Por debajo de (k,k) los elementos de A en la columna k serán 0
    % ¿para qué hacer la operación?
    A(i,k)=0;
    % Realizamos sobre las filas i=k+1,...,n la transformación elemental
    j = k+1:n;
    A(i,j) = A(i,j) - L(i,k)*A(k,j);
end
end
%A se ha convertido en triangular superior: es la U
U=A;
% Construimos P a partir de p: las filas 1:n de P son las de la permutación
P=eye(n); P=P(p,:);

```

En la segunda y última versión se juega con el hecho de que la matriz  $A$  va haciéndose triangular superior y en cada paso sus elementos por debajo de la diagonal se hacen ceros. De hecho, se van haciendo ceros a medida que se va construyendo la matriz  $L$  columna a columna con los multiplicadores. La idea es aprovechar la parte triangular inferior de  $A$  almacenando allí la parte triangular inferior de la  $L$ . De esta forma, toda referencia a  $L$  es innecesaria. La siguiente función de MATLAB es de una versión legible de la función LU de MATLAB.

```

function [L,U,P] = lutx(A)

%LUTX devuelve matrices L, triangular inferior con 1 en la diagonal,
%U, triangular superior y un vector p, tal que si P es la matriz de
%permutación que corresponde al vector p, se tiene PA=LU

% Calculamos el tamaño de A

```

```

[n,n] = size(A);
%Iniciamos p=[1 2 ... n]^T
p = (1:n)';
%Transformaciones elementales en las columnas 1, 2,..., n.
for k = 1:n-1
    %Calculamos el máximo, r, (en valor absoluto) de los elementos
    % A(k,k), A(k,k+1),...,A(k,n), y la posición, m, donde se encuentra
    [r,m] = max(abs(A(k:n,k)));
    % Calculamos la posición del máximo en toda la columna k-ésima de A
    m = m+k-1;
    %Si el máximo fuera cero es porque toda la columna es cero y A no sería
    %invertible
    if (A(m,k) ~= 0)
        %Si el máximo no está en la diagonal permutamos
        if (m ~= k)
            %permutamos las filas k y m
            A([k m],:) = A([m k],:);
            % y las posiciones k y m de p
            p([k m]) = p([m k]);
        end
        %Almacenamos en la columna k desde la posición diagonal para abajo
        %los elementos de L
        i = k+1:n;
        A(i,k) = A(i,k)/A(k,k);
        % Realizamos sobre las filas i=k+1,...,n la transformación elemental
        j = k+1:n;
        A(i,j) = A(i,j) - A(i,k)*A(k,j);
    end
end
%En la parte triangular inferior de A (sin la diagonal) está L
L = tril(A,-1) + eye(n,n);
%Y en la parte triangular superior (incluyendo la diagonal) está U
U = triu(A);
% Construimos P a partir de p: las filas 1:n de P son las de la permutación
P=eye(n); P=P(p,:);

```

Una vez se ha obtenido una descomposición  $LU$  de  $A$ , la resolución del sistema  $Ax = b$  es simple: se obtiene la solución del sistema  $Ly = b$  mediante sustitución hacia adelante; y a continuación la solución de  $Ux = y$  mediante sustitución hacia

atrás.

El algoritmo LU es el primer algoritmo que vemos en este curso. Se trata de un algoritmo sencillo pero eso no significa que sea bueno. Diseñar buenos algoritmos puede ser una tarea complicada y decidir si un algoritmo es bueno también. Hay dos propiedades fundamentales que se deben exigir a un buen algoritmo: que sea estable y sea poco costoso. Estudiaremos en capítulos posteriores lo que significa que un algoritmo es estable y analizaremos si la eliminación Gaussiana lo es.

Para un algoritmo como el que acabamos de ver, su coste se mide en función del número de operaciones que hay que realizar. A estas operaciones se les suele llamar *flops* (por *floating point operations*). Cada suma, resta, multiplicación y división es una *flop*. No distinguiremos entre operaciones con números reales o complejos y tampoco prestaremos atención a otras operaciones que realizan internamente los ordenadores (como asignación de memoria o realocación de variables, dedicación de los procesadores a las distintas tareas que se estén realizando en el momento de correr el algoritmo, etc). No obstante se debe saber que estos factores pueden afectar en gran medida al tiempo necesario para la implementación del algoritmo.

Normalmente el número de flops es un polinomio en el tamaño de la matriz. De este polinomio sólo se suele destacar el término de grado mayor. Así, se suele decir que el *coste del algoritmo LU es del orden de*  $\frac{2}{3}n^3$  y se escribe habitualmente  $\sim \frac{2}{3}n^3$  o simplemente  $O(n^3)$ . ¿Cómo se calcula este número? Observando que sólo hay operaciones aritméticas en dos líneas del algoritmo:

```
A(i,k) = A(i,k)/A(k,k);
A(i,j) = A(i,j) - A(i,k)*A(k,j)
```

La primera consiste de una división mientras que la segunda de una resta y una multiplicación. Ésta domina a aquella y teniendo en cuenta que la longitud del vector  $j$  es  $l = n - k + 1$  resulta que  $2l$  es el número de operaciones necesario para conseguir los elementos de la fila  $i$ . Y esto hay que repetirlo para  $i = k + 1 : n$  y para  $k = 1 : n - 1$ . Así pues, el número de operaciones está dominado por la expresión:

$$\sum_{k=1}^{n-1} \sum_{i=k+1}^n \sum_{j=k+1}^n 2 = \sum_{k=1}^{n-1} 2(n-k)^2 = \sum_{k=1}^{n-1} 2k^2 = 2 \frac{n(n-1)(2n-1)}{6} \sim \frac{2}{3}n^3$$

tal y como se ha dicho más arriba. Se dice entonces que el coste del algoritmo LU es de orden cúbico.