

Apuntes de álgebra lineal

EDUARDO LIZ MARZÁN

Junio de 2011.

Índice General

1	Introducción: estructura de cuerpo y números complejos	5
1.1	Operaciones internas y estructura de cuerpo.	5
1.2	Números complejos.	6
2	Matrices y determinantes	9
2.1	Introducción.	9
2.2	Definición y tipos de matrices.	9
2.3	Operaciones con matrices.	10
2.4	Propiedades de la trasposición de matrices.	14
2.5	Traza de una matriz.	15
2.6	Matrices elementales.	15
2.7	Forma escalonada y rango de una matriz.	17
2.8	Cálculo de la inversa.	20
2.9	Determinantes.	21
3	Sistemas de ecuaciones lineales	23
3.1	Introducción.	23
3.2	Expresión matricial.	23
3.3	Existencia de soluciones.	24
3.4	Conjuntos de soluciones.	25
3.5	Matrices cuadradas y uso de la factorización LU	27
4	Espacios vectoriales y aplicaciones lineales	33
4.1	Introducción.	33
4.2	Espacios y subespacios vectoriales.	33
4.3	Independencia lineal.	36
4.4	Bases y dimensión.	36
4.5	Cambio de base en \mathbb{R}^n	38
4.6	Definición de aplicación lineal y propiedades.	40
4.7	Núcleo e imagen de una aplicación lineal.	41
5	Diagonalización y funciones de matrices	43
5.1	Introducción.	43
5.2	Autovalores y autovectores.	43

5.3	Matrices diagonalizables.	47
5.4	Teorema de Cayley-Hamilton.	48
5.5	Funciones de matrices.	50
6	Espacios euclídeos	53
6.1	Introducción.	53
6.2	Espacios vectoriales con producto escalar.	53
6.3	Ortogonalidad	55
6.4	Proyección ortogonal.	56
6.5	Diagonalización ortogonal.	58
6.6	Formas cuadráticas sobre \mathbb{R}^n . Clasificación.	61
6.7	Mínimos cuadrados. Ajuste.	66
6.8	Descomposición en valores singulares.	69
	Referencias	75

Capítulo 1

Introducción: estructura de cuerpo y números complejos

1.1 Operaciones internas y estructura de cuerpo.

Una operación interna $*$ en un conjunto A es una correspondencia que asigna a cada par de elementos $a, b \in A$ un elemento $c = a * b \in A$.

Consideraremos dos tipos de operaciones internas, que denotaremos por suma (+) y producto (\cdot). Si A es un conjunto con una o dos operaciones internas, A puede tener distintas estructuras según las propiedades que verifiquen estas operaciones. Consideraremos las siguientes propiedades:

1. Propiedad asociativa: $(a * b) * c = a * (b * c)$, $\forall a, b, c \in A$. Esta propiedad permite operar más de dos elementos. En este caso escribiremos simplemente $a * b * c$.
2. Elemento neutro: Se dice que $(A, *)$ tiene elemento neutro si existe $e \in A$ tal que $a * e = e * a = a$, $\forall a \in A$. En la suma, el elemento neutro se llama cero (0) y en el producto se llama uno (1). El elemento neutro, si existe, es único.
3. Elemento simétrico: Se dice que $a \in A$ tiene elemento simétrico si existe $a' \in A$ tal que $a * a' = a' * a = e$. En el caso de la suma, el elemento simétrico se llama elemento opuesto y se denota por $-a$ ($a + (-a) = (-a) + a = 0$). En el caso del producto, se llama inverso y se denota por a^{-1} ($a \cdot a^{-1} = a^{-1} \cdot a = 1$).
4. Propiedad conmutativa: $a * b = b * a$, $\forall a, b \in A$. Si una operación producto verifica la propiedad conmutativa entonces el elemento inverso se denota por $1/a$.
5. Propiedad distributiva. Si A tiene definida una suma y un producto, se dice que el producto es distributivo con respecto a la suma si

$$\begin{aligned}a \cdot (b + c) &= a \cdot b + a \cdot c \\(a + b) \cdot c &= a \cdot c + b \cdot c,\end{aligned}$$

para todo $a, b, c \in A$.

Se dice que un conjunto con una operación interna $(A, *)$ es un **grupo conmutativo** si verifica las propiedades asociativa y conmutativa, tiene elemento neutro y todo elemento tiene simétrico. Dos ejemplos de grupos conmutativos son $(\mathbb{R}, +)$, $(\mathbb{C}, +)$, $(\mathbb{R} \setminus \{0\}, \cdot)$ y $(\mathbb{C} \setminus \{0\}, \cdot)$.

Observación. Si B es un subconjunto de A , se denota $A \setminus B = \{x \in A / x \notin B\}$. En particular, si $a \in A$, $A \setminus \{a\} = \{x \in A / x \neq a\}$.

Se dice que un conjunto con dos operaciones internas $(A, +, \cdot)$ es un **cuerpo conmutativo** si $(A, +)$ y $(A \setminus \{0\}, \cdot)$ son grupos conmutativos y se verifica la propiedad distributiva del producto respecto a la suma. Los conjuntos de números reales y números complejos $(\mathbb{R}, +, \cdot)$, $(\mathbb{C}, +, \cdot)$ son cuerpos conmutativos.

1.2 Números complejos.

Un **número complejo** es un par de números reales $z = (a, b)$. El número real a se llama parte real de z y b se llama parte imaginaria.

Si denotamos $1 = (1, 0)$, $i = (0, 1)$, se escribe $z = (a, b) = a(1, 0) + b(0, 1) = a + bi$ (Forma binómica). El número complejo $i = (0, 1)$ se llama unidad imaginaria. Así, denotaremos el conjunto de los números complejos como $\mathbb{C} = \{a + bi : a, b \in \mathbb{R}\}$.

Los números complejos se representan en un plano bidimensional. El eje horizontal se llama “eje real” y el eje vertical se llama “eje imaginario”.

Operaciones en \mathbb{C}

- **Suma.** Sean $z_1 = a_1 + b_1i$, $z_2 = a_2 + b_2i$ dos números complejos. Se define $z_1 + z_2 = (a_1 + a_2) + (b_1 + b_2)i$.
- **Producto.** El producto de números complejos se realiza en forma binómica, teniendo en cuenta que $i^2 = -1$, es decir, $(a_1 + b_1i)(a_2 + b_2i) = (a_1a_2 - b_1b_2) + (a_1b_2 + b_1a_2)i$.

Con estas dos operaciones, $(\mathbb{C}, +, \cdot)$ tiene estructura de cuerpo conmutativo: El elemento neutro de la suma es $0 = 0 + 0i$, y el elemento opuesto de $z = a + bi$ es $-z = -a - bi$.

El elemento neutro del producto es $1 = 1 + 0i$. Todo elemento distinto de cero tiene inverso para el producto. Para definir el inverso se suele usar el conjugado, que se define del siguiente modo: si $z = a + bi \in \mathbb{C}$, se define su conjugado como $\bar{z} = a - bi$. Obsérvese que $z\bar{z} = a^2 + b^2$ y por tanto

$$z^{-1} = \frac{1}{z} = \frac{a - bi}{a^2 + b^2},$$

que está bien definido para $z \neq 0$.

Módulo y argumento

Sea $z = a + bi \in \mathbb{C}$. Se define el módulo de z como el número real $|z| = \sqrt{a^2 + b^2}$. Obsérvese que $|z| \geq 0$, $\forall z \in \mathbb{C}$ y $|z| = 0 \Leftrightarrow z = 0$. Además, $z\bar{z} = |z|^2$, y $z^{-1} = \bar{z}/|z|^2$, $\forall z \neq 0$.

El módulo de z representa su distancia al origen en el plano complejo. Se define el argumento de $z = a + bi$ como el ángulo $\alpha \in (-\pi, \pi]$ que verifica $|z|\cos(\alpha) = a$ y $|z|\sin(\alpha) = b$. De este

modo, $z = |z|(\cos(\alpha) + \operatorname{sen}(\alpha)i)$, que es la llamada forma trigonométrica de z . El argumento representa el ángulo que forma el vector (a, b) en el plano complejo con el eje real.

Utilizando las fórmulas trigonométricas para el seno y el coseno de la suma, se obtiene que si $z_1 = |z_1|(\cos(\alpha_1) + \operatorname{sen}(\alpha_1)i)$ y $z_2 = |z_2|(\cos(\alpha_2) + \operatorname{sen}(\alpha_2)i)$ son dos números complejos entonces

$$z_1 z_2 = |z_1| |z_2| (\cos(\alpha_1 + \alpha_2) + \operatorname{sen}(\alpha_1 + \alpha_2)i),$$

es decir el módulo del producto es el producto de los módulos y el argumento del producto es la suma de los argumentos. De este modo, se obtiene inmediatamente que si $z = |z|(\cos(\alpha) + \operatorname{sen}(\alpha)i)$ entonces $z^n = |z|^n (\cos(n\alpha) + \operatorname{sen}(n\alpha)i)$, $\forall n \in \mathbb{N}$.

Forma exponencial

Si $b \in \mathbb{R}$, se define $e^{bi} = \cos(b) + \operatorname{sen}(b)i$. De este modo, se extiende la función exponencial real a \mathbb{C} manteniendo sus propiedades principales; en particular, si $z = a + bi$ entonces $e^z = e^{a+bi} = e^a e^{bi} = e^a (\cos(b) + \operatorname{sen}(b)i)$.

Teniendo en cuenta esto, si $z = |z|(\cos(\alpha) + \operatorname{sen}(\alpha)i)$, también se puede representar en la forma $z = |z|e^{i\alpha}$, que se llama forma exponencial de z .

Ejemplo: $1 + i = \sqrt{2}(\cos(\pi/4) + i \operatorname{sen}(\pi/4)) = \sqrt{2} e^{\frac{\pi}{4}i}$.

Capítulo 2

Matrices y determinantes

2.1 Introducción.

En este capítulo se introducen los conceptos básicos de la teoría de matrices, con especial atención a las operaciones elementales, que serán de mucha utilidad a lo largo del curso. Sus primeras aplicaciones (incluidas en este tema) son el cálculo del rango, la matriz inversa y el determinante.

2.2 Definición y tipos de matrices.

Iniciaremos esta sección definiendo lo que entenderemos por una matriz.

Definición 2.1 *Se llama matriz real de p filas y n columnas a cualquier agrupación de la forma*

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{pmatrix},$$

donde $a_{ij} \in \mathbb{R}$ para todo $i = 1, 2, \dots, p$, $j = 1, 2, \dots, n$.

También diremos que A es una matriz de tamaño $p \times n$ o de orden $p \times n$.

Denotaremos por $\mathcal{M}_{p \times n}(\mathbb{R})$ el conjunto de todas las matrices de p filas y n columnas con elementos en \mathbb{R} . En notación reducida, escribiremos $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$.

Si $A = (a_{ij})$, $B = (b_{ij})$ son dos matrices de tamaño $p \times n$, diremos que $A = B$ si $a_{ij} = b_{ij}$ para todo $i = 1, 2, \dots, p$, $j = 1, 2, \dots, n$.

Son especialmente importantes las matrices cuadradas, que se caracterizan por tener el mismo número de filas que de columnas.

Las matrices cuadradas más simples son las diagonales. Una matriz cuadrada $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es **diagonal** si $a_{ij} = 0$ para todo $i \neq j$, es decir,

$$A = \begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}.$$

Observación: Se llama diagonal de una matriz $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$ al vector de \mathbb{R}^m $\text{diag}(A) = (a_{11}, a_{22}, \dots, a_{mm})$, donde $m = \min\{p, n\}$.

También serán importantes las matrices triangulares.

- Una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es **triangular superior** si $a_{ij} = 0$ para todo $i > j$, es decir, si los elementos que están por debajo de la diagonal son todos cero. Por ejemplo,

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 3 & 4 \\ 0 & 0 & 2 \end{pmatrix}.$$

- Una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ es **triangular inferior** si $a_{ij} = 0$ para todo $i < j$, es decir, si los elementos que están por encima de la diagonal son todos cero.

Sea $A \in \mathcal{M}_{p \times n}(\mathbb{R})$. Se define su **traspuesta** y se denota A^t como la matriz cuyas columnas son las filas de A , es decir, si $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$, entonces $A^t = (b_{ij}) \in \mathcal{M}_{n \times p}(\mathbb{R})$, con $b_{ij} = a_{ji}$ para todo $i = 1, \dots, n$, $j = 1, \dots, p$.

En general, cuando hagamos operaciones con matrices que incluyan vectores, éstos se representarán en forma de columna. Si $v \in \mathbb{R}^n$ es un vector columna, el correspondiente vector fila es v^t :

$$v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathcal{M}_{n \times 1}(\mathbb{R}) \implies v^t = (v_1, v_2, \dots, v_n) \in \mathcal{M}_{1 \times n}(\mathbb{R}).$$

2.3 Operaciones con matrices.

Suma de matrices.

La suma es una operación interna en $\mathcal{M}_{p \times n}(\mathbb{R})$. Dadas dos matrices $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$, $B = (b_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$, se define su suma como la matriz $A + B = (a_{ij} + b_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$, es decir,

$$\begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{pmatrix} + \begin{pmatrix} b_{11} & b_{12} & \cdots & b_{1n} \\ b_{21} & b_{22} & \cdots & b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ b_{p1} & b_{p2} & \cdots & b_{pn} \end{pmatrix} = \begin{pmatrix} a_{11} + b_{11} & a_{12} + b_{12} & \cdots & a_{1n} + b_{1n} \\ a_{21} + b_{21} & a_{22} + b_{22} & \cdots & a_{2n} + b_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} + b_{p1} & a_{p2} + b_{p2} & \cdots & a_{pn} + b_{pn} \end{pmatrix}.$$

Es fácil comprobar que $(\mathcal{M}_{p \times n}(\mathbb{R}), +)$ tiene estructura de grupo conmutativo. El elemento neutro es la matriz nula

$$0 = \begin{pmatrix} 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & 0 \end{pmatrix} \in \mathcal{M}_{p \times n}(\mathbb{R}).$$

Producto de una matriz por un escalar.

Dada una matriz $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$ y un escalar $\lambda \in \mathbb{R}$, se define $\lambda A = \lambda(a_{ij}) = (\lambda a_{ij})$, es decir,

$$\lambda \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{pmatrix} = \begin{pmatrix} \lambda a_{11} & \lambda a_{12} & \cdots & \lambda a_{1n} \\ \lambda a_{21} & \lambda a_{22} & \cdots & \lambda a_{2n} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda a_{p1} & \lambda a_{p2} & \cdots & \lambda a_{pn} \end{pmatrix}.$$

Es fácil verificar las siguientes propiedades:

1. $\lambda(A + B) = \lambda A + \lambda B, \forall A, B \in \mathcal{M}_{p \times n}(\mathbb{R}), \forall \lambda \in \mathbb{R}$.
2. $(\lambda + \mu)A = \lambda A + \mu A, \forall A \in \mathcal{M}_{p \times n}(\mathbb{R}), \forall \lambda, \mu \in \mathbb{R}$.
3. $(\lambda\mu)A = \lambda(\mu A), \forall A \in \mathcal{M}_{p \times n}(\mathbb{R}), \forall \lambda, \mu \in \mathbb{R}$.

Producto de matrices.

Dadas dos matrices $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R}), B = (b_{ij}) \in \mathcal{M}_{n \times q}(\mathbb{R})$, se define su producto como la matriz $AB = (c_{ij}) \in \mathcal{M}_{p \times q}(\mathbb{R})$ dada por:

$$c_{ij} = \sum_{k=1}^n a_{ik}b_{kj} = a_{i1}b_{1j} + a_{i2}b_{2j} + \cdots + a_{in}b_{nj}, \forall i = 1, 2, \dots, p, \forall j = 1, 2, \dots, q.$$

Obsérvese que para poder realizar el producto AB es necesario que el número de columnas de A coincida con el número de filas de B . Un caso especialmente interesante se presenta cuando ambas matrices son vectores de \mathbb{R}^n . Sean

$$u = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} \in \mathcal{M}_{n \times 1}(\mathbb{R}) \quad ; \quad v = \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} \in \mathcal{M}_{n \times 1}(\mathbb{R}).$$

Entonces:

$$u^t v = (u_1, u_2, \dots, u_n) \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{pmatrix} = u_1 v_1 + u_2 v_2 + \cdots + u_n v_n \in \mathbb{R}$$

representa el producto escalar, mientras que

$$u v^t = \begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{pmatrix} (v_1, v_2, \dots, v_n) = \begin{pmatrix} u_1 v_1 & u_1 v_2 & \cdots & u_1 v_n \\ u_2 v_1 & u_2 v_2 & \cdots & u_2 v_n \\ \vdots & \vdots & \ddots & \vdots \\ u_n v_1 & u_n v_2 & \cdots & u_n v_n \end{pmatrix} \in \mathcal{M}_{n \times n}(\mathbb{R}).$$

Propiedades:

- El producto de matrices es asociativo, es decir, si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, $B \in \mathcal{M}_{n \times q}(\mathbb{R})$ y $C \in \mathcal{M}_{q \times r}(\mathbb{R})$, se cumple que $(AB)C = A(BC)$.
- El producto de matrices verifica la propiedad distributiva respecto a la suma, es decir, si $A, B \in \mathcal{M}_{p \times n}(\mathbb{R})$, $C, D \in \mathcal{M}_{n \times q}(\mathbb{R})$ entonces $A(C+D) = AC+AD$, $(A+B)C = AC+BC$.
- El producto de matrices tiene elemento neutro, llamado matriz identidad.

$$I = \begin{pmatrix} 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 \end{pmatrix} \in \mathcal{M}_{n \times n}(\mathbb{R}).$$

Se tiene que $AI = A$, $\forall A \in \mathcal{M}_{p \times n}(\mathbb{R})$ e $IB = B$, $\forall B \in \mathcal{M}_{n \times q}(\mathbb{R})$.

- No se cumple la propiedad conmutativa, es decir, si $A, B \in \mathcal{M}_{n \times n}(\mathbb{R})$, en general $AB \neq BA$.

Ejemplo:

$$\begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 4 & 3 \end{pmatrix} \neq \begin{pmatrix} 3 & 4 \\ 1 & 2 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}.$$

- Si $A, B \in \mathcal{M}_{n \times n}(\mathbb{R})$, en general $AB = 0 \not\Rightarrow A = 0$ o $B = 0$.

Ejemplo:

$$\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Interpretación del producto con vectores fila y vectores columna.

Sea $A \in \mathcal{M}_{p \times n}(\mathbb{R})$. Si denotamos sus columnas por u_1, u_2, \dots, u_n y sus filas como $v_1^t, v_2^t, \dots, v_p^t$, entonces podemos escribir A en las dos siguientes formas:

$$A = (u_1 | u_2 | \cdots | u_n) \quad ; \quad A = \begin{pmatrix} \frac{v_1^t}{\vdots} \\ \frac{v_p^t}{\vdots} \end{pmatrix}.$$

En ocasiones se puede describir el producto de matrices de forma más conveniente usando sus vectores fila y sus vectores columna.

1. Sea $A = (u_1|u_2|\cdots|u_n) \in \mathcal{M}_{p \times n}(\mathbb{R})$. Sea

$$b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in \mathcal{M}_{n \times 1}(\mathbb{R}).$$

Entonces:

$$Ab = (u_1|u_2|\cdots|u_n) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = b_1u_1 + b_2u_2 + \cdots + b_nu_n \in \mathcal{M}_{p \times 1}(\mathbb{R}).$$

2. Sean $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $B = (u_1|u_2|\cdots|u_q) \in \mathcal{M}_{n \times q}(\mathbb{R})$. Entonces:

$$AB = A(u_1|u_2|\cdots|u_q) = (Au_1|Au_2|\cdots|Au_q) \in \mathcal{M}_{p \times q}(\mathbb{R}).$$

3. Sean $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $B \in \mathcal{M}_{n \times q}(\mathbb{R})$. Denotemos por u_1, u_2, \dots, u_n las columnas de A y por $v_1^t, v_2^t, \dots, v_n^t$ las filas de B . Entonces:

$$AB = (u_1|u_2|\cdots|u_n) \begin{pmatrix} \frac{v_1^t}{\hline} \\ \frac{v_2^t}{\hline} \\ \vdots \\ \frac{v_n^t}{\hline} \end{pmatrix} = u_1v_1^t + u_2v_2^t + \cdots + u_nv_n^t \in \mathcal{M}_{p \times q}(\mathbb{R}).$$

Matriz inversa y potencia de una matriz.

Para matrices cuadradas tiene sentido definir el concepto de matriz inversa y el de potencia de una matriz.

Definición 2.2 Una matriz cuadrada $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ se dice **invertible** si existe una matriz $B \in \mathcal{M}_{n \times n}(\mathbb{R})$ tal que $AB = BA = I$, donde I es la matriz identidad. En caso de existir, B se llama matriz inversa de A y se denota por A^{-1} . Las matrices invertibles también se llaman **regulares** o **no singulares**.

La siguiente propiedad se deduce inmediatamente de la definición:

Propiedad: Sean $A, B \in \mathcal{M}_{n \times n}(\mathbb{R})$. Si A y B son invertibles entonces AB también lo es y además $(AB)^{-1} = B^{-1}A^{-1}$.

Definición 2.3 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $k \in \mathbb{N}$. Se define A^k por inducción del siguiente modo: $A^2 = AA$, y, en general, $A^{k+1} = A^kA$, para todo $k \geq 2$, es decir A^k resulta de multiplicar A por sí misma k veces. Por convenio, $A^0 = I$, $A^1 = A$.

En general es difícil encontrar la expresión general de A^k en función de k . Sin embargo, es sencillo para matrices diagonales:

Propiedad: Si A es diagonal entonces A^k también es diagonal. Además,

$$\begin{pmatrix} a_{11} & 0 & \cdots & 0 \\ 0 & a_{22} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn} \end{pmatrix}^k = \begin{pmatrix} a_{11}^k & 0 & \cdots & 0 \\ 0 & a_{22}^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & a_{nn}^k \end{pmatrix}.$$

Definición 2.4 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Se dice que $B \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una **raíz k -ésima** de A si $B^k = A$.

Ejemplo:

La matriz $B = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$ es una raíz cuadrada de la matriz identidad $I \in \mathcal{M}_{2 \times 2}(\mathbb{R})$ ya que $B^2 = I$.

2.4 Propiedades de la trasposición de matrices.

Recordemos que si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ entonces A^t es la matriz cuyas columnas son las filas de A .

Se cumplen las siguientes propiedades:

1. $(A^t)^t = A, \forall A \in \mathcal{M}_{p \times n}(\mathbb{R})$.
2. $(A + B)^t = A^t + B^t, \forall A, B \in \mathcal{M}_{p \times n}(\mathbb{R})$.
3. $(\lambda A)^t = \lambda A^t, \forall A \in \mathcal{M}_{p \times n}(\mathbb{R}), \forall \lambda \in \mathbb{R}$.
4. $(AB)^t = B^t A^t, \forall A \in \mathcal{M}_{p \times n}(\mathbb{R}), \forall B \in \mathcal{M}_{n \times q}(\mathbb{R})$.
5. Si A es inversible entonces $(A^t)^{-1} = (A^{-1})^t$.
6. $(A^t)^k = (A^k)^t, \forall A \in \mathcal{M}_{n \times n}(\mathbb{R}), \forall k \in \mathbb{N}$.

En relación con la trasposición de matrices tenemos las siguientes matrices especiales:

Definición 2.5 Una matriz $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ es **simétrica** si $A^t = A$, es decir, si $a_{ij} = a_{ji}, \forall i, j = 1, \dots, n$.

Ejemplo:

La matriz $A = \begin{pmatrix} 0 & -1 & 1 \\ -1 & 2 & 3 \\ 1 & 3 & 1 \end{pmatrix}$ es simétrica.

Propiedades:

1. Si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ entonces $A^t A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es simétrica.
2. Si A es simétrica entonces A^k es simétrica para todo $k \in \mathbb{N}$.

Definición 2.6 Una matriz $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ es **antisimétrica** si $A^t = -A$, es decir, si $a_{ij} = -a_{ji}$, $\forall i, j = 1, \dots, n$.

Es inmediato comprobar que si una matriz $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ es antisimétrica, entonces $a_{ii} = 0$, $\forall i = 1, \dots, n$.

Ejemplo:

La matriz $A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$ es antisimétrica.

Definición 2.7 Una matriz $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ es **ortogonal** si $AA^t = A^t A = I$, es decir, si A es inversible y $A^t = A^{-1}$.

Ejemplo:

Si α es cualquier número real, la siguiente matriz es ortogonal:

$$A = \begin{pmatrix} \cos(\alpha) & -\text{sen}(\alpha) \\ \text{sen}(\alpha) & \cos(\alpha) \end{pmatrix}.$$

2.5 Traza de una matriz.

Sea $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$. Se llama **traza** de A , y se denota $\text{tr}(A)$, a la suma de sus elementos diagonales, es decir, $\text{tr}(A) = \sum_{i=1}^n a_{ii} = a_{11} + a_{22} + \dots + a_{nn}$.

Propiedades:

1. $\text{tr}(A + B) = \text{tr}(A) + \text{tr}(B)$, $\forall A, B \in \mathcal{M}_{n \times n}(\mathbb{R})$.
2. $\text{tr}(\lambda A) = \lambda \text{tr}(A)$, $\forall A \in \mathcal{M}_{n \times n}(\mathbb{R})$, $\forall \lambda \in \mathbb{R}$.
3. $\text{tr}(AB) = \text{tr}(BA)$, $\forall A \in \mathcal{M}_{p \times n}(\mathbb{R})$, $\forall B \in \mathcal{M}_{n \times p}(\mathbb{R})$.

2.6 Matrices elementales.

Definición 2.8 Sea $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$. Se llaman **operaciones elementales** sobre las filas o columnas de A a cualquiera de las siguientes transformaciones:

1. Permutar dos filas o dos columnas de A .

2. Sumar a una fila (o columna) de A un múltiplo de otra fila (o columna) de A .
3. Multiplicar una fila o columna de A por un escalar no nulo.

Definición 2.9 Una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una **matriz elemental** si se obtiene como resultado de efectuar una operación elemental sobre las filas o columnas de la matriz identidad.

Tipos de matrices elementales.

Distinguiremos seis tipos de matrices elementales según los tipos de operaciones elementales definidos arriba y dependiendo de si la operación se realiza sobre las filas o sobre las columnas de I . Así,

1. F_{ij} es la matriz obtenida al permutar las filas i y j en I .
2. $F_i(\lambda)$ es la matriz obtenida al multiplicar la fila i de I por un escalar $\lambda \neq 0$.
3. $F_{ij}(\lambda)$ es la matriz obtenida al sumar a la fila i de I la fila j multiplicada por el escalar λ .
4. K_{ij} es la matriz obtenida al permutar las columnas i y j en I .
5. $K_i(\lambda)$ es la matriz obtenida al multiplicar la columna i de I por un escalar $\lambda \neq 0$.
6. $K_{ij}(\lambda)$ es la matriz obtenida al sumar a la columna i de I la columna j multiplicada por el escalar λ .

Ejemplos:

Tomando $I \in \mathcal{M}_{3 \times 3}(\mathbb{R})$, tenemos

$$F_{23} = K_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, \quad K_2(3) = F_2(3) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$$F_{13}(2) = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad K_{13}(2) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 2 & 0 & 1 \end{pmatrix}.$$

Efectos de las matrices elementales.

Las operaciones elementales sobre las filas y columnas de una matriz A pueden obtenerse como resultado de multiplicar por una matriz elemental:

1. Realizar una operación elemental sobre las filas de $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ es equivalente a multiplicar A por la izquierda por la correspondiente matriz elemental de filas $F \in \mathcal{M}_{p \times p}(\mathbb{R})$.
2. Realizar una operación elemental sobre las columnas de $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ es equivalente a multiplicar A por la derecha por la correspondiente matriz elemental de columnas $K \in \mathcal{M}_{n \times n}(\mathbb{R})$.

Ejemplos:

$$\text{Sea } A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}.$$

1. Permutar las columnas 1 y 3 de A es equivalente a multiplicar A por la derecha por K_{13} :

$$AK_{13} = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} = \begin{pmatrix} 3 & 2 & 1 \\ 6 & 5 & 4 \end{pmatrix}.$$

2. Restar a la fila 2 de A la fila 1 multiplicada por 3 es equivalente a multiplicar A por la izquierda por $F_{21}(-3)$:

$$F_{21}(-3)A = \begin{pmatrix} 1 & 0 \\ -3 & 1 \end{pmatrix} \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix} = \begin{pmatrix} 1 & 2 & 3 \\ 1 & -1 & -3 \end{pmatrix}.$$

Inversas de las matrices elementales.

Es muy sencillo comprobar que todas las matrices elementales son inversibles y además su inversa es la matriz elemental equivalente a la “transformación inversa”. Así,

1. Por filas:

$$(F_{ij})^{-1} = F_{ij} \quad , \quad (F_i(\lambda))^{-1} = F_i(1/\lambda) \quad , \quad (F_{ij}(\lambda))^{-1} = F_{ij}(-\lambda).$$

2. Por columnas:

$$(K_{ij})^{-1} = K_{ij} \quad , \quad (K_i(\lambda))^{-1} = K_i(1/\lambda) \quad , \quad (K_{ij}(\lambda))^{-1} = K_{ij}(-\lambda).$$

2.7 Forma escalonada y rango de una matriz.

Definición 2.10 Sea $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$. Supongamos que la fila i de A no tiene todos los elementos iguales a cero. Se llama **entrada principal** de la fila i al primer elemento de dicha fila distinto de cero, es decir, al elemento a_{ij} tal que $a_{ij} \neq 0$, $a_{ik} = 0 \forall k < j$.

Definición 2.11 Se dice que la matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ está en **forma escalonada** si cumple las dos siguientes condiciones:

1. Si hay alguna fila de ceros, está al final.
2. Si hay varias filas distintas de cero, entonces la entrada principal de cada fila no nula está más a la izquierda que la de la siguiente fila.

Definición 2.12 Se dice que la matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ está en **forma escalonada reducida** si cumple las siguientes condiciones:

1. Está en forma escalonada.
2. Todas las entradas principales son iguales a 1.
3. En cada columna donde hay una entrada principal, el resto de los elementos son ceros.

Ejemplo: La matriz

$$A = \begin{pmatrix} \boxed{1} & -1 & 0 & 2 & 0 \\ 0 & 0 & \boxed{1} & 3 & 0 \\ 0 & 0 & 0 & 0 & \boxed{1} \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

está en forma escalonada reducida. Se han resaltado sus entradas principales.

El siguiente resultado es clave para las aplicaciones de las operaciones elementales:

Teorema 2.1 (Reducción de Gauss-Jordan) Toda matriz se puede transformar en una matriz en forma escalonada reducida mediante operaciones elementales por filas.

Definición 2.13 Para cada matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, la matriz obtenida mediante el teorema anterior es única y recibe el nombre de **forma escalonada reducida de A**. La denotaremos por $\text{rref}(A)$.

Ejemplo: Hallar la forma escalonada reducida de

$$A = \begin{pmatrix} -1 & -1 & 0 & 3 & -2 \\ 3 & 3 & 2 & -1 & 0 \\ -3 & -3 & -2 & 1 & 0 \\ 2 & 2 & 3 & 0 & -2 \end{pmatrix}.$$

$$\begin{aligned} A &= \begin{pmatrix} -1 & -1 & 0 & 3 & -2 \\ 3 & 3 & 2 & -1 & 0 \\ -3 & -3 & -2 & 1 & 0 \\ 2 & 2 & 3 & 0 & -2 \end{pmatrix} \xrightarrow{\substack{F_{21}(3) \\ F_{31}(-3), F_{41}(2)}}} \begin{pmatrix} -1 & -1 & 0 & 3 & -2 \\ 0 & 0 & 2 & 8 & -6 \\ 0 & 0 & -2 & -8 & 6 \\ 0 & 0 & 3 & 6 & -6 \end{pmatrix} \\ &\xrightarrow{\substack{F_{32}(1) \\ F_{42}(-3/2)}}} \begin{pmatrix} -1 & -1 & 0 & 3 & -2 \\ 0 & 0 & 2 & 8 & -6 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -6 & 3 \end{pmatrix} \xrightarrow{F_{34}} \begin{pmatrix} -1 & -1 & 0 & 3 & -2 \\ 0 & 0 & 2 & 8 & -6 \\ 0 & 0 & 0 & -6 & 3 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \\ &\xrightarrow{\substack{F_1(-1) \\ F_2(1/2), F_3(-1/6)}}} \begin{pmatrix} 1 & 1 & 0 & -3 & 2 \\ 0 & 0 & 1 & 4 & -3 \\ 0 & 0 & 0 & 1 & -1/2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{\substack{F_{23}(-4) \\ F_{13}(3)}}} \begin{pmatrix} 1 & 1 & 0 & 0 & 1/2 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1/2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Por tanto,

$$\text{rref}(A) = \begin{pmatrix} \boxed{1} & 1 & 0 & 0 & 1/2 \\ 0 & 0 & \boxed{1} & 0 & -1 \\ 0 & 0 & 0 & \boxed{1} & -1/2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Rango de una matriz.

Definición 2.14 Sea $A \in \mathcal{M}_{p \times n}(\mathbb{R})$. Se define el **rango** de A como el número de filas no nulas de la forma escalonada reducida de A . Se denota $\text{rg}(A)$.

Ejemplo: En el ejemplo anterior, $\text{rg}(A) = 3$.

Observación: En la práctica no es preciso calcular la forma escalonada reducida de A . El rango de filas de A coincide con el número de filas no nulas de cualquier matriz escalonada obtenida realizando operaciones elementales sobre las filas de A . De hecho, para calcular el rango de A se pueden combinar operaciones elementales por filas y por columnas hasta obtener una matriz en forma escalonada.

La siguiente propiedad proporciona un método para determinar si una matriz tiene inversa usando operaciones elementales.

Proposición 2.1 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz cuadrada. Las siguientes afirmaciones son equivalentes:

- (1) A es inversible.
- (2) $\text{rref}(A) = I$.
- (3) $\text{rg}(A) = n$.

Demostración. Recordemos que $\text{rref}(A)$ se obtiene haciendo operaciones elementales sobre las filas de A . Por tanto, $\text{rref}(A) = FA$, donde F es una matriz que resulta de multiplicar matrices elementales. En particular, F es inversible. Veamos que se cumplen las equivalencias:

(1) \implies (2): Como A es inversible, $\text{rref}(A) = FA$ también es inversible y por tanto no tiene filas de ceros. Necesariamente $\text{rref}(A) = I$.

(2) \implies (3): Como $\text{rref}(A) = I$, $\text{rref}(A)$ tiene n filas no nulas y por tanto $\text{rg}(A) = n$.

(3) \implies (1): Como $\text{rg}(A) = n$, $\text{rref}(A)$ tiene n filas no nulas y por tanto $\text{rref}(A) = I$. Esto quiere decir que existe una matriz F tal que $FA = \text{rref}(A) = I$. Por definición, A es inversible y $F = A^{-1}$. \square

2.8 Cálculo de la inversa.

Como consecuencia de que la forma escalonada reducida de las matrices inversibles es la identidad, se tiene el siguiente resultado:

Proposición 2.2 *Toda matriz inversible $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ se puede transformar en la matriz identidad mediante operaciones elementales por filas.*

Esta proposición permite calcular la inversa de A utilizando operaciones elementales del siguiente modo: sean F_1, F_2, \dots, F_k las matrices elementales de filas por las que debemos multiplicar A para llegar a la identidad, es decir, $F_k \dots F_2 F_1 A = I$. Entonces $A^{-1} = F_k \dots F_2 F_1$.

En la práctica, se procede del siguiente modo: si escribimos la matriz ampliada $(A|I)$, el resultado de aplicar F_1, F_2, \dots, F_k sobre esta matriz es $(I|A^{-1})$:

$$(A|I) \xrightarrow{F_1, F_2, \dots, F_k} (F_k \dots F_2 F_1 A | F_k \dots F_2 F_1 I) = (I|A^{-1}).$$

Ejemplo:

Para calcular la inversa de

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 3 \end{pmatrix},$$

realizamos las siguientes operaciones elementales:

$$\begin{aligned} (A|I) &= \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 1 & 2 & 0 & 0 & 1 & 0 \\ 1 & 0 & 3 & 0 & 0 & 1 \end{array} \right) \xrightarrow{F_{21}(-1)} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 1 & 0 & 3 & 0 & 0 & 1 \end{array} \right) \\ &\xrightarrow{F_{31}(-1)} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & -1 & 2 & -1 & 0 & 1 \end{array} \right) \xrightarrow{F_{32}(1)} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & -1 & -1 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 & 1 \end{array} \right) \\ &\xrightarrow{F_{23}(1)} \left(\begin{array}{ccc|ccc} 1 & 1 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -3 & 2 & 1 \\ 0 & 0 & 1 & -2 & 1 & 1 \end{array} \right) \xrightarrow{F_{13}(-1)} \left(\begin{array}{ccc|ccc} 1 & 1 & 0 & 3 & -1 & -1 \\ 0 & 1 & 0 & -3 & 2 & 1 \\ 0 & 0 & 1 & -2 & 1 & 1 \end{array} \right) \\ &\xrightarrow{F_{12}(-1)} \left(\begin{array}{ccc|ccc} 1 & 0 & 0 & 6 & -3 & -2 \\ 0 & 1 & 0 & -3 & 2 & 1 \\ 0 & 0 & 1 & -2 & 1 & 1 \end{array} \right) = (I|A^{-1}). \end{aligned}$$

Por tanto,

$$A^{-1} = \begin{pmatrix} 6 & -3 & -2 \\ -3 & 2 & 1 \\ -2 & 1 & 1 \end{pmatrix}.$$

Observación: En ningún caso se pueden combinar operaciones elementales de filas y columnas para calcular la inversa.

2.9 Determinantes.

Las operaciones elementales también se usan como un método eficaz para calcular el determinante de una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, teniendo en cuenta las siguientes propiedades:

- Sumar a una fila o columna de una matriz un múltiplo de otra fila o columna no varía el valor del determinante.
- Permutar dos filas o dos columnas de una matriz hace que su determinante cambie de signo.
- Si A es una matriz triangular entonces su determinante es el producto de los elementos de la diagonal.

De este modo, realizando operaciones elementales en A obtenemos una matriz en forma triangular cuyo determinante se calcula haciendo uso de la propiedad c).

Ejemplo:

$$\left| \begin{array}{ccc|c} 1 & 1 & 2 & F_{21}(-1) \\ 1 & 1 & 0 & = \\ 2 & 1 & 2 & F_{31}(-2) \end{array} \right| \left| \begin{array}{ccc|c} 1 & 1 & 2 & F_{23} \\ 0 & 0 & -2 & = - \\ 0 & -1 & -2 & \end{array} \right| \left| \begin{array}{ccc|c} 1 & 0 & 2 \\ 0 & -1 & -2 \\ 0 & 0 & -2 \end{array} \right| = -2.$$

En ocasiones conviene combinar este método con el desarrollo por los elementos de una fila o una columna (regla de Laplace).

Sea $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$. Sea \widetilde{A}_{ij} la matriz que se obtiene suprimiendo en A la fila i y la columna j . Entonces, para cada fila i de A , se tiene:

$$\det(A) = \sum_{j=1}^n (-1)^{i+j} a_{ij} \det(\widetilde{A}_{ij}).$$

Esta fórmula permite expresar el determinante de una matriz de orden n en función del determinante de n matrices de orden $(n - 1)$. También se verifica una fórmula análoga para cada columna de A . En particular, se tienen las siguientes consecuencias:

- Si $n=2$,

$$\left| \begin{array}{cc} a & b \\ c & d \end{array} \right| = ad - bc.$$

- Si A tiene una fila o una columna de ceros entonces $|A| = 0$.
- Si el único elemento no nulo de la fila i es a_{ik} entonces $\det(A) = (-1)^{i+k} a_{ik} \det(\widetilde{A}_{ik})$.

Otras propiedades de los determinantes:

- $|AB| = |A| |B|$, $\forall A, B \in \mathcal{M}_{n \times n}(\mathbb{R})$.

2. $|A^t| = |A|, \forall A \in \mathcal{M}_{n \times n}(\mathbb{R})$.

3. Si $\lambda \in \mathbb{R}$ entonces

$$\begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda a_{i1} & \lambda a_{i2} & \cdots & \lambda a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix} = \lambda \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{vmatrix}$$

La misma propiedad es válida si una columna está multiplicada por el escalar λ .

4. $|\lambda A| = \lambda^n |A|, \forall A \in \mathcal{M}_{n \times n}(\mathbb{R}), \forall \lambda \in \mathbb{R}$. En particular, $|-A| = (-1)^n |A|$.

5. Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ entonces A es inversible si y sólo si $|A| \neq 0$. Además, en ese caso, $|A^{-1}| = 1/|A|$.

Prueba de la propiedad 5.

Si A es inversible, entonces $A^{-1}A = I$ y por tanto $|A^{-1}| |A| = |A^{-1}A| = |I| = 1$. De aquí se obtiene que $|A| \neq 0$ y además $|A^{-1}| = 1/|A|$.

Supongamos ahora que $|A| \neq 0$ y consideremos su forma escalonada reducida $\text{rref}(A)$. Existe una matriz inversible F tal que $\text{rref}(A) = FA$, y por tanto $|\text{rref}(A)| = |F| |A| \neq 0$.

En consecuencia, $\text{rref}(A)$ no puede tener filas de ceros y se concluye que A es inversible. \square

Capítulo 3

Sistemas de ecuaciones lineales

3.1 Introducción.

Este capítulo está dedicado a la resolución de sistemas de ecuaciones lineales, lo que incluye el estudio de la compatibilidad del sistema (existencia de soluciones), la determinación del conjunto de soluciones y la interpretación geométrica de dicho conjunto. El método principal de resolución es el método de Gauss, basado en operaciones elementales sobre las filas de la matriz ampliada del sistema.

3.2 Expresión matricial.

Un **sistema** de p ecuaciones lineales con n incógnitas en \mathbb{R} es un conjunto de expresiones:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \dots &= \dots \\ \dots &= \dots \\ a_{p1}x_1 + a_{p2}x_2 + \cdots + a_{pn}x_n &= b_p, \end{aligned}$$

donde los elementos $a_{ij} \in \mathbb{R}$ se llaman *coeficientes del sistema*, $b_i \in \mathbb{R}$ se llaman *términos independientes* y x_i se llaman *incógnitas*.

El sistema es **homogéneo** si $b_i = 0, \forall i = 1, 2, \dots, p$. En otro caso diremos que es no homogéneo.

El sistema se puede expresar en la forma matricial $Ax = b$, donde

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} \end{pmatrix} \in \mathcal{M}_{p \times n}(\mathbb{R}) \quad ; \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_p \end{pmatrix} \in \mathbb{R}^p \quad ; \quad x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} .$$

La matriz A se llama *matriz de coeficientes* del sistema y b es el término independiente.

La matriz

$$(A|b) = \left(\begin{array}{cccc|c} a_{11} & a_{12} & \cdots & a_{1n} & b_1 \\ a_{21} & a_{22} & \cdots & a_{2n} & b_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ a_{p1} & a_{p2} & \cdots & a_{pn} & b_p \end{array} \right) \in \mathcal{M}_{p \times (n+1)}(\mathbb{R})$$

se llama *matriz ampliada del sistema*. Cada una de las ecuaciones se puede identificar con la correspondiente fila de la matriz $(A|b)$. Obsérvese que el número de columnas de A coincide con el número de incógnitas del sistema.

3.3 Existencia de soluciones.

Definición 3.1 Un vector $v = (v_1, v_2, \dots, v_n) \in \mathbb{R}^n$ es una **solución** del sistema si $Av = b$.

Resolver el sistema es determinar el conjunto de sus soluciones (que es un subconjunto de \mathbb{R}^n). Si no existe ninguna solución, el sistema es **incompatible**. Si existe alguna solución, diremos que el sistema es **compatible determinado** si la solución es única y **compatible indeterminado** si existe más de una solución.

Eliminación gaussiana.

La siguiente propiedad permitirá estudiar con facilidad si un sistema es compatible y calcular el conjunto de sus soluciones.

Proposición 3.1 Sea $Ax = b$ un sistema de p ecuaciones lineales con n incógnitas. Si efectuamos operaciones elementales sobre las filas de la matriz ampliada $(A|b)$ hasta obtener una nueva matriz $(A'|b')$ entonces los sistemas $Ax = b$ y $A'x = b'$ son equivalentes, es decir, tienen el mismo conjunto de soluciones.

Demostración. Sea $F = F_k \dots F_2 F_1$, donde F_1, F_2, \dots, F_k son las matrices elementales correspondientes a las operaciones por filas sobre $(A|b)$. Entonces $(A'|b') = (FA|Fb)$ y el nuevo sistema es $FAx = Fb$, que es equivalente a $Ax = b$ ya que F es inversible. \square

Utilizando esta proposición, para resolver un sistema se realizan operaciones elementales sobre las filas de $(A|b)$ hasta obtener su forma escalonada reducida $(A'|b')$. Sea $r = \text{rg}(A|b) = \text{rg}(A'|b')$. El sistema $A'x = b'$ se resuelve de forma inmediata, despejando las r incógnitas correspondientes a las entradas principales en función de las $(n - r)$ restantes. De este modo, tenemos:

- Si $\text{rg}(A) \neq \text{rg}(A|b)$ entonces el sistema es incompatible porque en el sistema $A'x = b'$ hay una ecuación $0 = 1$.
- Si $\text{rg}(A) = \text{rg}(A|b) = n$ ($n =$ número de incógnitas $=$ número de columnas de A) entonces el sistema es compatible determinado.
- Si $\text{rg}(A) = \text{rg}(A|b) < n$ entonces el sistema es compatible indeterminado y el conjunto de soluciones se puede escribir en función de $(n - r)$ parámetros.

3.4 Conjuntos de soluciones.

Una de las características especiales de los sistemas de ecuaciones lineales es que aunque el conjunto de soluciones puede ser infinito, siempre queda determinado por un conjunto finito de vectores de \mathbb{R}^n .

Comenzamos analizando el caso de sistemas homogéneos.

Sistemas homogéneos.

Consideremos un sistema homogéneo $Ax = 0$, donde $A \in \mathcal{M}_{p \times n}(\mathbb{R})$. En primer lugar, observemos que un sistema homogéneo siempre es compatible, ya que $x = 0$ es solución. El conjunto de soluciones se denomina núcleo de A y se denota por $\text{Ker}(A)$, es decir,

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

Por tanto sólo hay dos posibilidades:

- Si $\text{rg}(A) = n$ entonces el sistema es compatible determinado y su única solución es el vector cero ($\text{Ker}(A) = \{0\}$).
- Si $\text{rg}(A) = r < n$ entonces el sistema es compatible indeterminado y el núcleo de A es el conjunto de todas las combinaciones lineales de $k = n - r$ vectores de \mathbb{R}^n u_1, u_2, \dots, u_k , es decir,

$$\text{Ker}(A) = \{\lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_k u_k / \lambda_i \in \mathbb{R}, i = 1, \dots, k\}.$$

Se dice que $\text{Ker}(A)$ está generado por los vectores u_1, u_2, \dots, u_k y se denota

$$\text{Ker}(A) = \langle \{u_1, u_2, \dots, u_k\} \rangle.$$

Estos vectores se determinan despejando las incógnitas correspondientes a las entradas principales de la forma escalonada reducida de A en función del resto.

Ejemplo: Consideremos el sistema

$$\begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

Realizando operaciones elementales sobre las filas de la matriz A , tenemos:

$$\begin{aligned} A &= \begin{pmatrix} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 0 \\ 1 & 0 & 2 & 2 \end{pmatrix} \xrightarrow{F_{21}(-1)} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & -1 \\ 1 & 0 & 2 & 2 \end{pmatrix} \xrightarrow{F_{31}(-1)} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & -1 & 1 & 1 \end{pmatrix} \\ &\xrightarrow{F_{32}(1)} \begin{pmatrix} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \xrightarrow{F_{12}(-1)} \begin{pmatrix} 1 & 0 & 2 & 2 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = A' = \text{rref}(A). \end{aligned}$$

Como $\text{rg}(A) = \text{rg}(A') = 2 < 4 = \text{número de incógnitas}$, el sistema es compatible indeterminado. Además, el conjunto de soluciones de $Ax = 0$ coincide con el conjunto de soluciones del sistema equivalente $A'x = 0$, es decir, del sistema

$$\begin{aligned}x + 2z + 2t &= 0 \\y - z - t &= 0.\end{aligned}$$

Despejando las incógnitas x e y en función de z y t , tenemos que el conjunto de soluciones es:

$$\begin{aligned}\text{Ker}(A) &= \{(x, y, z, t) \in \mathbb{R}^4 / x = -2z - 2t, y = z + t\} = \{(-2z - 2t, z + t, z, t) / z, t \in \mathbb{R}\} = \\&= \{z(-2, 1, 1, 0) + t(-2, 1, 0, 1) / z, t \in \mathbb{R}\} = \langle \{(-2, 1, 1, 0), (-2, 1, 0, 1)\} \rangle.\end{aligned}$$

El conjunto de soluciones está formado por las combinaciones lineales de $u_1 = (-2, 1, 1, 0)$ y $u_2 = (-2, 1, 0, 1)$.

Sistemas no homogéneos.

Consideremos ahora un sistema no homogéneo $Ax = b$, con $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, $b \in \mathbb{R}^p$.

El sistema es compatible indeterminado si $\text{rg}(A) = r = \text{rg}(A|b) < n$. En este caso el conjunto de soluciones está determinado por los $k = n - r$ generadores del núcleo de A y un vector p llamado **solución particular**. En concreto, se tiene el siguiente resultado:

Proposición 3.2 Si $\text{rg}(A) = r = \text{rg}(A|b) < n$, el conjunto de soluciones del sistema $Ax = b$ es

$$S = \{p + \lambda_1 u_1 + \lambda_2 u_2 + \cdots + \lambda_k u_k / \lambda_i \in \mathbb{R}, i = 1, \dots, k\} := p + \langle \{u_1, u_2, \dots, u_k\} \rangle,$$

donde p es una solución de $Ax = b$ (es decir, $Ap = b$) y $\langle \{u_1, u_2, \dots, u_k\} \rangle = \text{Ker}(A)$. En notación abreviada, escribiremos el conjunto de soluciones en la forma $S = p + \text{Ker}(A)$.

Demostración. Como el conjunto de soluciones es $S = \{x \in \mathbb{R}^n / Ax = b\}$, se tiene:

$$\begin{aligned}z \in S &\iff Az = b = Ap \iff A(z - p) = Az - Ap = 0 \iff z - p \in \text{Ker}(A) \iff \\&\iff z = p + u, u \in \text{Ker}(A) \iff z \in p + \text{Ker}(A).\end{aligned}$$

□

Ejemplo: Consideremos el sistema

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 2 & 0 \\ 1 & 0 & 2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Realizando operaciones elementales sobre las filas de la matriz ampliada $(A|b)$, tenemos:

$$(A|b) = \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 2 & 1 \end{array} \right) \xrightarrow{F_{21}(-1)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 2 & 1 \end{array} \right) \xrightarrow{F_{31}(-1)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \end{array} \right)$$

$$\xrightarrow{F_{32}(1)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \xrightarrow{F_{12}(-1)} \left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) = (A'|b').$$

En primer lugar, $\text{rg}(A|b) = \text{rg}(A'|b') = 2 < 3 = \text{número de incógnitas}$, y por tanto el sistema es compatible indeterminado. Además, el conjunto de soluciones de $Ax = b$ coincide con el conjunto de soluciones de $A'x = b'$, es decir, del sistema

$$\begin{aligned} x + 2z &= 1 \\ y - z &= 0. \end{aligned}$$

Despejando $x = 1 - 2z$, $y = z$, tenemos que el conjunto de soluciones en función del parámetro z es

$$\begin{aligned} S &= \{(x, y, z) \in \mathbb{R}^3 / y = z, x = 1 - 2z\} = \{(1 - 2z, z, z) / z \in \mathbb{R}\} = \\ &= \{(1, 0, 0) + z(-2, 1, 1) / z \in \mathbb{R}\} = \underbrace{(1, 0, 0)}_p + \underbrace{\{(-2, 1, 1)\}}_{\text{Ker}(A)}. \end{aligned}$$

3.5 Matrices cuadradas y uso de la factorización LU.

Cuando A es una matriz cuadrada, es más sencillo determinar si el sistema $Ax = b$ es compatible determinado:

Proposición 3.3 Sean $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $b \in \mathbb{R}^n$. El sistema $Ax = b$ tiene solución única si y sólo si $\text{rg}(A) = n$.

Demostración. Si $\text{rg}(A) = n$ entonces también se cumple que $\text{rg}(A|b) = n$, ya que la matriz $(A|b)$ tiene n filas y, por tanto, $n = \text{rg}(A) \leq \text{rg}(A|b) \leq n$. \square

Obsérvese que en este caso la única solución del sistema homogéneo asociado $Ax = 0$ es la solución trivial, es decir, $\text{Ker}(A) = \{0\}$. En consecuencia, las siguientes propiedades son equivalentes para una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$:

1. El sistema $Ax = b$ es compatible determinado para cada $b \in \mathbb{R}^n$.
2. $\text{Ker}(A) = \{0\}$.
3. $\text{rg}(A) = n$.
4. A es inversible.
5. $\det(A) \neq 0$.
6. $\text{rref}(A) = I$.

Observación: Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es invertible, entonces la única solución del sistema $Ax = b$ se puede escribir en la forma $x = A^{-1}b$. Sin embargo, en la práctica no se suele calcular la inversa de A para resolver el sistema.

Factorización LU .

La factorización LU consiste en descomponer una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ en el producto $A = LU$, donde $L \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz triangular inferior con todos los elementos diagonales iguales a 1, y $U \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz triangular superior. Diremos que A admite factorización LU si es posible encontrar estas dos matrices.

El método de cálculo de L y U se basa en la eliminación gaussiana. Para poder obtener L y U por este procedimiento será necesario pedir condiciones adicionales a la matriz A .

Definición 3.2 Sea $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$. Para cada $k = 1, 2, \dots, n$, se llama **menor principal** de orden k de A y se denota Δ_k al siguiente determinante:

$$\Delta_k = \begin{vmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{vmatrix}$$

Proposición 3.4 Si todos los menores principales de A son distintos de cero entonces A admite factorización LU . Además, en este caso, dicha factorización es única.

Cálculo de la factorización LU .

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz en las condiciones de la proposición anterior. Entonces es posible transformar la matriz A en una matriz triangular superior U mediante operaciones elementales sobre las filas de A del tipo $F_{ij}(\lambda)$, con $i > j$, es decir, sin efectuar permutaciones de filas y utilizando sólo las filas superiores para modificar las inferiores.

Sean F_1, F_2, \dots, F_k las correspondientes matrices elementales de filas tales que $F_k \dots F_2 F_1 A = U$. Entonces $L = (F_k \dots F_2 F_1)^{-1} = F_1^{-1} F_2^{-1} \dots F_k^{-1}$ es triangular inferior, sus elementos diagonales son iguales a 1 y además $A = LU$.

Ejemplo: Consideremos la matriz:

$$A = \begin{pmatrix} 2 & -1 & 0 & 1 \\ 4 & -4 & 1 & 5 \\ -2 & 1 & -1 & 0 \\ -2 & 5 & -4 & -1 \end{pmatrix} \in \mathcal{M}_{4 \times 4}(\mathbb{R}).$$

Veamos que A admite factorización LU .

Los menores principales de la matriz A son:

$$\begin{aligned}\Delta_1 &= 2 \neq 0 \\ \Delta_2 &= \begin{vmatrix} 2 & -1 \\ 4 & -4 \end{vmatrix} = -4 \neq 0 \\ \Delta_3 &= \begin{vmatrix} 2 & -1 & 0 \\ 4 & -4 & 1 \\ -2 & 1 & -1 \end{vmatrix} = 4 \neq 0 \\ \Delta_4 &= \begin{vmatrix} 2 & -1 & 0 & 1 \\ 4 & -4 & 1 & 5 \\ -2 & 1 & -1 & 0 \\ -2 & 5 & -4 & -1 \end{vmatrix} = 16 \neq 0.\end{aligned}$$

Todos los menores principales de A son no nulos y por tanto admite factorización LU . Para calcular dicha factorización, en primer lugar determinaremos la matriz triangular superior U mediante operaciones elementales sobre las filas de la matriz A del tipo $F_{ij}(\lambda)$, con $i > j$. Así,

$$\begin{aligned}\begin{pmatrix} 2 & -1 & 0 & 1 \\ 4 & -4 & 1 & 5 \\ -2 & 1 & -1 & 0 \\ -2 & 5 & -4 & -1 \end{pmatrix} &\xrightarrow{F_{21}(-2), F_{31}(1), F_{41}(1)} \begin{pmatrix} 2 & -1 & 0 & 1 \\ 0 & -2 & 1 & 3 \\ 0 & 0 & -1 & 1 \\ 0 & 4 & -4 & 0 \end{pmatrix} \\ &\xrightarrow{F_{42}(2)} \begin{pmatrix} 2 & -1 & 0 & 1 \\ 0 & -2 & 1 & 3 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -2 & 6 \end{pmatrix} \xrightarrow{F_{43}(-2)} \begin{pmatrix} 2 & -1 & 0 & 1 \\ 0 & -2 & 1 & 3 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 4 \end{pmatrix} = U.\end{aligned}$$

De esto se deduce que

$$[F_{43}(-2)F_{42}(2)F_{41}(1)F_{31}(1)F_{21}(-2)] A = U$$

y entonces

$$\begin{aligned}L &= [F_{43}(-2)F_{42}(2)F_{41}(1)F_{31}(1)F_{21}(-2)]^{-1} = \\ &= F_{21}(2)F_{31}(-1)F_{41}(-1)F_{42}(-2)F_{43}(2).\end{aligned}$$

Calcular el producto de estas matrices elementales es equivalente a realizar las correspondientes operaciones elementales a la matriz identidad:

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{F_{43}(2), F_{42}(-2), F_{41}(-1)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & -2 & 2 & 1 \end{pmatrix}$$

$$\xrightarrow{F_{31}(-1)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & -2 & 2 & 1 \end{pmatrix} \xrightarrow{F_{21}(2)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & -2 & 2 & 1 \end{pmatrix} = L.$$

Observación: En la práctica no es necesario comprobar previamente que todos los menores principales de A son no nulos. Esto es equivalente a que se pueda obtener la matriz U mediante operaciones elementales sobre las filas de A del tipo $F_{ij}(\lambda)$, con $i > j$, y además los elementos diagonales de U sean distintos de cero.

Uso de la factorización LU .

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz cuadrada de rango n . Supongamos que A admite factorización LU . Entonces resolver el sistema de ecuaciones lineales $Ax = b$ es equivalente a resolver consecutivamente los sistemas $Lz = b$, $Ux = z$. (En efecto, $Ax = LUx = Lz = b$).

Ejemplo: Sean

$$A = \begin{pmatrix} 2 & -1 & 0 & 1 \\ 4 & -4 & 1 & 5 \\ -2 & 1 & -1 & 0 \\ -2 & 5 & -4 & -1 \end{pmatrix} ; \quad b = \begin{pmatrix} -5 \\ -14 \\ 1 \\ 1 \end{pmatrix}.$$

Vamos a resolver el sistema $Ax = b$ usando la factorización LU .

Ya hemos calculado la factorización LU de la matriz A:

$$A = \begin{pmatrix} 2 & -1 & 0 & 1 \\ 4 & -4 & 1 & 5 \\ -2 & 1 & -1 & 0 \\ -2 & 5 & -4 & -1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & -2 & 2 & 1 \end{pmatrix} \begin{pmatrix} 2 & -1 & 0 & 1 \\ 0 & -2 & 1 & 3 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & 0 & 4 \end{pmatrix} = LU.$$

Como $A = LU$, la resolución del sistema $Ax = b$ es equivalente a la resolución sucesiva de dos sistemas triangulares:

$$Ax = b \iff L \underbrace{Ux}_z = b \iff \begin{cases} Lz = b \\ Ux = z \end{cases}$$

La solución $z = (z_1, z_2, z_3, z_4)^t$ del sistema $Lz = b$ viene dada por

$$\begin{aligned} z_1 &= -5 \\ 2z_1 + z_2 &= -14 \implies z_2 = -4 \\ -z_1 + z_3 &= 1 \implies z_3 = -4 \\ -z_1 - 2z_2 + 2z_3 + z_4 &= 1 \implies z_4 = -4 \end{aligned}$$

Calculamos ahora la solución del sistema $Ux = z$:

$$4x_4 = -4 \implies x_4 = -1$$

$$-x_3 + x_4 = -4 \implies x_3 = 3$$

$$-2x_2 + x_3 + 3x_4 = -4 \implies x_2 = 2$$

$$2x_1 - x_2 + x_4 = -5 \implies x_1 = -1$$

Se puede comprobar que $x = (x_1, x_2, x_3, x_4) = (-1, 2, 3, -1)$ es la solución del sistema original $Ax = b$.

Capítulo 4

Espacios vectoriales y aplicaciones lineales

4.1 Introducción.

En este capítulo introduciremos la definición de espacio vectorial y los principales conceptos relacionados, como la independencia lineal, generadores, base y dimensión. También se interpretan las matrices como aplicaciones lineales.

4.2 Espacios y subespacios vectoriales.

Definición 4.1 Se llama **espacio vectorial** sobre \mathbb{R} o **espacio vectorial real** a un conjunto V dotado de dos operaciones:

- Una operación interna (suma), de tal forma que $(V, +)$ es un grupo conmutativo.
- Una operación externa (producto por escalares) que asigna a cada escalar $\lambda \in \mathbb{R}$ y a cada elemento $v \in V$ un nuevo elemento $\lambda v \in V$, de tal forma que se cumplen las siguientes propiedades:

1. $\lambda(v + w) = \lambda v + \lambda w, \forall \lambda \in \mathbb{R}, \forall v, w \in V.$

2. $(\lambda + \mu)v = \lambda v + \mu v, \forall \lambda, \mu \in \mathbb{R}, \forall v \in V.$

3. $(\lambda\mu)v = \lambda(\mu v), \forall \lambda, \mu \in \mathbb{R}, \forall v \in V.$

4. $1v = v, \forall v \in V$, donde 1 es el elemento neutro del producto en \mathbb{R} .

A los elementos de V los llamaremos *vectores* y a los elementos de \mathbb{R} los llamaremos *escalares*. Generalmente denotaremos a estos últimos con letras del alfabeto griego. Si hay posibilidad de confusión, al elemento neutro de la suma en V lo denotaremos por θ para distinguirlo del cero de \mathbb{R} .

Ejemplos:

1. \mathbb{R}^n es un espacio vectorial real con las operaciones usuales de suma y producto por escalares.
2. El conjunto $\mathcal{M}_{p \times n}(\mathbb{R})$ de las matrices reales de p filas y n columnas es un espacio vectorial sobre \mathbb{R} con las operaciones definidas en el capítulo 1.
3. El conjunto $\Pi_n(\mathbb{R})$ de los polinomios en una variable de grado menor o igual que n y con coeficientes en \mathbb{R} es un espacio vectorial real con las operaciones habituales de suma de polinomios y producto de un escalar por un polinomio.

$$\Pi_n(\mathbb{R}) = \{a_0 + a_1x + \cdots + a_nx^n / a_0, a_1, \dots, a_n \in \mathbb{R}\}.$$

Subespacios vectoriales.

Definición 4.2 Sea V un espacio vectorial. Un subconjunto U de V es un **subespacio vectorial** de V si cumple las siguientes propiedades:

- (1) $0 \in U$.
- (2) $u_1 + u_2 \in U, \forall u_1, u_2 \in U$.
- (3) $\lambda u \in U, \forall \lambda \in \mathbb{R}, \forall u \in U$.

Las propiedades (2) y (3) de la definición se pueden sustituir por la siguiente:

- (4) $\lambda_1 u_1 + \lambda_2 u_2 \in U, \forall \lambda_1, \lambda_2 \in \mathbb{R}, \forall u_1, u_2 \in U$.

Ejemplos:

1. El conjunto $U = \{A \in \mathcal{M}_{n \times n}(\mathbb{R}) / A^t = A\}$ es un subespacio vectorial de $\mathcal{M}_{n \times n}(\mathbb{R})$.

Es evidente que $0 \in U$, ya que $0^t = 0$. Veamos que se cumple la propiedad (4): si $A, B \in U$ entonces $A^t = A, B^t = B$, y por tanto, si $\lambda, \mu \in \mathbb{R}$ entonces:

$$(\lambda A + \mu B)^t = \lambda A^t + \mu B^t = \lambda A + \mu B,$$

es decir, $(\lambda A + \mu B) \in U, \forall \lambda, \mu \in \mathbb{R}$.

2. El conjunto $W = \{A \in \mathcal{M}_{2 \times 2}(\mathbb{R}) / \det(A) = 0\}$ no es un subespacio vectorial de $\mathcal{M}_{2 \times 2}(\mathbb{R})$. Aunque $0 \in W$, veamos que no se cumple la propiedad (2); para ello basta tomar

$$A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, A_2 = \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}.$$

Es claro que A_1 y A_2 pertenecen a W ya que $\det(A_1) = \det(A_2) = 0$. Sin embargo,

$$\det(A_1 + A_2) = \begin{vmatrix} 1 & 0 \\ 0 & 1 \end{vmatrix} = 1 \neq 0 \implies A_1 + A_2 \notin W.$$

Recordemos que si v_1, v_2, \dots, v_n son n vectores de un espacio vectorial V y $\lambda_1, \dots, \lambda_n$ son números reales, entonces cualquier vector de la forma

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n$$

se llama **combinación lineal** de v_1, v_2, \dots, v_n .

Tenemos la siguiente caracterización de los subespacios vectoriales:

Proposición 4.1 *Un subconjunto no vacío U de un espacio vectorial V es un subespacio vectorial si y sólo si todas las combinaciones lineales de vectores de U pertenecen a U .*

Definición 4.3 *Sea U un subespacio vectorial de un espacio vectorial V . Se dice que un subconjunto S de U es un **conjunto de generadores** de U si todo vector de U es combinación lineal de vectores de S . Se denota $U = \langle S \rangle$. Si S es un conjunto de generadores de U , diremos que U es el subespacio generado por S .*

Proposición 4.2 *Si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, entonces $\text{Ker}(A) = \{x \in \mathbb{R}^n / Ax = 0\}$ es un subespacio vectorial de \mathbb{R}^n .*

Demostración. Es claro que $0 \in \text{Ker}(A)$ ya que $A0 = 0$. Además, si $x_1, x_2 \in \text{Ker}(A)$ y $\lambda_1, \lambda_2 \in \mathbb{R}$, entonces

$$A(\lambda_1 x_1 + \lambda_2 x_2) = \lambda_1 Ax_1 + \lambda_2 Ax_2 = \lambda_1 0 + \lambda_2 0 = 0,$$

de modo que $\lambda_1 x_1 + \lambda_2 x_2 \in \text{Ker}(A)$. □

Hallar un conjunto de generadores de $\text{Ker}(A)$ es equivalente a resolver el sistema homogéneo $Ax = 0$.

Ejemplo:

Sea $U = \{(x, y, z) \in \mathbb{R}^3 / x + y + z = 0\}$. Podemos escribir:

$$\begin{aligned} U &= \{(-y - z, y, z) / y, z \in \mathbb{R}\} = \{y(-1, 1, 0) + z(-1, 0, 1) / y, z \in \mathbb{R}\} \\ &= \langle \{(-1, 1, 0), (-1, 0, 1)\} \rangle. \end{aligned}$$

En muchas ocasiones la forma más sencilla de probar que un subconjunto U de un espacio vectorial V es un subespacio consiste en encontrar un conjunto de generadores.

Ejemplo: Sea $U = \{p(x) \in \Pi_2(\mathbb{R}) / p(1) = 0\}$.

Consideremos un polinomio arbitrario $p(x) = a + bx + cx^2 \in \Pi_2(\mathbb{R})$. Entonces:

$$p(x) \in U \iff p(1) = 0 \iff a + b + c = 0.$$

Podemos reescribir U como:

$$\begin{aligned} U &= \{a + bx + cx^2 \in \Pi_2(\mathbb{R}) / a + b + c = 0\} = \{a + bx + cx^2 \in \Pi_2(\mathbb{R}) / c = -a - b\} = \\ &= \{a + bx + (-a - b)x^2 / a, b \in \mathbb{R}\} = \{a(1 - x^2) + b(x - x^2) / a, b \in \mathbb{R}\} = \langle \{1 - x^2, x - x^2\} \rangle. \end{aligned}$$

Por tanto, U es el subespacio vectorial de $\Pi_2(\mathbb{R})$ generado por $1 - x^2$ y $x - x^2$.

4.3 Independencia lineal.

Sea V un espacio vectorial y S un subconjunto de V . Se dice que un vector $v \in V$ depende linealmente de los vectores de S si v es combinación lineal de vectores de S , es decir, si existen $\lambda_1, \dots, \lambda_n \in \mathbb{R}$, $v_1, v_2, \dots, v_n \in S$ tales que $v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n$.

Definición 4.4 *Un conjunto de vectores es linealmente independiente o libre si ninguno de ellos es combinación lineal del resto.*

Definición 4.5 *Sea $S = \{v_1, v_2, \dots, v_n\}$ un conjunto de vectores de un espacio vectorial V . Se llama **rango** de S al mayor número de vectores linealmente independientes que hay en S . Se denota $\text{rg}(S)$.*

De la definición anterior se deduce inmediatamente que un conjunto $S = \{v_1, v_2, \dots, v_n\}$ es libre si y sólo si $\text{rg}(S) = n$.

Proposición 4.3 *Si S es un conjunto de p vectores de \mathbb{R}^n , entonces $\text{rg}(S) = \text{rg}(A)$, donde $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ es la matriz cuyas filas son los vectores de S .*

Demostración. Es consecuencia de que la independencia lineal de un conjunto de vectores no varía por operaciones elementales y el conjunto de filas no nulas de una matriz escalonada es linealmente independiente. \square

Ejemplo:

Sea $S = \{(1, 2, 1, 1), (-1, 1, 0, 0), (1, 5, 2, 2)\}$. Entonces:

$$\text{rg}(S) = \text{rg} \begin{pmatrix} 1 & 2 & 1 & 1 \\ -1 & 1 & 0 & 0 \\ 1 & 5 & 2 & 2 \end{pmatrix} \begin{array}{l} F_{21}(1) \\ \\ F_{31}(-1) \end{array} = \text{rg} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 3 & 1 & 1 \\ 0 & 3 & 1 & 1 \end{pmatrix} \begin{array}{l} F_{32}(-1) \\ \\ \end{array} = \text{rg} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 3 & 1 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = 2.$$

Observación: Si sólo se realizan operaciones elementales por filas en A para determinar una matriz escalonada A' y obtener el rango de S entonces el subespacio generado por S coincide con el subespacio generado por las filas no nulas de A' . Esta propiedad no es cierta si se combinan operaciones de filas y columnas para calcular el rango.

En el ejemplo anterior,

$$U = \langle S \rangle = \langle \{(1, 2, 1, 1), (-1, 1, 0, 0), (1, 5, 2, 2)\} \rangle = \langle \{(1, 2, 1, 1), (0, 3, 1, 1)\} \rangle .$$

4.4 Bases y dimensión.

Definición 4.6 *Un conjunto de vectores \mathcal{B} de un espacio vectorial V es una **base** de V si \mathcal{B} es libre y $V = \langle \mathcal{B} \rangle$.*

Ejemplos:

1. El conjunto $\mathcal{C} = \{(1, 0, \dots, 0), (0, 1, 0, \dots, 0), \dots, (0, 0, \dots, 0, 1)\}$ es una base de \mathbb{R}^n llamada **base canónica**.
2. El conjunto $\mathcal{B} = \{1, x, x^2, \dots, x^n\}$ es una base del espacio de polinomios $\Pi_n(\mathbb{R})$.
3. El conjunto

$$\mathcal{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$$

es una base de $\mathcal{M}_{2 \times 2}(\mathbb{R})$.

Dimensión.

Aunque un espacio vectorial tiene infinitas bases, todas ellas tienen el mismo número de vectores. Se llama **dimensión** de V al número de vectores de cualquier base de V . Se denota $\dim(V)$.

Ejemplos:

Para los espacios vectoriales que hemos mencionado anteriormente, se tiene:

$$\dim(\mathbb{R}^n) = n, \quad \dim(\Pi_n(\mathbb{R})) = n + 1, \quad \dim(\mathcal{M}_{2 \times 2}(\mathbb{R})) = 4.$$

Observación: Si $V = \{0\}$ entonces no existe ninguna base de V y, por convenio, definiremos $\dim(V) = 0$.

Cálculo de la dimensión.

- En primer lugar, si $V = \langle \{v_1, v_2, \dots, v_p\} \rangle$ entonces $\dim(V) = \text{rg}(\{v_1, v_2, \dots, v_p\})$.

Ejemplo:

Sea $U = \langle \{(1, 2, 1, 1), (0, 1, -1, -1), (0, 0, 0, 1)\} \rangle$. Entonces

$$\dim(U) = \text{rg} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & 1 & -1 & -1 \\ 0 & 0 & 0 & 1 \end{pmatrix} = 3.$$

- Si $U = \text{Ker}(A)$, con $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, entonces U es un subespacio de \mathbb{R}^n de dimensión $d = n - \text{rg}(A)$.

Ejemplo:

$$\text{Sea } U = \left\{ (x, y, z, t) \in \mathbb{R}^4 / \begin{array}{l} x + 2y + z + t = 0 \\ x - y = 0 \\ x + 5y + 2z + 2t = 0 \end{array} \right\} = \text{Ker} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & 5 & 2 & 2 \end{pmatrix}.$$

$$\begin{aligned} \dim(U) &= 4 - \text{rg} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 1 & -1 & 0 & 0 \\ 1 & 5 & 2 & 2 \end{pmatrix} \begin{matrix} F_{21}(-1) \\ = \\ F_{31}(-1) \end{matrix} = 4 - \text{rg} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & -3 & -1 & -1 \\ 0 & 3 & 1 & 1 \end{pmatrix} = \\ & \begin{matrix} F_{32}(1) \\ = \end{matrix} 4 - \text{rg} \begin{pmatrix} 1 & 2 & 1 & 1 \\ 0 & -3 & -1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} = 4 - 2 = 2. \end{aligned}$$

Esta propiedad se puede extender a cualquier espacio vectorial de dimensión finita V : Si U es un subespacio de V entonces la dimensión de U es igual a la dimensión de V menos el número de ecuaciones linealmente independientes que definen a U .

Por ejemplo, si $U = \{A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R}) / a_{ii} = 0, \forall i = 1, 2, \dots, n\}$ entonces

$$\dim(U) = \dim(\mathcal{M}_{n \times n}(\mathbb{R})) - n = n^2 - n.$$

4.5 Cambio de base en \mathbb{R}^n .

La siguiente propiedad es una consecuencia inmediata de la definición de base y permite introducir el concepto de vector de coordenadas:

Proposición 4.4 *Sea $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ una base de \mathbb{R}^n . Cada $x \in \mathbb{R}^n$ se puede escribir de modo único como*

$$x = \lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_n u_n.$$

El vector $(\lambda_1, \lambda_2, \dots, \lambda_n)$ se llama vector de coordenadas de x respecto de la base \mathcal{B} y se suele denotar $x = (\lambda_1, \lambda_2, \dots, \lambda_n)_{\mathcal{B}}$.

Ejemplo: En \mathbb{R}^3 se considera la base $\mathcal{B} = \{(1, 1, 1), (1, 2, 0), (0, 0, 1)\}$.

Calculamos las coordenadas de $x = (1, 0, 0)$ respecto de \mathcal{B} :

Si $(1, 0, 0) = (\alpha, \beta, \gamma)_{\mathcal{B}}$ entonces:

$$(1, 0, 0) = \alpha(1, 1, 1) + \beta(1, 2, 0) + \gamma(0, 0, 1) = (\alpha + \beta, \alpha + 2\beta, \alpha + \gamma) \iff$$

$$\iff \begin{cases} \alpha + \beta = 1 \\ \alpha + 2\beta = 0 \\ \alpha + \gamma = 0 \end{cases} \iff \begin{cases} \alpha = 2 \\ \beta = -1 \\ \gamma = -2. \end{cases}$$

Por tanto, $(1, 0, 0) = (2, -1, -2)_{\mathcal{B}}$.

Si \mathcal{B} es una base de un espacio vectorial V y $x = (\lambda_1, \lambda_2, \dots, \lambda_n)_{\mathcal{B}}$ entonces denotaremos

$$x_{\mathcal{B}} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} \in \mathcal{M}_{n \times 1}(\mathbb{R}).$$

Observemos que si consideramos la base canónica \mathcal{C} , entonces las coordenadas de un vector $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ respecto de \mathcal{C} son precisamente (x_1, x_2, \dots, x_n) , es decir,

$$x_{\mathcal{C}} = x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} \in \mathcal{M}_{n \times 1}(\mathbb{R}).$$

A continuación veremos cómo cambian las coordenadas de un vector x al cambiar de base.

Sea $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ una base de \mathbb{R}^n . Se llama **matriz de cambio de base** de \mathcal{B} a la base canónica \mathcal{C} a la matriz $P \in \mathcal{M}_{n \times n}(\mathbb{R})$ cuyas columnas son los vectores de \mathcal{B} , es decir,

$$P = (u_1 | u_2 | \dots | u_n).$$

Ejemplo: Sea $\mathcal{B} = \{(1, 1, 1), (1, 2, 0), (0, 0, 1)\}$. La matriz de cambio de base de \mathcal{B} a \mathcal{C} es

$$P = P_{\mathcal{B}\mathcal{C}} = \left(\begin{array}{c|c|c} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{array} \right).$$

La propiedad que caracteriza a la matriz de cambio de base es la siguiente:

Proposición 4.5 Si $P = P_{\mathcal{B}\mathcal{C}}$ es la matriz de cambio de base de \mathcal{B} a \mathcal{C} entonces

$$P_{\mathcal{B}\mathcal{C}} x_{\mathcal{B}} = x_{\mathcal{C}}, \forall x \in \mathbb{R}^n.$$

Demostración. Sea $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ y $(\lambda_1, \lambda_2, \dots, \lambda_n)$ su vector de coordenadas respecto de \mathcal{B} . Entonces:

$$x = x_{\mathcal{C}} = \lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_n u_n = (u_1 | u_2 | \dots | u_n) \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix} = P x_{\mathcal{B}}.$$

El cambio de base de \mathcal{C} a \mathcal{B} se puede hacer utilizando la siguiente propiedad:

Proposición 4.6 Sea \mathcal{B} una base de \mathbb{R}^n . Entonces $P_{\mathcal{B}\mathcal{C}}$ es inversible y además $(P_{\mathcal{B}\mathcal{C}})^{-1} = P_{\mathcal{C}\mathcal{B}}$.

Demostración. Como las columnas de $P_{\mathcal{B}\mathcal{C}}$ son los vectores de la base \mathcal{B} , claramente son linealmente independientes y por tanto $P_{\mathcal{B}\mathcal{C}}$ es inversible.

Por otra parte, para cada $x \in \mathbb{R}^n$, se tiene:

$$P_{\mathcal{B}\mathcal{C}} x_{\mathcal{B}} = x_{\mathcal{C}} \implies x_{\mathcal{B}} = (P_{\mathcal{B}\mathcal{C}})^{-1} x_{\mathcal{C}}.$$

De aquí se deduce que $(P_{\mathcal{B}\mathcal{C}})^{-1} = P_{\mathcal{C}\mathcal{B}}$. □

Ejemplo:

La matriz de cambio de base de $\mathcal{C} = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ a $\mathcal{B} = \{(1, 1, 1), (1, 2, 0), (0, 0, 1)\}$ es

$$P_{\mathcal{C}\mathcal{B}} = (P_{\mathcal{B}\mathcal{C}})^{-1} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 2 & -1 & 0 \\ -1 & 1 & 0 \\ -2 & 1 & 1 \end{pmatrix}.$$

4.6 Definición de aplicación lineal y propiedades.

Una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ se puede identificar con la aplicación $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ definida por $L(x) = Ax$, donde $x \in \mathbb{R}^n$ es un vector columna.

Esta aplicación recibe el nombre de aplicación lineal. En general, una aplicación $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ es lineal si cumple las siguientes propiedades:

1. $L(x + y) = L(x) + L(y), \forall x, y \in \mathbb{R}^n$.
2. $L(\lambda x) = \lambda L(x), \forall \lambda \in \mathbb{R}, \forall x \in \mathbb{R}^n$.

De estas propiedades se obtiene por inducción que

$$L(\lambda_1 v_1 + \lambda_2 v_2 + \cdots + \lambda_n v_n) = \lambda_1 L(v_1) + \lambda_2 L(v_2) + \cdots + \lambda_n L(v_n),$$

para todo $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}$, y $v_1, v_2, \dots, v_n \in \mathbb{R}^n$.

En otras palabras, $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ es una aplicación lineal si la imagen de la combinación lineal de n vectores de \mathbb{R}^n es igual a la combinación lineal de las imágenes.

Matriz asociada a una aplicación lineal.

Al igual que una matriz define una aplicación lineal, veremos que una aplicación lineal $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ siempre se puede escribir en la forma $L(x) = Ax$ para una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$.

Teorema 4.1 *Sea $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ una aplicación lineal. Entonces existe una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ tal que $L(x) = Ax, \forall x \in \mathbb{R}^n$.*

Demostración. Denotemos por $C = \{e_1, e_2, \dots, e_n\}$ la base canónica de \mathbb{R}^n .

Sea $x = (x_1, x_2, \dots, x_n) = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n \in \mathbb{R}^n$. Como L es una aplicación lineal:

$$\begin{aligned} L(x) &= L(x_1 e_1 + x_2 e_2 + \cdots + x_n e_n) = x_1 L(e_1) + x_2 L(e_2) + \cdots + x_n L(e_n) = \\ &= (L(e_1) | L(e_2) | \cdots | L(e_n)) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = Ax. \end{aligned}$$

□

La matriz A del teorema anterior se llama **matriz asociada** a L . Sus columnas son las imágenes de los vectores de la base canónica. En la práctica, la matriz asociada a una aplicación lineal se puede obtener directamente.

Ejemplo: Sea $L : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ definida por $L(x, y, z) = (x + 2y - z, y + 4z)$. Entonces:

$$L(x, y, z) = \begin{pmatrix} x + 2y - z \\ y + 4z \end{pmatrix} = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 4 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix}.$$

La matriz asociada a L es

$$A = \begin{pmatrix} 1 & 2 & -1 \\ 0 & 1 & 4 \end{pmatrix} \in \mathcal{M}_{2 \times 3}(\mathbb{R}).$$

4.7 Núcleo e imagen de una aplicación lineal.

Sea $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ una aplicación lineal. Se define el **núcleo** de L como

$$\text{Ker}(L) = \{x \in \mathbb{R}^n / L(x) = 0\}.$$

Es claro que si A es la matriz asociada a L entonces $\text{Ker}(L) = \text{Ker}(A) = \{x \in \mathbb{R}^n / Ax = 0\}$.

La **imagen** de L se define como el subespacio formado por todos los vectores de \mathbb{R}^p que son imagen de algún vector de \mathbb{R}^n por la aplicación L :

$$\text{Im}(L) = \{L(x) / x \in \mathbb{R}^n\}.$$

Si $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ es una base de \mathbb{R}^n entonces $\text{Im}(L) = \langle \{L(u_1), L(u_2), \dots, L(u_n)\} \rangle$.

En particular, tomando la base canónica, se obtiene que la imagen de L está generada por las columnas de la matriz asociada. Por extensión, se llama imagen de una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ al subespacio generado por sus columnas.

Ejemplo: Se considera la aplicación lineal $L : \mathbb{R}^4 \rightarrow \mathbb{R}^3$ definida por

$$L(x, y, z, t) = (x + y + z, y - 2z + t, 2x + y + 4z - t).$$

Vamos a calcular una base de $\text{Ker}(L)$ y otra de $\text{Im}(L)$.

La matriz asociada es

$$A = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 2 & 1 & 4 & -1 \end{pmatrix}.$$

Por tanto, $\text{Ker}(L) = \text{Ker}(A) = \{x \in \mathbb{R}^4 / Ax = 0\}$. Para resolver el sistema, hacemos operaciones elementales sobre las filas de la matriz de coeficientes:

$$\begin{aligned} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 2 & 1 & 4 & -1 \end{pmatrix} &\xrightarrow{F_{31}(-2)} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & -1 & 2 & -1 \end{pmatrix} \xrightarrow{F_{32}(1)} \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix} \\ &\xrightarrow{F_{12}(-1)} \begin{pmatrix} 1 & 0 & 3 & -1 \\ 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 0 \end{pmatrix}. \end{aligned}$$

Así,

$$\begin{aligned} \text{Ker}(L) &= \left\{ (x, y, z, t) \in \mathbb{R}^4 / \begin{array}{l} x = -3z + t \\ y = 2z - t \end{array} \right\} = \{(-3z + t, 2z - t, z, t) / z, t \in \mathbb{R}\} = \\ &= \{z(-3, 2, 1, 0) + t(1, -1, 0, 1) / z, t \in \mathbb{R}\} = \langle \{(-3, 2, 1, 0), (1, -1, 0, 1)\} \rangle . \end{aligned}$$

Por tanto, $\dim(\text{Ker}(L)) = 2$ y una base de $\text{Ker}(L)$ es

$$\mathcal{B}_1 = \{(-3, 2, 1, 0), (1, -1, 0, 1)\} .$$

Por otra parte, la imagen de L está generada por las columnas de A :

$$\text{Im}(L) = \langle \{(1, 0, 2), (1, 1, 1), (1, -2, 4), (0, 1, -1)\} \rangle .$$

Para calcular una base de la imagen de L hacemos operaciones elementales para eliminar los vectores linealmente dependientes:

$$\begin{pmatrix} 1 & 0 & 2 \\ 1 & 1 & 1 \\ 1 & -2 & 4 \\ 0 & 1 & -1 \end{pmatrix} \xrightarrow{\begin{array}{l} F_{21}(-1) \\ F_{31}(-1) \end{array}} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & -2 & 2 \\ 0 & 1 & -1 \end{pmatrix} \xrightarrow{\begin{array}{l} F_{32}(2) \\ F_{42}(-1) \end{array}} \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & -1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} .$$

Por tanto, $\dim(\text{Im}(L)) = 2$ y una base de $\text{Im}(L)$ es

$$\mathcal{B}_2 = \{(1, 0, 2), (0, 1, -1)\} .$$

Inversas de aplicaciones lineales.

El siguiente resultado muestra qué aplicaciones lineales son inversibles y cómo calcular la aplicación inversa.

Proposición 4.7 *Sea $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$ una aplicación lineal y sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ su matriz asociada. Entonces L es inversible si y sólo si A es inversible. Además, la matriz asociada a L^{-1} es A^{-1} .*

Ejemplo:

Consideremos la aplicación lineal $L : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ dada por $L(x, y) = (x + y, 2x + y)$. Su matriz asociada es

$$A = \begin{pmatrix} 1 & 1 \\ 2 & 1 \end{pmatrix} .$$

Como $|A| = -1 \neq 0$, A es inversible y por tanto L es inversible.

La matriz asociada a L^{-1} es

$$A^{-1} = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} ,$$

y en consecuencia la aplicación inversa $L^{-1} : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ está definida por

$$L^{-1}(x, y) = A^{-1} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -1 & 1 \\ 2 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} -x + y \\ 2x - y \end{pmatrix} .$$

Capítulo 5

Diagonalización y funciones de matrices

5.1 Introducción.

Los conceptos principales de este capítulo son los de autovalor y autovector de una matriz cuadrada. Se introduce el polinomio característico para el cálculo de autovalores y se dan aplicaciones a la diagonalización de matrices y al cálculo de funciones de matrices.

5.2 Autovalores y autovectores.

Definición 5.1 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Un vector x es un **autovector** de A si $x \neq 0$ y existe un escalar λ tal que $Ax = \lambda x$. El escalar λ se llama **autovalor** de A asociado al autovector x .

Aunque en la mayoría de las aplicaciones que veremos este curso trabajaremos con autovalores reales y por tanto el autovector es un vector de \mathbb{R}^n , veremos que es posible que el escalar λ sea complejo. En ese caso el autovector asociado será un vector $x \in \mathbb{C}^n$.

Definición 5.2 El conjunto de todos los autovalores de una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ se llama **espectro** de A y se denota $Sp(A)$.

Ejemplo 1:

Consideremos la matriz

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \in \mathcal{M}_{3 \times 3}(\mathbb{R}).$$

Veamos que $\lambda = 3$ es un autovalor de A y $v = (1, 1, 1)$ es un autovector asociado a dicho autovalor :

$$Av = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = \begin{pmatrix} 3 \\ 3 \\ 3 \end{pmatrix} = 3 \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}.$$

Ejemplo 2:

La matriz

$$A = \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}$$

no tiene autovalores reales. Sin embargo, $\lambda = i \in \text{Sp}(A)$:

$$\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} i \\ 1 \end{pmatrix} = \begin{pmatrix} -1 \\ i \end{pmatrix} = i \begin{pmatrix} i \\ 1 \end{pmatrix}.$$

Cálculo de autovalores: polinomio característico.

La forma de calcular los autovalores de una matriz la proporciona el siguiente resultado:

Teorema 5.1 *Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y sea λ un escalar. Entonces $\lambda \in \text{Sp}(A) \iff \det(A - \lambda I) = 0$. En consecuencia, $\text{Sp}(A) = \{\lambda \in \mathbb{C} / \det(A - \lambda I) = 0\}$.*

Demostración.

Observemos que

$$Ax = \lambda x \iff Ax - \lambda x = 0 \iff (A - \lambda I)x = 0 \iff x \in \text{Ker}(A - \lambda I).$$

Por tanto,

$$\lambda \in \text{Sp}(A) \iff \text{Ker}(A - \lambda I) \neq \{0\} \iff |A - \lambda I| = 0.$$

□

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, se llama **polinomio característico** de A al polinomio definido por $q_A(x) = \det(A - xI)$. El teorema anterior dice que los autovalores de A son las raíces de su polinomio característico.

Ejemplo: Sea

$$A = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix} \in \mathcal{M}_{3 \times 3}(\mathbb{R}).$$

El polinomio característico de A es

$$q_A(x) = |A - xI| = \begin{vmatrix} 1-x & 2 \\ 2 & 1-x \end{vmatrix} = x^2 - 2x - 3.$$

Los autovalores de A son las raíces $q_A(x)$. En este caso, como

$$x^2 - 2x - 3 = 0 \iff x = \frac{2 \pm \sqrt{16}}{2},$$

los autovalores de A son $\lambda_1 = 3$, $\lambda_2 = -1$.

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ entonces su polinomio característico tiene grado exactamente n y su coeficiente principal es $(-1)^n$. Es decir,

$$q_A(x) = (-1)^n x^n + a_{n-1} x^{n-1} + \cdots + a_1 x + a_0.$$

Recordamos ahora algunas notas sobre raíces de polinomios necesarias para enunciar otros resultados sobre el polinomio característico.

Definición 5.3 Sea $p(x)$ un polinomio de grado n con coeficientes en \mathbb{R} . Se dice que λ es una raíz de $p(x)$ de multiplicidad k si existe un polinomio $p_1(x)$ tal que $p(x) = (x - \lambda)^k p_1(x)$ y $p_1(\lambda) \neq 0$.

Es bien sabido que un polinomio $p(x)$ de grado n con coeficientes reales tiene exactamente n raíces en \mathbb{C} contadas con su multiplicidad, es decir,

$$p(x) = c(x - \lambda_1)^{\alpha_1} (x - \lambda_2)^{\alpha_2} \cdots (x - \lambda_r)^{\alpha_r},$$

donde $c \in \mathbb{R}$, $\lambda_1, \lambda_2, \dots, \lambda_r \in \mathbb{C}$, $\alpha_1, \alpha_2, \dots, \alpha_r \in \mathbb{N}$ y $\alpha_1 + \alpha_2 + \cdots + \alpha_r = n$.

Definición 5.4 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y sea $\lambda \in Sp(A)$. Se llama **multiplicidad algebraica** de λ a la multiplicidad de λ como raíz de $q_A(x)$, es decir al número natural α tal que $q_A(x) = (x - \lambda)^\alpha p(x)$, $p(\lambda) \neq 0$. Se denota $m.a.(\lambda)$.

Por tanto, una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ tiene exactamente n autovalores (contados con su multiplicidad), aunque algunos de ellos pueden no ser reales.

Cálculo de autovectores. Subespacios propios.

Definición 5.5 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y sea $\lambda \in Sp(A)$. Si $\lambda \in \mathbb{R}$ entonces los autovectores asociados son vectores de \mathbb{R}^n . Se llama **subespacio propio** asociado a λ al conjunto

$$V(\lambda) = \{x \in \mathbb{R}^n / Ax = \lambda x\} = \text{Ker}(A - \lambda I).$$

Definición 5.6 Se llama **multiplicidad geométrica** de λ a la dimensión del subespacio propio $V(\lambda)$, es decir,

$$m.g.(\lambda) = \dim(V(\lambda)) = \dim(\text{Ker}(A - \lambda I)).$$

Observación: Recordemos que si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ entonces $\dim(\text{Ker}(A)) = n - \text{rg}(A)$. Por tanto,

$$m.g.(\lambda) = \dim(\text{Ker}(A - \lambda I)) = n - \text{rg}(A - \lambda I).$$

Si $\lambda \in Sp(A)$, tanto la multiplicidad algebraica como la multiplicidad geométrica de λ son al menos 1. De hecho se tiene el siguiente resultado:

Proposición 5.1 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y sea $\lambda \in Sp(A)$. Entonces $1 \leq m.g.(\lambda) \leq m.a.(\lambda) \leq n$.

Corolario 5.1 Si $\lambda \in \text{Sp}(A)$ y $\text{m.a.}(\lambda) = 1$ entonces $\text{m.g.}(\lambda) = \text{m.a.}(\lambda) = 1$.

Ejemplo:

Se considera la matriz

$$A = \begin{pmatrix} 0 & 1 & 1 \\ -1 & 1 & 0 \\ 1 & 0 & 1 \end{pmatrix}.$$

Calculamos el polinomio característico de A :

$$\begin{aligned} |A - xI| &= \begin{vmatrix} -x & 1 & 1 \\ -1 & 1-x & 0 \\ 1 & 0 & 1-x \end{vmatrix} \stackrel{F_{32}(1)}{=} \begin{vmatrix} -x & 1 & 1 \\ -1 & 1-x & 0 \\ 0 & 1-x & 1-x \end{vmatrix} = \\ & \stackrel{K_{23}(-1)}{=} \begin{vmatrix} -x & 0 & 1 \\ -1 & 1-x & 0 \\ 0 & 0 & 1-x \end{vmatrix} = (1-x) \begin{vmatrix} -x & 0 \\ -1 & 1-x \end{vmatrix} = -x(1-x)^2. \end{aligned}$$

Por tanto, $\text{Sp}(A) = \{0, 1\}$, con $\text{m.a.}(0) = 1$, $\text{m.a.}(1) = 2$.

Como $\text{m.a.}(0) = 1$, se tiene que $\text{m.g.}(0) = \text{m.a.}(0) = 1$.

A continuación calculamos la multiplicidad geométrica del autovalor $\lambda = 1$:

$$\text{m.g.}(1) = 3 - \text{rg}(A - I) = 3 - \text{rg} \begin{pmatrix} -1 & 1 & 1 \\ -1 & 0 & 0 \\ 1 & 0 & 0 \end{pmatrix} = 3 - 2 = 1.$$

Los subespacios propios asociados a 0 y 1 son:

$$V(0) = \text{Ker}(A) = \{(x, y, z) \in \mathbb{R}^3 / y = x, z = -x\} = \langle \{(1, 1, -1)\} \rangle.$$

$$V(1) = \text{Ker}(A - I) = \{(x, y, z) \in \mathbb{R}^3 / x = 0, z = -y\} = \langle \{(0, 1, -1)\} \rangle.$$

Propiedades:

1. Si $D = (d_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz diagonal entonces los autovalores de D son los elementos diagonales d_1, d_2, \dots, d_n .
2. Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $\text{Sp}(A) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ (cada autovalor aparece tantas veces como indica su multiplicidad algebraica), entonces:

- $\det(A) = \prod_{i=1}^n \lambda_i = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$
- $\text{tr}(A) = \sum_{i=1}^n \lambda_i = \lambda_1 + \lambda_2 + \cdots + \lambda_n$.

Esta propiedad es útil para comprobar si los autovalores se han calculado correctamente, ya que su suma debe coincidir con la traza de la matriz.

5.3 Matrices diagonalizables.

Definición 5.7 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Se dice que A es **diagonalizable** si existen dos matrices $P, D \in \mathcal{M}_{n \times n}(\mathbb{R})$ tales que P es inversible, D es diagonal y $A = PDP^{-1}$.

Denotemos por

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} ; \quad P = (u_1 | u_2 | \dots | u_n).$$

Obsérvese que

$$A = PDP^{-1} \iff AP = PD \iff (Au_1 | Au_2 | \dots | Au_n) = (\lambda_1 u_1 | \lambda_2 u_2 | \dots | \lambda_n u_n).$$

Esto quiere decir que si A es diagonalizable entonces los elementos diagonales de la matriz D son los autovalores de A (contados con su multiplicidad) y las columnas de la matriz P son los correspondientes autovectores asociados (en el mismo orden). Para poder construir D y P es necesario que todos los autovalores de A sean reales y que cada autovalor proporcione tantos autovectores linealmente independientes como indica su multiplicidad algebraica. En resumen, se tiene el siguiente resultado:

Teorema 5.2 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Entonces:

(a) A es diagonalizable si y sólo si todos los autovalores de A son reales y además

$$\text{m.a.}(\lambda) = \text{m.g.}(\lambda), \forall \lambda \in Sp(A).$$

(b) Si A es diagonalizable, las matrices P y D tales que $A = PDP^{-1}$ se construyen del siguiente modo:

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix} ; \quad P = (u_1 | u_2 | \dots | u_n),$$

donde $\lambda_1, \lambda_2, \dots, \lambda_n$ son los autovalores de A (contados con su multiplicidad) y u_1, u_2, \dots, u_n son los correspondientes autovectores asociados.

La diagonalización se puede aplicar al cálculo de potencias y raíces cuadradas de matrices.

Proposición 5.2 Si $A = PDP^{-1}$ entonces $A^k = PD^k P^{-1}$, $\forall k \geq 1$.

Proposición 5.3 Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es diagonalizable y todos sus autovalores son mayores o iguales que cero entonces se puede calcular una raíz cuadrada de A en la forma $A^{1/2} = PD^{1/2}P^{-1}$, donde

$$D^{1/2} = \begin{pmatrix} \sqrt{\lambda_1} & 0 & \cdots & 0 \\ 0 & \sqrt{\lambda_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sqrt{\lambda_n} \end{pmatrix}.$$

Ejemplo: Hallar una raíz cuadrada de la matriz

$$A = \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}.$$

En este caso $\text{Sp}(A) = \{0, 3\}$, con m.a. (0) = m.g. (0) = 2. Además,

$$\text{Ker}(A) = \langle \{(1, 0, -1), (0, 1, -1)\} \rangle, \quad \text{Ker}(A - 3I) = \langle \{(1, 1, 1)\} \rangle.$$

Por tanto, podemos tomar

$$D = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & -1 & 1 \end{pmatrix},$$

de tal forma que $A = PDP^{-1}$. De este modo, la matriz

$$\begin{aligned} B = PD^{1/2}P^{-1} &= \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & \sqrt{3} \end{pmatrix} \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} = \\ &= \frac{1}{\sqrt{3}} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{aligned}$$

es una raíz cuadrada de A (es decir, $B^2 = A$).

5.4 Teorema de Cayley-Hamilton.

Polinomios de matrices.

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Sea $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_kx^k$. Se define

$$p(A) = a_0I + a_1A + a_2A^2 + \cdots + a_kA^k \in \mathcal{M}_{n \times n}(\mathbb{R}).$$

Diremos que $p(x)$ es un polinomio anulador de A si $p(A)$ es la matriz cero.

Ejemplo: El polinomio $p(x) = x^2 - 2x$ es un polinomio anulador de la matriz

$$A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}.$$

En efecto,

$$p(A) = A^2 - 2A = \begin{pmatrix} 2 & 2 \\ 2 & 2 \end{pmatrix} - 2 \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

Teorema 5.3 (Teorema de Cayley-Hamilton) Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $q_A(x)$ su polinomio característico. Entonces $q_A(A) = 0$, es decir, $q_A(x)$ es un polinomio anulador de A .

Del teorema de Cayley-Hamilton se deduce que para calcular cualquier polinomio de una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es suficiente calcular las $(n - 1)$ primeras potencias de A .

Corolario 5.2 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Si $p(x)$ es un polinomio de grado $k \geq n$ entonces existe un polinomio $r(x)$ de grado menor que n tal que $p(A) = r(A)$.

Demostración. Dividiendo $p(x)$ entre $q_A(x)$, se tiene que $p(x) = q_A(x)d(x) + r(x)$, donde el resto $r(x)$ tiene grado menor que n . Utilizando el teorema de Cayley-Hamilton:

$$p(A) = \underbrace{q_A(A)}_0 d(A) + r(A) = r(A).$$

□

Para calcular $r(x)$ no es necesario efectuar la división. Observemos que si λ es un autovalor de A entonces $p(\lambda) = q_A(\lambda)d(\lambda) + r(\lambda) = r(\lambda)$, ya que $q_A(\lambda) = 0$. Es decir, los polinomios $p(x)$ y $r(x)$ deben tomar el mismo valor sobre todos los autovalores de A . Del mismo modo, si la multiplicidad algebraica de λ es m entonces

$$p^{(k)}(\lambda) = r^{(k)}(\lambda), \quad \forall \lambda \in \text{Sp}(A), \quad \forall k = 1, 2, \dots, m - 1.$$

Esta propiedad permite calcular $r(x)$ resolviendo un sistema de ecuaciones lineales.

Ejemplo: Calcular un polinomio $r(x)$ de grado 1 tal que $r(A) = p(A)$, donde $p(x) = x^{10} - 2x + 1$ y

$$A = \begin{pmatrix} -1 & 1 \\ -2 & 2 \end{pmatrix}.$$

Como los autovalores de A son $\lambda_1 = 0$, $\lambda_2 = 1$, el polinomio $r(x) = a + bx$ de grado 1 debe cumplir las relaciones:

$$\begin{aligned} r(0) &= a = p(0) = 1 \\ r(1) &= a + b = p(1) = 0. \end{aligned}$$

Por tanto $a = 1$, $b = -1$ y $r(x) = 1 - x$.

Finalmente,

$$p(A) = r(A) = I - A = \begin{pmatrix} 2 & -1 \\ 2 & -1 \end{pmatrix}.$$

5.5 Funciones de matrices.

En esta sección usaremos la idea anterior para obtener funciones de matrices para una clase de funciones más general que los polinomios. En concreto, consideraremos funciones analíticas, entre las cuales están las funciones racionales, las raíces k -ésimas, la exponencial, el logaritmo y las funciones trigonométricas más comunes. Con ayuda de la última observación se pueden calcular estas funciones de matrices como combinaciones lineales de las $n - 1$ primeras potencias de A .

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y sea $f : D \rightarrow \mathbb{R}$ una función analítica definida en un dominio real D . Supongamos que para cada autovalor λ de A están definidos los valores $f^{(k)}(\lambda)$ para todo $k = 0, 1, \dots, m - 1$, donde $m = \text{m.a.}(\lambda)$, $f^{(0)}(\lambda) = f(\lambda)$. Entonces es posible encontrar un polinomio $r(x) = a_0 + a_1x + \dots + a_{n-1}x^{n-1}$ de grado menor que n tal que

$$f^{(k)}(\lambda) = r^{(k)}(\lambda), \quad \forall \lambda \in \text{Sp}(A), \quad \forall k = 0, 1, \dots, \text{m.a.}(\lambda) - 1.$$

Denotaremos $V_{f,A} = \{f^{(k)}(\lambda) / \lambda \in \text{Sp}(A), k = 0, 1, \dots, \text{m.a.}(\lambda) - 1\}$.

Definición 5.8 Sean $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y f una función de tal forma que existen todos los valores del conjunto $V_{f,A}$. Entonces diremos que f está definida sobre A y se define $f(A)$ como el valor del polinomio $r(x)$ en A , es decir,

$$f(A) = r(A) = a_0I + a_1A + \dots + a_{n-1}A^{n-1}.$$

Obsérvese que los n coeficientes a_i de $r(x)$ se determinan resolviendo un sistema de n ecuaciones lineales con n incógnitas.

Ejemplo 1: Se consideran la función $f(x) = e^x$ y la matriz

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}.$$

En este caso $\text{Sp}(A) = \{0\}$, con $\text{m.a.}(0) = 3$. Entonces existe un polinomio $r(x) = a + bx + cx^2$ de grado menor o igual que dos tal que

$$\begin{aligned} r(0) &= a = f(0) = e^0 = 1 \\ r'(0) &= b = f'(0) = 1 \\ r''(0) &= 2c = f''(0) = 1. \end{aligned}$$

Por tanto $a = 1$, $b = 1$, $c = 1/2$ y $r(x) = 1 + x + (1/2)x^2$.

Finalmente,

$$\begin{aligned} e^A = f(A) = r(A) &= I + A + \frac{1}{2}A^2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} + \begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix} + \frac{1}{2} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} = \\ &= \begin{pmatrix} 1 & 1 & 3/2 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{pmatrix}. \end{aligned}$$

Ejemplo 2: No es posible calcular una raíz cuadrada de la matriz

$$A = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}.$$

En efecto, consideremos la función $f(x) = \sqrt{x} = x^{1/2}$. Como $\text{Sp}(A) = \{0\}$ con m.a. $(0) = 2$, para calcular $f(A) = A^{1/2}$ necesitamos determinar los valores de $f(0)$ y $f'(0)$.

Pero no existe $f'(0)$ ya que $f'(x) = 1/(2\sqrt{x})$.

Observación: La condición de que existan todos los valores del conjunto $V_{f,A}$ no siempre es necesaria para definir $f(A)$. Por ejemplo, aunque

$$B = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$$

también tiene $\text{Sp}(B) = \{0\}$ con m.a. $(0) = 2$, es posible calcular una raíz cuadrada de B (por ejemplo, la propia B).

Autovalores de $f(A)$.

Los autovalores de la matriz $f(A)$ se pueden obtener sin calcularla explícitamente.

Proposición 5.4 Si $\lambda_1, \lambda_2, \dots, \lambda_n$ son los autovalores de A (contados con su multiplicidad) entonces los autovalores de $f(A)$ son $f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)$.

Casos particulares:

1. $\text{Sp}(A^k) = \{\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k\}, \forall k \in \mathbb{N}$.
2. $\text{Sp}(A^{-1}) = \{1/\lambda_1, 1/\lambda_2, \dots, 1/\lambda_n\}$. (Si A es inversible).

En particular, la proposición 5.4 permite obtener el determinante y la traza de $f(A)$ sin calcular la función de la matriz. Si $\lambda_1, \lambda_2, \dots, \lambda_n$ son los autovalores de A contados con su multiplicidad, entonces:

$$\begin{aligned} \det(f(A)) &= f(\lambda_1)f(\lambda_2) \cdots f(\lambda_n) \\ \text{tr}(f(A)) &= f(\lambda_1) + f(\lambda_2) + \cdots + f(\lambda_n). \end{aligned}$$

Funciones de matrices usando la diagonalización.

El siguiente resultado es consecuencia de la forma que tienen las potencias de las matrices diagonales:

Proposición 5.5 *Si D es diagonal,*

$$D = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & \lambda_n \end{pmatrix},$$

y f es una función definida sobre D entonces

$$f(D) = \begin{pmatrix} f(\lambda_1) & 0 & \cdots & 0 \\ 0 & f(\lambda_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & f(\lambda_n) \end{pmatrix}.$$

Ejemplo: Si $f(x) = e^x$ y

$$A = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$$

entonces

$$e^A = f(A) = \begin{pmatrix} f(0) & 0 & 0 \\ 0 & f(0) & 0 \\ 0 & 0 & f(0) \end{pmatrix} = \begin{pmatrix} e^0 & 0 & 0 \\ 0 & e^0 & 0 \\ 0 & 0 & e^0 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Este resultado proporciona una forma alternativa para calcular funciones de matrices cuando A es diagonalizable:

Proposición 5.6 *Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es diagonalizable, es decir, $A = PDP^{-1}$ con D diagonal, entonces $f(A) = Pf(D)P^{-1}$.*

Capítulo 6

Espacios euclídeos

6.1 Introducción.

En este tema se introduce el producto escalar y algunos conceptos importantes asociados a él, como la norma y la ortogonalidad. Esto permite desarrollar nuevas aplicaciones del álgebra matricial, como la diagonalización ortogonal, la descomposición en valores singulares, la clasificación de formas cuadráticas y el método de mínimos cuadrados para obtener soluciones aproximadas de sistemas de ecuaciones lineales sobredeterminados.

6.2 Espacios vectoriales con producto escalar.

Definición 6.1 Sea V un espacio vectorial real. Una aplicación $\langle \cdot, \cdot \rangle : V \times V \rightarrow \mathbb{R}$ es un **producto escalar** o **producto interior** si cumple las siguientes propiedades:

- 1) $\langle x_1 + x_2, y \rangle = \langle x_1, y \rangle + \langle x_2, y \rangle, \forall x_1, x_2, y \in V.$
- 2) $\langle \lambda x, y \rangle = \lambda \langle x, y \rangle, \forall x, y \in V, \lambda \in \mathbb{R}.$
- 3) $\langle x, y \rangle = \langle y, x \rangle, \forall x, y \in V.$
- 4) $\langle x, x \rangle > 0, \forall x \in V, x \neq \theta.$

Proposición 6.1 De las propiedades (1)-(4) se deducen las siguientes:

- 5) $\langle x, y_1 + y_2 \rangle = \langle x, y_1 \rangle + \langle x, y_2 \rangle, \forall x, y_1, y_2 \in V.$
- 6) $\langle x, \lambda y \rangle = \lambda \langle x, y \rangle, \forall x, y \in V, \lambda \in \mathbb{R}.$
- 7) $\langle x, \theta \rangle = 0, \forall x \in V.$

Observación: Las propiedades (1), (2), (5) y (6) se resumen diciendo que el producto escalar real es una forma bilineal; la propiedad (3) dice que es simétrico y la propiedad (4) que es definido positivo.

Norma inducida.

El producto escalar en un espacio vectorial V permite definir una norma:

Proposición 6.2 *Sea V un espacio vectorial con producto escalar. La aplicación $\|\cdot\| : V \rightarrow \mathbb{R}$ dada por*

$$\|v\| = +\sqrt{\langle v, v \rangle}, \quad \forall v \in V$$

define una norma sobre V ; esto es, cumple las siguientes propiedades:

- i) $\|x\| > 0, \forall x \in V \setminus \{\theta\}$ y $\|\theta\| = 0$.*
- ii) $\|\lambda x\| = |\lambda| \cdot \|x\|, \forall x \in V, \forall \lambda \in \mathbb{R}$.*
- iii) $\|x + y\| \leq \|x\| + \|y\|, \forall x, y \in V$.*

Recuérdese que si x, y son dos vectores de V entonces $\|x - y\|$ representa la distancia de x a y . En particular, la norma de x representa su distancia a cero.

Definición 6.2 *Los espacios vectoriales con producto escalar dotados de la norma inducida se llaman **espacios euclídeos**.*

Ejemplos:**I. Producto escalar usual de \mathbb{R}^n .**

Se define el *producto escalar usual* en \mathbb{R}^n como

$$\langle x, y \rangle = x^t y = \sum_{i=1}^n x_i y_i, \quad \forall x, y \in \mathbb{R}^n.$$

La norma asociada a $\langle \cdot, \cdot \rangle$ coincide con la norma usual de \mathbb{R}^n :

$$\|x\| = +\sqrt{x^t x} = +\sqrt{x_1^2 + x_2^2 + \cdots + x_n^2}, \quad \forall x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n.$$

II. Producto escalar de funciones. Sea V el espacio vectorial de las funciones reales continuas definidas en $[-1, 1]$. Se suele usar el siguiente producto escalar:

$$\langle f, g \rangle = \int_{-1}^1 f(x)g(x) dx.$$

La norma inducida es

$$\|f\| = \left(\int_{-1}^1 f^2(x) dx \right)^{1/2}.$$

6.3 Ortogonalidad

Definición 6.3 Sea V un espacio vectorial con producto escalar.

1. Se dice que dos vectores x e y son ortogonales si $\langle x, y \rangle = 0$.
2. Un vector x es ortogonal a un subespacio U de V si x es ortogonal a todos los vectores de U .
3. Un conjunto de vectores $\mathcal{S} = \{v_1, v_2, \dots, v_k\}$ de V es ortogonal si $\langle v_i, v_j \rangle = 0$, $\forall i \neq j$.
4. Un conjunto de vectores $\mathcal{S} = \{v_1, v_2, \dots, v_k\}$ de V es ortonormal si es ortogonal y $\|v_i\| = 1$, $\forall i = 1, 2, \dots, k$.

Observación: Los vectores de norma uno se llaman vectores unitarios. De cada vector v distinto de cero se puede obtener un vector unitario con su misma dirección y sentido sin más que dividir por su norma.

Ortonormalización.

El siguiente resultado muestra cómo obtener un conjunto ortonormal de un conjunto libre:

Teorema 6.1 (Ortonormalización de Gram-Schmidt) Sea V un espacio vectorial con un producto interior y sea $S = \{v_1, v_2, \dots, v_p\}$ un conjunto libre de vectores de V . Existe un conjunto ortonormal $T = \{u_1, u_2, \dots, u_p\}$ tal que $\langle S \rangle = \langle T \rangle$. Es más,

$$\langle \{v_1, \dots, v_k\} \rangle = \langle \{u_1, \dots, u_k\} \rangle, \quad \forall k = 1, \dots, p.$$

Descripción del proceso de ortonormalización de Gram-Schmidt.

Sea $S = \{v_1, v_2, \dots, v_p\}$ el conjunto libre de vectores de V que queremos ortonormalizar. Se procede del siguiente modo:

- (1) Se construye u_1 dividiendo v_1 por su norma:

$$u_1 = \frac{1}{\|v_1\|} v_1.$$

- (2) Para cada $i \geq 2$ se construye u_i en dos etapas:

- (2.1) Se calcula un vector \tilde{u}_i dado por:

$$\tilde{u}_i = v_i - \sum_{j=1}^{i-1} \langle v_i, u_j \rangle u_j = v_i - \langle v_i, u_1 \rangle u_1 - \dots - \langle v_i, u_{i-1} \rangle u_{i-1}.$$

(2.2) Se normaliza el vector \tilde{u}_i :

$$u_i = \frac{1}{\|\tilde{u}_i\|} \tilde{u}_i.$$

Ejemplo:

Vamos a ortonormalizar el subconjunto $S = \{(1, 0, 1), (1, 1, 1)\}$ de \mathbb{R}^3 .

Denotemos por $v_1 = (1, 0, 1)$, $v_2 = (1, 1, 1)$. Entonces:

$$u_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{2}}(1, 0, 1) = \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right);$$

$$\tilde{u}_2 = v_2 - \langle v_2, u_1 \rangle u_1 = (1, 1, 1) - \frac{2}{\sqrt{2}} \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right) = (1, 1, 1) - (1, 0, 1) = (0, 1, 0);$$

$$u_2 = \frac{\tilde{u}_2}{\|\tilde{u}_2\|} = (0, 1, 0).$$

El conjunto $T = \{u_1, u_2\} = \left\{ \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right), (0, 1, 0) \right\}$ es ortonormal y genera el mismo subespacio vectorial que S .

6.4 Proyección ortogonal.

Sea $x \in \mathbb{R}^n$ y sea U un subespacio de \mathbb{R}^n con $\dim(U) = p < n$. Se llama **proyección ortogonal** de x sobre el subespacio U al único vector $u_x \in U$ tal que $(x - u_x)$ es ortogonal a U . El vector $v_x = x - u_x$ se llama **componente normal** de x respecto a U y su norma representa la **mínima distancia** de x al subespacio U , es decir, $d(x, U) = \|x - u_x\|$.

Cálculo de la proyección ortogonal.

Proposición 6.3 *Sea U un subespacio vectorial de \mathbb{R}^n y $\mathcal{B} = \{u_1, \dots, u_p\}$ una base ortonormal de U . Entonces la proyección ortogonal de un vector x sobre U es*

$$u_x = u_1 u_1^t x + u_2 u_2^t x + \dots + u_p u_p^t x = Px,$$

donde

$$P = u_1 u_1^t + u_2 u_2^t + \dots + u_p u_p^t = (u_1 | u_2 | \dots | u_p) \begin{pmatrix} u_1^t \\ u_2^t \\ \vdots \\ u_p^t \end{pmatrix} \in \mathcal{M}_{n \times n}(\mathbb{R})$$

se llama *matriz de proyección ortogonal*.

Demostración. En primer lugar, $u_x = u_1(u_1^t x) + u_2(u_2^t x) + \dots + u_p(u_p^t x) \in U$ por ser combinación lineal de vectores de una base de U .

Por otra parte, $(x - u_x)$ es ortogonal a U ya que es ortogonal a los vectores de la base \mathcal{B} . Por ejemplo, usando que \mathcal{B} es ortonormal, se tiene:

$$u_1^t u_x = u_1^t (u_1 u_1^t x + u_2 u_2^t x + \cdots + u_p u_p^t x) = (u_1^t u_1) u_1^t x + (u_1^t u_2) u_2^t x + \cdots + (u_1^t u_p) u_p^t x = u_1^t x.$$

Por tanto, $u_1^t (x - u_x) = u_1^t x - u_1^t u_x = 0$.

Del mismo modo se prueba para u_2, \dots, u_p □

Ejemplo: Hallar la matriz de proyección ortogonal sobre el subespacio

$$U = \{(x, y, z) \in \mathbb{R}^3 / x + y - z = 0\}.$$

En primer lugar, calculamos una base de U :

$$U = \{(x, y, z) \in \mathbb{R}^3 / x + y - z = 0\} = \{(x, y, x + y) / x, y \in \mathbb{R}\} = \langle \{(1, 0, 1), (0, 1, 1)\} \rangle.$$

Una base de U es $\mathcal{B}_U = \{(1, 0, 1), (0, 1, 1)\}$.

Aplicamos el proceso de Gram-Schmidt a los vectores $v_1 = (1, 0, 1)$, $v_2 = (0, 1, 1)$ para obtener una base ortonormal $\mathcal{B}_U = \{u_1, u_2\}$ de U :

$$u_1 = \frac{v_1}{\|v_1\|} = \frac{1}{\sqrt{2}} \begin{pmatrix} 1 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{pmatrix};$$

$$\tilde{u}_2 = v_2 - \langle v_2, u_1 \rangle u_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 1 \\ 1/2 \end{pmatrix};$$

$$u_2 = \frac{\tilde{u}_2}{\|\tilde{u}_2\|} = \begin{pmatrix} -1/\sqrt{6} \\ 2/\sqrt{6} \\ 1/\sqrt{6} \end{pmatrix}.$$

La matriz de proyección ortogonal sobre U es:

$$P = u_1 u_1^t + u_2 u_2^t = (u_1 | u_2) \begin{pmatrix} u_1^t \\ u_2^t \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ -1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} =$$

$$= \begin{pmatrix} 1/2 + 1/6 & 0 - 2/6 & 1/2 - 1/6 \\ 0 - 2/6 & 0 + 4/6 & 0 + 2/6 \\ 1/2 - 1/6 & 0 + 2/6 & 1/2 + 1/6 \end{pmatrix} = \begin{pmatrix} 2/3 & -1/3 & 1/3 \\ -1/3 & 2/3 & 1/3 \\ 1/3 & 1/3 & 2/3 \end{pmatrix}.$$

Caso particular:

Sea u un vector unitario y sea $U = \langle \{u\} \rangle$. La matriz de proyección ortogonal sobre U es $P_U = uu^t$. Es fácil comprobar que P_U tiene rango 1, ya que todas sus filas son múltiplos de u . En el caso general, el rango de P_U coincide con la dimensión de U .

Ejemplo: Construir la matriz de proyección ortogonal sobre $W = \langle \{(2, 2, 1)\} \rangle$.

Para ello calculamos un vector unitario u en la dirección de $v = (2, 2, 1)$ dividiendo por su norma:

$$u = \frac{v}{\|v\|} = \begin{pmatrix} 2/3 \\ 2/3 \\ 1/3 \end{pmatrix}.$$

Por tanto,

$$P = uu^t = \begin{pmatrix} 2/3 \\ 2/3 \\ 1/3 \end{pmatrix} (2/3, 2/3, 1/3) = \frac{1}{9} \begin{pmatrix} 4 & 4 & 2 \\ 4 & 4 & 2 \\ 2 & 2 & 1 \end{pmatrix}.$$

6.5 Diagonalización ortogonal.

Recordemos que una matriz $P \in \mathcal{M}_{n \times n}(\mathbb{R})$ es ortogonal si $P^{-1} = P^t$, es decir $P^t P = I$.

Definición 6.4 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Se dice que A es ortogonalmente diagonalizable si existen dos matrices $P, D \in \mathcal{M}_{n \times n}(\mathbb{R})$ tales que P es ortogonal, D es diagonal y $A = PDP^t$. En tal caso, se dice que la descomposición $A = PDP^t$ es una diagonalización ortogonal de A .

Teorema 6.2 (Teorema espectral para matrices simétricas) Una matriz real $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es ortogonalmente diagonalizable si y sólo si A es simétrica.

Descomposición Espectral.

Sea $A = PDP^t$ la diagonalización ortogonal de una matriz simétrica A de rango r . Sean $\lambda_1, \lambda_2, \dots, \lambda_r$ sus autovalores no nulos, contados con su multiplicidad. Si u_1, u_2, \dots, u_n son las columnas de P entonces, usando el producto de matrices por bloques, se tiene:

$$\begin{aligned} A = PDP^t &= (u_1 | u_2 | \dots | u_n) \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \begin{pmatrix} u_1^t \\ u_2^t \\ \vdots \\ u_n^t \end{pmatrix} = \\ &= \lambda_1 u_1 u_1^t + \lambda_2 u_2 u_2^t + \dots + \lambda_n u_n u_n^t = \lambda_1 u_1 u_1^t + \lambda_2 u_2 u_2^t + \dots + \lambda_r u_r u_r^t, \end{aligned}$$

ya que $\lambda_{r+1} = \dots = \lambda_n = 0$.

De esta manera se descompone A en la suma de r matrices $u_i u_i^t$ de rango uno. Esta descomposición se llama **descomposición espectral de A** . Obsérvese que cada sumando es el producto de un autovalor por la matriz de proyección sobre el subespacio generado por el autovector correspondiente.

Cálculo de la diagonalización ortogonal de una matriz simétrica.

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica. Veamos cómo construir las matrices P y D tales que $A = PDP^t$.

La matriz D se construye en la forma habitual, es decir, es una matriz diagonal cuyos elementos diagonales son los autovalores de A , repetidos un número de veces igual a su multiplicidad algebraica. Una observación importante es que **todos los autovalores de una matriz simétrica son reales**.

Como $A = PDP^t = PDP^{-1}$, las columnas de la matriz P deben ser autovectores de A . Necesitamos además que P sea ortogonal. La siguiente caracterización de las matrices ortogonales será útil:

Proposición 6.4 *Una matriz $P \in \mathcal{M}_{n \times n}(\mathbb{R})$ es ortogonal si y sólo si sus columnas son una base ortonormal de \mathbb{R}^n (respecto al producto escalar usual).*

Demostración. Denotemos por u_1, u_2, \dots, u_n las columnas de P . Dado que $\text{rg}(P) = n$, el conjunto $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ es una base de \mathbb{R}^n . Además,

$$P^t P = \begin{pmatrix} \frac{u_1^t}{|u_1|} \\ \frac{u_2^t}{|u_2|} \\ \vdots \\ \frac{u_n^t}{|u_n|} \end{pmatrix} (|u_1| |u_2| \cdots |u_n|) = I \iff \left\{ \begin{array}{l} u_i^t u_j = 0, \text{ si } i \neq j \\ u_i^t u_i = 1, \forall i = 1, 2, \dots, n \end{array} \right\} \iff \mathcal{B} \text{ es ortonormal.}$$

□

En virtud de la Proposición 6.4, necesitamos conseguir una base de autovectores de A que además sea ortonormal. La siguiente propiedad hace que esto sea posible:

Lema 6.1 *Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica. Si x_1 y x_2 son autovectores asociados a dos autovalores distintos de A entonces x_1 y x_2 son ortogonales.*

Demostración. Sean $\lambda_1 \neq \lambda_2$ dos autovalores de A y sean $x_1 \in V(\lambda_1)$, $x_2 \in V(\lambda_2)$. Teniendo en cuenta que $A = A^t$ y $\lambda_1, \lambda_2 \in \mathbb{R}$:

$$\lambda_1 \langle x_1, x_2 \rangle = \langle \lambda_1 x_1, x_2 \rangle = \langle Ax_1, x_2 \rangle = (Ax_1)^t x_2 = x_1^t A^t x_2 = x_1^t Ax_2 = x_1^t \lambda_2 x_2 = \lambda_2 \langle x_1, x_2 \rangle.$$

Por tanto, $\lambda_1 \langle x_1, x_2 \rangle = \lambda_2 \langle x_1, x_2 \rangle$. Como $\lambda_1 \neq \lambda_2$, necesariamente $\langle x_1, x_2 \rangle = 0$. □

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica. Teniendo en cuenta las propiedades anteriores, los pasos para calcular una diagonalización ortogonal $A = PDP^t$ son los siguientes:

- (1) Se calculan los autovalores de A . Los elementos diagonales de la matriz D son los autovalores de A (repetidos tantas veces como indica su multiplicidad algebraica).
- (2) Para cada autovalor $\lambda \in \text{Sp}(A)$ se halla una base del subespacio propio asociado $V(\lambda)$ y se le aplica el proceso de ortonormalización de Gram-Schmidt para obtener una base ortonormal de $V(\lambda)$.

- (3) La matriz P es la que tiene por columnas los elementos de las bases ortonormales de $V(\lambda_1)$, $V(\lambda_2), \dots, V(\lambda_k)$ (donde $\lambda_1, \lambda_2, \dots, \lambda_k$ son los autovalores distintos de A) colocadas en el mismo orden que ocupan los correspondientes autovalores en la diagonal de D .

Ejemplo:

Hallar una diagonalización ortogonal de la matriz $A \in \mathcal{M}_{4 \times 4}(\mathbb{R})$ dada por

$$A = \begin{pmatrix} -2 & -1 & 0 & 2 \\ -1 & -2 & 0 & -2 \\ 0 & 0 & -3 & 0 \\ 2 & -2 & 0 & 1 \end{pmatrix}$$

Dado que A es una matriz simétrica real, es ortogonalmente diagonalizable, es decir, existen dos matrices $P, D \in \mathcal{M}_{4 \times 4}(\mathbb{R})$ tales que P es ortogonal, D es diagonal y $A = PDP^t$. La matriz diagonal D tiene como elementos diagonales los autovalores de A .

El polinomio característico de A es $q_A(x) = (-3 - x)^3(3 - x)$ (hágase como ejercicio).

Por tanto los autovalores de A son $\lambda_1 = -3$ y $\lambda_2 = 3$, con m.a.(-3)=3, m.a.(3)=1, y la matriz D es

$$D = \begin{pmatrix} -3 & 0 & 0 & 0 \\ 0 & -3 & 0 & 0 \\ 0 & 0 & -3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

Los vectores columna de la matriz ortogonal $P = (u_1|u_2|u_3|u_4)$ constituyen una base ortonormal de \mathbb{R}^4 formada por autovectores de A . Para determinarlos, aplicaremos el procedimiento de ortonormalización de Gram-Schmidt a sendas bases de los subespacios propios asociados a $\lambda_1 = -3$ y $\lambda_2 = 3$.

Resolviendo el correspondiente sistema homogéneo, se tiene:

$$\text{Ker}(A + 3I) = \langle \{(1, 1, 0, 0), (0, 2, 0, 1), (0, 0, 1, 0)\} \rangle.$$

Si denotamos $v_1 = (1, 1, 0, 0)$, $v_2 = (0, 2, 0, 1)$, $v_3 = (0, 0, 1, 0)$ entonces los tres primeros vectores columna u_1, u_2, u_3 de la matriz P se calculan del siguiente modo:

$$u_1 = \frac{v_1}{\|v_1\|} = (1/\sqrt{2}, 1/\sqrt{2}, 0, 0)$$

$$\tilde{u}_2 = v_2 - \langle v_2, u_1 \rangle u_1 = (-1, 1, 0, 1) \quad ; \quad u_2 = \frac{\tilde{u}_2}{\|\tilde{u}_2\|} = (-1/\sqrt{3}, 1/\sqrt{3}, 0, 1/\sqrt{3})$$

$$\tilde{u}_3 = v_3 - \langle v_3, u_1 \rangle u_1 - \langle v_3, u_2 \rangle u_2 = (0, 0, 1, 0) \quad ; \quad u_3 = \frac{\tilde{u}_3}{\|\tilde{u}_3\|} = (0, 0, 1, 0).$$

Del mismo modo,

$$\text{Ker}(A - 3I) = \langle \{(1, -1, 0, 2)\} \rangle = \langle \{v_4\} \rangle,$$

de modo que el vector columna u_4 de P viene dado por

$$u_4 = \frac{v_4}{\|v_4\|} = (1/\sqrt{6}, -1/\sqrt{6}, 0, 2/\sqrt{6}).$$

Así, la matriz ortogonal

$$P = (u_1|u_2|u_3|u_4) = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{3} & 0 & 1/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & 0 & -1/\sqrt{6} \\ 0 & 0 & 1 & 0 \\ 0 & 1/\sqrt{3} & 0 & 2/\sqrt{6} \end{pmatrix}$$

cumple que $A = PDP^t$.

La descomposición espectral de A es

$$A = -3u_1u_1^t - 3u_2u_2^t - 3u_3u_3^t + 3u_4u_4^t.$$

En particular, la matriz A se puede escribir como $A = -3P_1 + 3P_2$, donde

$$P_1 = u_1u_1^t + u_2u_2^t + u_3u_3^t = \begin{pmatrix} 5/6 & 1/6 & 0 & -1/3 \\ 1/6 & 5/6 & 0 & 1/3 \\ 0 & 0 & 1 & 0 \\ -1/3 & 1/3 & 0 & 1/3 \end{pmatrix};$$

$$P_2 = u_4u_4^t = \begin{pmatrix} 1/6 & -1/6 & 0 & 1/3 \\ -1/6 & 1/6 & 0 & -1/3 \\ 0 & 0 & 0 & 0 \\ 1/3 & -1/3 & 0 & 2/3 \end{pmatrix}$$

son las matrices de proyección sobre los subespacios propios $V(-3)$ y $V(3)$ respectivamente.

6.6 Formas cuadráticas sobre \mathbb{R}^n . Clasificación.

Definición 6.5 Una forma cuadrática sobre \mathbb{R}^n es una aplicación $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por

$$\omega(x) = x^tAx, \quad \forall x \in \mathbb{R}^n,$$

donde $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz simétrica.

Observación: Si $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ entonces la forma cuadrática $\omega(x) = x^tAx$ se expresa como:

$$\omega(x_1, x_2, \dots, x_n) = (x_1, x_2, \dots, x_n) \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = \sum_{i,j=1}^n a_{ij}x_ix_j.$$

Recíprocamente, si tenemos una expresión extendida de la forma cuadrática como la anterior, podemos encontrar una única matriz simétrica $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ tal que $\omega(x) = x^t Ax$, $\forall x \in \mathbb{R}^n$.

Ejemplo:

Sea $\omega(x_1, x_2, x_3) = 2x_1^2 + 3x_2^2 + x_3^2 - 4x_1x_2 + 2x_1x_3 - 2x_2x_3$. Entonces:

$$\omega(x_1, x_2, x_3) = (x_1, x_2, x_3) \begin{pmatrix} 2 & -2 & 1 \\ -2 & 3 & -1 \\ 1 & -1 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix} = x^t Ax.$$

Clasificación de formas cuadráticas.

Definición 6.6 Sea $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ una forma cuadrática. Diremos que

1. ω es definida positiva si $\omega(x) > 0$, $\forall x \neq 0$,
2. ω es definida negativa si $\omega(x) < 0$, $\forall x \neq 0$,
3. ω es semidefinida positiva si $\omega(x) \geq 0$, $\forall x \in \mathbb{R}^n$,
4. ω es semidefinida negativa si $\omega(x) \leq 0$, $\forall x \in \mathbb{R}^n$,
5. ω es indefinida en cualquier otro caso, es decir, si existen dos vectores $x, y \in \mathbb{R}^n$ tales que $\omega(x) > 0$, $\omega(y) < 0$.

Definición 6.7 Una matriz simétrica $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ se dice definida positiva, definida negativa, semidefinida positiva, semidefinida negativa o indefinida según lo sea la forma cuadrática $\omega_A : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por $\omega_A(x) = x^t Ax$.

Formas cuadráticas degeneradas y no degeneradas

Definición 6.8 Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica y sea $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ la forma cuadrática definida por $\omega(x) = x^t Ax$, $\forall x \in \mathbb{R}^n$. Se dice que ω es **no degenerada** si $\text{rg}(A) = n$. En otro caso se dice que ω es degenerada. Obsérvese que ω es no degenerada si y sólo si $|A| \neq 0$.

Las formas cuadráticas definidas positivas y definidas negativas son siempre no degeneradas, mientras que las semidefinidas son degeneradas. Las formas cuadráticas indefinidas pueden ser degeneradas o no degeneradas, de modo que debe indicarse este aspecto en su clasificación.

Ejemplos:

1. $\omega(x, y, z) = x^2 + y^2 + z^2$ es definida positiva ya que $x^2 + y^2 + z^2 \geq 0$, $\forall (x, y, z) \in \mathbb{R}^3$ y además $x^2 + y^2 + z^2 = 0 \iff x = y = z = 0$.

2. $\omega(x, y, z) = x^2 + y^2 - z^2$ es indefinida ya que, por ejemplo, $\omega(1, 0, 0) = 1 > 0$ y $\omega(0, 0, 1) = -1 < 0$. Además es no degenerada ya que

$$\omega(x, y, z) = (x, y, z) \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = x^t A x,$$

con $|A| = -1 \neq 0$.

Sin embargo, en general es difícil determinar la clasificación de ω si aparecen “términos cruzados”. Por ejemplo, la forma cuadrática

$$\omega(x_1, x_2, x_3) = 2x_1^2 + 3x_2^2 + x_3^2 - 4x_1x_2 + 2x_1x_3 - 2x_2x_3$$

es definida positiva, pero no es inmediato deducirlo a simple vista.

Uso de la diagonalización ortogonal.

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica. Recordemos que A es ortogonalmente diagonalizable, es decir, existen dos matrices $P, D \in \mathcal{M}_{n \times n}(\mathbb{R})$ tales que D es diagonal, P es ortogonal y $A = PDP^t$.

Sea $x \in \mathbb{R}^n$. Entonces:

$$\omega(x) = x^t A x = x^t P D P^t x = (P^t x)^t D (P^t x).$$

Si denotamos $y = P^t x$ entonces la forma cuadrática se escribe en la nueva variable como

$$\omega(y) = y^t D y = \sum_{i=1}^n \lambda_i y_i^2,$$

donde $y = (y_1, y_2, \dots, y_n)$ y $\lambda_1, \lambda_2, \dots, \lambda_n$ son los autovalores de A contados con su multiplicidad.

De aquí se deduce el siguiente resultado:

Teorema 6.3 *Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica. Entonces:*

1. *A es definida positiva si y sólo si $\lambda > 0, \forall \lambda \in Sp(A)$.*
2. *A es definida negativa si y sólo si $\lambda < 0, \forall \lambda \in Sp(A)$.*
3. *A es semidefinida positiva si y sólo si $\lambda \geq 0, \forall \lambda \in Sp(A)$.*
4. *A es semidefinida negativa si y sólo si $\lambda \leq 0, \forall \lambda \in Sp(A)$.*
5. *En cualquier otro caso, A es indefinida.*

Ejemplo:

La matriz

$$A = \begin{pmatrix} 2 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 2 \end{pmatrix}$$

es definida positiva ya que $\text{Sp}(A) = \{1, 4\}$, con m.a. (1) = 2, m.a. (4) = 1.

Uso de los menores principales.

Las formas cuadráticas no degeneradas se pueden clasificar analizando el signo de los menores principales de la matriz.

Teorema 6.4 *Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica. Las siguientes afirmaciones son equivalentes:*

1. A es definida positiva.
2. Todos los autovalores de A son estrictamente positivos.
3. Todos los menores diagonales principales de A son mayores que cero.

Ejemplo: Consideremos la forma cuadrática $\omega : \mathbb{R}^3 \rightarrow \mathbb{R}$ definida por $\omega(x) = x^t Ax$, donde

$$A = \begin{pmatrix} 2 & -2 & 1 \\ -2 & 3 & -1 \\ 1 & -1 & 1 \end{pmatrix}$$

Los menores principales de A son:

$$\Delta_1 = 2 > 0$$

$$\Delta_2 = \begin{vmatrix} 2 & -2 \\ -2 & 3 \end{vmatrix} = 2 > 0$$

$$\Delta_3 = \begin{vmatrix} 2 & -2 & 1 \\ -2 & 3 & -1 \\ 1 & -1 & 1 \end{vmatrix} = 1 > 0.$$

Como todos son positivos, A es definida positiva.

El resultado anterior se puede aplicar también a matrices definidas negativas, teniendo en cuenta que A es definida negativa si y sólo si $B = -A$ es definida positiva y que si $A_k \in \mathcal{M}_{k \times k}(\mathbb{R})$ entonces $\det(-A_k) = (-1)^k \det(A_k)$. De este modo se obtiene el siguiente resultado:

Proposición 6.5 *Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica. A es definida negativa si y sólo si los menores diagonales principales de orden impar son menores que cero y los de orden par son mayores que cero.*

El uso de los menores principales se puede resumir en el siguiente resultado:

Teorema 6.5 Si $|A| \neq 0$ entonces la forma cuadrática $\omega(x) = x^t Ax$ es no degenerada y sólo puede ser definida positiva, definida negativa o indefinida. En este caso, la clasificación se puede hacer usando los menores principales:

- (a) Si todos los menores principales de A son positivos entonces ω es definida positiva.
- (b) Si los menores principales de orden impar son negativos y los de orden par son positivos entonces ω es definida negativa.
- (c) En cualquier otro caso, ω es indefinida.

Ejemplo: Consideremos la matriz

$$A = \begin{pmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}.$$

Como $|A| = 2 \neq 0$, estamos en el caso no degenerado y sólo puede ser definida positiva, definida negativa o indefinida. Como el primer menor principal es $\Delta_1 = 0$, A no puede ser definida positiva ni definida negativa. En consecuencia, A es indefinida. (Alternativamente, se puede comprobar que $\text{Sp}(A) = \{-1, 2\}$, con m.a. $(-1) = 2$, m.a. $(2) = 1$.)

Otras observaciones útiles.

En el caso degenerado ($|A| = 0$) la forma cuadrática $\omega(x) = x^t Ax$ sólo puede ser indefinida o semidefinida, pero en este caso la clasificación no se puede deducir de los menores principales y en general hay que recurrir al cálculo de autovalores. Sin embargo, el siguiente resultado permite resolver de forma inmediata algunas situaciones.

Proposición 6.6 Sea $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica y sea $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ la forma cuadrática definida por $\omega(x) = x^t Ax$, $\forall x \in \mathbb{R}^n$.

Si $B = \{e_1, e_2, \dots, e_n\}$ es la base canónica de \mathbb{R}^n entonces $\omega(e_i) = a_{ii}$, $\forall i = 1, 2, \dots, n$.

Esta proposición permite llegar a ciertas conclusiones sin más que observar los elementos diagonales de A . En particular:

Corolario 6.1 Sea $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica y sea $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ la forma cuadrática definida por $\omega(x) = x^t Ax$, $\forall x \in \mathbb{R}^n$.

Si existen i, j tales que $a_{ii} > 0$, $a_{jj} < 0$ entonces necesariamente ω es indefinida.

Ejemplo: Consideremos la matriz

$$A = \begin{pmatrix} 1 & -1 & 1 \\ -1 & -1 & 1 \\ 1 & 1 & -1 \end{pmatrix}.$$

Como $|A| = 0$, estamos en el caso degenerado y sólo puede ser semidefinida o indefinida. Ahora bien, como $a_{11} = 1 > 0$, $a_{22} = -1 < 0$, podemos concluir que A es indefinida.

6.7 Mínimos cuadrados. Ajuste.

Consideremos un sistema de ecuaciones lineales $Ax = b$, donde $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $b \in \mathbb{R}^p$. Recordemos que la imagen de A es

$$\text{Im}(A) = \{Ax / x \in \mathbb{R}^n\} \subset \mathbb{R}^p.$$

La compatibilidad del sistema $Ax = b$ se caracteriza en términos del subespacio $\text{Im}(A)$ de forma sencilla.

Proposición 6.7 *El sistema $Ax = b$ es compatible si y sólo si $b \in \text{Im}(A)$.*

Demostración.

$$Ax = b \text{ es compatible} \iff \exists x \in \mathbb{R}^n / Ax = b \iff b \in \text{Im}(A). \quad \square$$

En el caso de que el sistema sea incompatible, buscaremos una “solución aproximada”. Una posibilidad es determinar el vector $y \in \text{Im}(A)$ cuya distancia al término independiente b sea la menor posible. Los vectores $x \in \mathbb{R}^n$ tales que $Ax = y$ serán lo que llamaremos soluciones del sistema $Ax = b$ en el sentido de mínimos cuadrados. Así, se tiene la siguiente definición:

Definición 6.9 *Sean $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $b \in \mathbb{R}^p$. Se dice que $x_0 \in \mathbb{R}^n$ es una **solución en el sentido de mínimos cuadrados** del sistema $Ax = b$ si se verifica:*

$$\|Ax_0 - b\| = \min\{\|Ax - b\| / x \in \mathbb{R}^n\}.$$

Recordemos que la distancia mínima de b a $\text{Im}(A)$ es la distancia de b a la proyección ortogonal de b sobre $\text{Im}(A)$ y por tanto x_0 es una solución de $Ax = b$ en el sentido de mínimos cuadrados si y sólo si $v = Ax_0 - b$ es ortogonal al subespacio $\text{Im}(A)$. Esto permite probar el siguiente resultado:

Teorema 6.6 *Sean $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $b \in \mathbb{R}^p$. Un vector x_0 es una solución en el sentido de mínimos cuadrados de $Ax = b$ si y sólo si*

$$A^t Ax_0 = A^t b.$$

Demostración. El vector x_0 es solución de mínimos cuadrados de $Ax = b$ si y sólo si $(Ax_0 - b)$ es ortogonal a $\text{Im}(A)$, es decir:

$$\langle Ax, Ax_0 - b \rangle = 0, \quad \forall x \in \mathbb{R}^n,$$

lo que equivale a

$$0 = (Ax)^t (Ax_0 - b) = x^t A^t (Ax_0 - b) = \langle x, A^t (Ax_0 - b) \rangle, \quad \forall x \in \mathbb{R}^n.$$

Esto sólo es posible si $A^t (Ax_0 - b) = 0$, es decir, $A^t Ax_0 = A^t b$. □

Definición 6.10 *El sistema de ecuaciones lineales $A^tAx = A^tb$ cuyas soluciones son las soluciones en el sentido de mínimos cuadrados del sistema $Ax = b$ se conoce con el nombre de sistema de ecuaciones normales del sistema $Ax = b$.*

El siguiente resultado es una consecuencia de que en \mathbb{R}^n siempre es posible calcular la proyección ortogonal de un vector v sobre un subespacio U . Además, si $v \in U$ entonces la proyección ortogonal es el propio v .

Teorema 6.7 *Sean $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $b \in \mathbb{R}^p$. El sistema de ecuaciones lineales $A^tAx = A^tb$ es un sistema compatible. Además:*

- (1) *Si $Ax = b$ es compatible entonces el conjunto de soluciones de $A^tAx = A^tb$ coincide con el conjunto de soluciones de $Ax = b$.*
- (2) *Si $Ax = b$ es incompatible entonces el conjunto de soluciones de $A^tAx = A^tb$ coincide con el conjunto de soluciones de $Ax = b$ en el sentido de mínimos cuadrados.*

Ajuste polinómico de datos mediante mínimos cuadrados.

Supongamos que se calcula experimentalmente el valor de una cierta cantidad y que se supone que es función polinómica de otra cantidad x :

$$y = p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_nx^n.$$

Si se realizan k experimentos en los que se obtienen las mediciones y_1, y_2, \dots, y_k para los datos de entrada respectivos x_1, x_2, \dots, x_k , los coeficientes del polinomio $p(x)$ vendrían dados por las soluciones del sistema de ecuaciones lineales

$$\begin{cases} y_1 = a_0 + a_1x_1 + a_2x_1^2 + \cdots + a_nx_1^n \\ y_2 = a_0 + a_1x_2 + a_2x_2^2 + \cdots + a_nx_2^n \\ \vdots \\ y_k = a_0 + a_1x_k + a_2x_k^2 + \cdots + a_nx_k^n, \end{cases}$$

o, en forma matricial,

$$\underbrace{\begin{pmatrix} 1 & x_1 & x_1^2 & \cdots & x_1^n \\ 1 & x_2 & x_2^2 & \cdots & x_2^n \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_k & x_k^2 & \cdots & x_k^n \end{pmatrix}}_A \underbrace{\begin{pmatrix} a_0 \\ a_1 \\ a_2 \\ \vdots \\ a_n \end{pmatrix}}_x = \underbrace{\begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_k \end{pmatrix}}_b.$$

Si el sistema $Ax = b$ es compatible entonces la gráfica del polinomio cuyos coeficientes son la solución del sistema pasa por todos los puntos $(x_1, y_1), (x_2, y_2), \dots, (x_k, y_k)$. Si no es compatible, la solución del sistema de ecuaciones normales $A^tAx = A^tb$ proporciona los coeficientes del

polinomio de grado n que mejor ajusta los datos en el sentido de mínimos cuadrados.

Observación: Si el polinomio $p(x)$ que buscamos es de grado 1 se dice que el ajuste es *lineal*. Si $p(x)$ es de grado 2, se dice que el ajuste es *cuadrático*.

Ejemplo: Encontrar la recta y la parábola de ajuste en el sentido de mínimos cuadrados para los siguientes datos:

$$\begin{array}{c|cccc} x & -2 & -1 & 1 & 2 \\ \hline y & 3 & 1 & 1 & 5 \end{array}$$

La recta tiene la forma $y = a_0 + a_1x$, de modo que buscamos la solución de mínimos cuadrados del sistema

$$\begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 1 \\ 5 \end{pmatrix}.$$

Denotando por

$$A = \begin{pmatrix} 1 & -2 \\ 1 & -1 \\ 1 & 1 \\ 1 & 2 \end{pmatrix}, \quad b = \begin{pmatrix} 3 \\ 1 \\ 1 \\ 5 \end{pmatrix},$$

el sistema de ecuaciones normales $A^t Ax = A^t b$ es

$$\begin{pmatrix} 4 & 0 \\ 0 & 10 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \end{pmatrix}.$$

Por tanto, $a_0 = 5/2$, $a_1 = 2/5$ y la recta es $y = \frac{5}{2} + \frac{2}{5}x$.

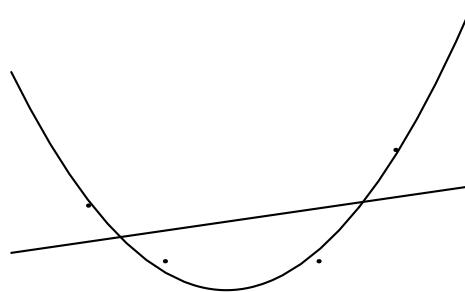


Figura 6.1: Aproximaciones lineal y cuadrática de los datos.

Si ahora buscamos la parábola $y = a_0 + a_1x + a_2x^2$ que ajusta mejor estos datos en el sentido de mínimos cuadrados, planteamos el sistema

$$\begin{pmatrix} 1 & -2 & 4 \\ 1 & -1 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 4 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 3 \\ 1 \\ 1 \\ 5 \end{pmatrix}.$$

El sistema de ecuaciones normales es

$$\begin{pmatrix} 4 & 0 & 10 \\ 0 & 10 & 0 \\ 10 & 0 & 34 \end{pmatrix} \begin{pmatrix} a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 10 \\ 4 \\ 34 \end{pmatrix},$$

y tiene como solución $(a_0, a_1, a_2) = (0, 2/5, 1)$. En consecuencia, la ecuación de la parábola de ajuste es

$$y = a_0 + a_1x + a_2x^2 = \frac{2}{5}x + x^2.$$

En la figura 6.1 se representan los puntos y las aproximaciones lineal y cuadrática. Se observa que ésta última es mucho más precisa.

6.8 Descomposición en valores singulares.

Sea $A \in \mathcal{M}_{p \times n}(\mathbb{R})$. Entonces $A^t A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz simétrica. En particular, todos los autovalores de $A^t A$ son reales. Además son no negativos:

Proposición 6.8 *Todos los autovalores de $A^t A$ son mayores o iguales que cero.*

Demostración. Sea $\lambda \in \text{Sp}(A^t A)$ y x un autovector asociado. Entonces:

$$\|Ax\|^2 = \langle Ax, Ax \rangle = x^t A^t Ax = \lambda x^t x = \lambda \|x\|^2 \implies \lambda = \frac{\|Ax\|^2}{\|x\|^2} \geq 0.$$

□

Definición 6.11 *Sean $A \in \mathcal{M}_{p \times n}(\mathbb{R})$. Se llaman **valores singulares** de A a las raíces cuadradas positivas de $A^t A$, es decir, si $\text{Sp}(A^t A) = \{\lambda_1, \dots, \lambda_n\}$ entonces los valores singulares de A son $\sqrt{\lambda_1}, \dots, \sqrt{\lambda_n}$. Se suelen denotar $\sigma_1, \dots, \sigma_n$ y se ordenan de tal forma que $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_n \geq 0$.*

Ejemplo: Calcular los valores singulares de

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \in \mathcal{M}_{4 \times 3}(\mathbb{R}).$$

$$A^t A = \begin{pmatrix} 0 & 1 & -1 & 1 \\ 0 & 1 & 1 & -1 \\ 1 & 0 & 2 & 2 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 9 \end{pmatrix}.$$

Los autovalores de $A^t A$ son 2, 4 y 9, de modo que los valores singulares de A son

$$\sigma_1 = \sqrt{9} = 3$$

$$\sigma_2 = \sqrt{4} = 2$$

$$\sigma_3 = \sqrt{2}.$$

Una de las principales aplicaciones de los valores singulares es que permiten obtener una descomposición de A como suma de r matrices de rango 1, donde $r = \text{rg}(A)$.

Teorema 6.8 Descomposición en valores singulares. *Sea $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ con $\text{rg}(A) = r$ y valores singulares no nulos $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$. Entonces existen dos matrices ortogonales $U \in \mathcal{M}_{p \times p}(\mathbb{R})$, $V \in \mathcal{M}_{n \times n}(\mathbb{R})$ y una matriz $\Sigma \in \mathcal{M}_{p \times n}(\mathbb{R})$ tales que $A = U\Sigma V^t$, donde*

$$\Sigma = \left(\begin{array}{c|c} D & 0 \\ \hline 0 & 0 \end{array} \right), \text{ con } D = \begin{pmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{pmatrix} \in \mathcal{M}_{r \times r}(\mathbb{R}).$$

Ejemplo:

En el ejemplo anterior,

$$\Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}$$

Observación: El rango de A coincide con el número de valores singulares no nulos de A (contados con su multiplicidad).

Podemos obtener una expresión extendida de la descomposición en valores singulares de modo similar al que utilizamos para definir la descomposición espectral de una matriz simétrica:

Teorema 6.9 *Sea $A = U\Sigma V^t$ una descomposición en valores singulares de una matriz A de rango r . Si u_1, u_2, \dots, u_r y v_1, v_2, \dots, v_r son las r primeras columnas de U y V respectivamente entonces*

$$A = U\Sigma V^t = \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \cdots + \sigma_r u_r v_r^t.$$

Definición 6.12 *Sea $A = \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \cdots + \sigma_r u_r v_r^t$ la descomposición en valores singulares de una matriz A de rango r . Si k es cualquier número entero positivo menor que r , se llama **aproximación de rango k** de A a la matriz A_k que se obtiene sumando los k primeros términos de la expresión anterior, es decir,*

$$A_k = \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \cdots + \sigma_k u_k v_k^t.$$

De entre todas las matrices de rango k que tienen el mismo tamaño que A , la matriz A_k es la que más se parece a A en cierto sentido. Concretamente, se puede definir una norma en el espacio de matrices $\mathcal{M}_{p \times n}(\mathbb{R})$ del siguiente modo:

$$\|A\| = \sigma_1 = \max\{\sigma_1, \sigma_2, \dots, \sigma_n\}.$$

Es decir, la norma de A es el mayor de sus valores singulares. Dicha norma se llama **norma espectral** de A .

Se puede probar que $\|A - A_k\| = \sigma_{k+1} = \min\{\|A - B\| / B \in \mathcal{M}_{p \times n}(\mathbb{R}), \text{rg}(B) = k\}$.

La descomposición en valores singulares de una matriz A se suele llamar $\text{SVD}(A)$ (las iniciales de la traducción al inglés “singular value decomposition”).

A continuación se describe el método para calcular tanto la SVD de A como sus aproximaciones de rango k para cada $k < r$.

Cálculo de la SVD y la aproximación de rango k .

Sea $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ con $\text{rg}(A) = r$.

- (1) Los vectores v_1, v_2, \dots, v_r se obtienen calculando bases ortonormales de los subespacios propios asociados a los autovalores no nulos de $A^t A$, ordenados de mayor a menor.
- (2) Denotemos $V = (v_1 | v_2 | \dots | v_n)$ y $U = (u_1 | u_2 | \dots | u_p)$. Como $A = U \Sigma V^t$, se deduce que $AV = U \Sigma$ y por tanto $Av_i = \sigma_i u_i$, $\forall i = 1, 2, \dots, r$. En consecuencia, las primeras r columnas de U se obtienen directamente de las de V mediante las fórmulas

$$u_i = \frac{1}{\sigma_i} Av_i, \quad \forall i = 1, 2, \dots, r.$$

- (3) Una vez que hemos calculado las r primeras columnas de U y V , podemos obtener la SVD de A y sus aproximaciones de rango k :

$$A = (u_1 | u_2 | \dots | u_r) \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_r \end{pmatrix} \begin{pmatrix} \frac{v_1^t}{\sigma_1} \\ \frac{v_2^t}{\sigma_2} \\ \vdots \\ \frac{v_r^t}{\sigma_r} \end{pmatrix} = \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \dots + \sigma_r u_r v_r^t;$$

$$A_k = (u_1 | u_2 | \dots | u_k) \begin{pmatrix} \sigma_1 & 0 & \dots & 0 \\ 0 & \sigma_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma_k \end{pmatrix} \begin{pmatrix} \frac{v_1^t}{\sigma_1} \\ \frac{v_2^t}{\sigma_2} \\ \vdots \\ \frac{v_k^t}{\sigma_k} \end{pmatrix} = \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \dots + \sigma_k u_k v_k^t.$$

Ejemplo: Calcular una descomposición en valores singulares de la matriz

$$A = \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \in \mathcal{M}_{4 \times 3}(\mathbb{R})$$

y su aproximación de rango dos A_2 .

Ya hemos calculado las matrices $A^t A$ y Σ :

$$A^t A = \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 9 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & \sqrt{2} \\ 0 & 0 & 0 \end{pmatrix}.$$

Por tanto, $\text{rg}(A) = 3$ y los vectores v_1, v_2, v_3 se obtienen calculando una base ortonormal de cada uno de los subespacios propios de $A^t A$. Dado que

$$\begin{aligned} V(9) &= \text{Ker}(A^t A - 9I) = \langle \{(0, 0, 1)\} \rangle, \\ V(4) &= \text{Ker}(A^t A - 4I) = \langle \{(1, -1, 0)\} \rangle, \\ V(2) &= \text{Ker}(A^t A - 2I) = \langle \{(1, 1, 0)\} \rangle, \end{aligned}$$

se obtiene sin más que dividir cada vector por su norma que $\mathcal{B}_1 = \{(0, 0, 1)\}$ es una base ortonormal de $V(9)$, $\mathcal{B}_2 = \{(1/\sqrt{2}, -1/\sqrt{2}, 0)\}$ es una base ortonormal de $V(4)$ y $\mathcal{B}_3 = \{(1/\sqrt{2}, 1/\sqrt{2}, 0)\}$ es una base ortonormal de $V(2)$.

Por tanto,

$$V = (v_1|v_2|v_3) = \begin{pmatrix} 0 & 1/\sqrt{2} & 1/\sqrt{2} \\ 0 & -1/\sqrt{2} & 1/\sqrt{2} \\ 1 & 0 & 0 \end{pmatrix}.$$

Los vectores u_1, u_2 y u_3 se calculan directamente:

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{9} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 0 \\ 2/3 \\ 2/3 \end{pmatrix};$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{4} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix};$$

$$u_3 = \frac{1}{\sigma_3} A v_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

La descomposición en valores singulares de A es $A = 3u_1v_1^t + 2u_2v_2^t + \sqrt{2}u_3v_3^t$.

La aproximación de rango 2 de A se obtiene tomando los dos primeros sumandos en la expresión anterior:

$$\begin{aligned} A_2 = 3u_1v_1^t + 2u_2v_2^t &= 3 \begin{pmatrix} 1/3 \\ 0 \\ 2/3 \\ 2/3 \end{pmatrix} (0, 0, 1) + 2 \begin{pmatrix} 0 \\ 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix} (1/\sqrt{2}, -1/\sqrt{2}, 0) = \\ &= \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & 0 & 2 \\ 0 & 0 & 2 \end{pmatrix} + \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & -1 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix}. \end{aligned}$$

Referencias

Algunos libros donde buscar más información y, en particular, muchos ejemplos y aplicaciones del álgebra lineal:

- D. C. LAY, “Álgebra Lineal y sus Aplicaciones” (3^a ed.), Pearson Educación, 2007.
- G. NAKOS Y D. JOYNER, “Álgebra Lineal con aplicaciones”, Thomson, 1999.
- D. POOLE, “Álgebra Lineal con aplicaciones” (2^a ed.), Thomson, 2007.