

Apuntes de álgebra lineal

EDUARDO LIZ MARZÁN

Enero de 2015.

Índice general

1. Introducción	7
1.1. Operaciones internas y estructura de cuerpo.	7
1.2. Números complejos.	8
1.3. Vectores.	10
2. Matrices y determinantes	13
2.1. Introducción.	13
2.2. Definición y tipos de matrices.	13
2.3. Operaciones con matrices.	14
2.4. Trasposición de matrices.	18
2.5. Matrices elementales.	19
2.6. Forma escalonada y rango de una matriz.	21
2.7. Cálculo de la inversa.	23
2.8. Determinantes.	24
2.9. Formas cuadráticas.	26
3. Sistemas de ecuaciones lineales	31
3.1. Introducción.	31
3.2. Expresión matricial.	31
3.3. Existencia de soluciones.	32
3.4. Conjuntos de soluciones.	33
3.5. Matrices cuadradas y uso de la factorización LU	35
3.6. Mínimos cuadrados. Ajuste.	39
4. Espacios vectoriales y aplicaciones lineales	43
4.1. Introducción.	43
4.2. Espacios y subespacios vectoriales.	43
4.3. Independencia lineal.	45
4.4. Bases y dimensión.	46
4.5. Cambio de base en \mathbb{R}^n	47
4.6. Bases ortonormales.	49
4.7. Definición de aplicación lineal y matriz asociada.	50
4.8. Núcleo e imagen de una aplicación lineal.	52
4.9. Inversas de aplicaciones lineales.	53

4.10. Transformaciones ortogonales.	54
4.11. Proyección ortogonal.	54
5. Diagonalización y funciones de matrices	57
5.1. Introducción.	57
5.2. Autovalores y autovectores.	57
5.3. Matrices diagonalizables.	60
5.4. Diagonalización ortogonal.	62
5.5. Clasificación de formas cuadráticas usando la diagonalización ortogonal.	64
5.6. Descomposición espectral	65
5.7. Descomposición en valores singulares.	65
5.8. Teorema de Cayley-Hamilton.	69
5.9. Funciones de matrices.	71
Referencias	75

Introducción

Existen muchos libros de álgebra lineal (véanse, por ejemplo, las referencias al final de este documento), por lo que escribir uno más no tiene mucho sentido. Estos apuntes deben considerarse una ayuda para que los alumnos tengan el material del curso organizado.

Escribí la primera versión cuando impartía la asignatura de álgebra lineal en la *Escuela de Ingeniería de Telecomunicación* de la Universidad de Vigo y desde el curso 2010/2011 se siguen en las titulaciones de *Ingeniería de la Energía* e *Ingeniería de los Recursos Mineros y Energéticos*, que comparten las actividades docentes en el primer curso. El programa se desarrolla en 40 horas de grupo A (de aproximadamente 50 alumnos) y 10 horas de grupo B (aproximadamente 20 alumnos). En estas últimas se resuelven problemas y se realizan algunas prácticas de ordenador.

A lo largo de los años los apuntes han experimentado varias modificaciones, algunas de ellas como consecuencia de comentarios de los alumnos y de algunos compañeros. En especial quiero agradecer mis discusiones con Elvira Hernández García, Profesora Titular de la E.T.S.I. Industriales de la UNED (Madrid).

Eduardo Liz Marzán
Vigo, enero de 2015.

Capítulo 1

Introducción

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 cumplan 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 general, 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 en una operación producto se cumple la propiedad conmutativa entonces el elemento inverso se suele denotar 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 cumple 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 cumple 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**.

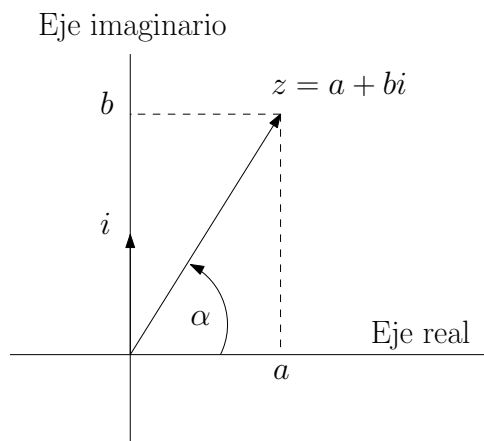


Figura 1.1: Representación de un número complejo $z = a + bi$ en el plano complejo. El ángulo α es el argumento de z . La unidad imaginaria i se sitúa en el eje imaginario y tiene módulo 1.

Operaciones en \mathbb{C}

- **Suma.** Sean $z_1 = a_1 + b_1i$, $z_2 = a_2 + b_2i$ dos números complejos. Se define su suma como

$$z_1 + z_2 = (a_1 + a_2) + (b_1 + b_2)i.$$

- **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 $z = a + bi \in \mathbb{C}$, su conjugado es $\bar{z} = a - bi$. Obsérvese que $z\bar{z} = a^2 + b^2$ y por tanto

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

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 $|z| = 0 \Leftrightarrow z = 0$. Usando el módulo, el inverso de un número complejo $z \neq 0$ se expresa como $z^{-1} = \bar{z}/|z|^2$.

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,

$$z = a + bi = |z|(\cos(\alpha) + \sin(\alpha)i),$$

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 (ver Figura 1.1).

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

$$z_1 z_2 = |z_1| |z_2| (\cos(\alpha_1 + \alpha_2) + \sin(\alpha_1 + \alpha_2)i),$$

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 $z = |z|(\cos(\alpha) + \sin(\alpha)i)$ entonces $z^n = |z|^n (\cos(n\alpha) + \sin(n\alpha)i)$, $\forall n \in \mathbb{N}$.

Forma exponencial

Si $b \in \mathbb{R}$, se define $e^{bi} = \cos(b) + \sin(b)i$. (En realidad esta fórmula se obtiene usando desarrollos en serie de las funciones exponencial, seno y coseno).

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

Las fórmulas para el producto y las potencias de números complejos resultan más sencillas cuando se utiliza la forma exponencial:

$$\begin{aligned} z_1 z_2 &= (|z_1| e^{\alpha_1 i}) \cdot (|z_2| e^{\alpha_2 i}) = |z_1| |z_2| e^{(\alpha_1 + \alpha_2) i}. \\ z^n &= (|z| e^{\alpha i})^n = |z|^n (e^{\alpha i})^n = |z|^n e^{n\alpha i}. \end{aligned}$$

1.3. Vectores.

Se define \mathbb{R}^2 como el conjunto de los pares ordenados de números reales, es decir:

$$\mathbb{R}^2 = \{(x_1, x_2) / x_1, x_2 \in \mathbb{R}\}.$$

Cada elemento (x_1, x_2) de \mathbb{R}^2 es un punto en el plano bidimensional; la proyección sobre el eje horizontal es la coordenada x_1 y la proyección sobre el eje vertical es la coordenada x_2 . El punto (x_1, x_2) se llama vector de \mathbb{R}^2 y se puede representar por una flecha con origen en $(0, 0)$ y extremo en (x_1, x_2) .

La suma de dos vectores de \mathbb{R}^2 se realiza coordenada a coordenada; si $x = (x_1, x_2)$ e $y = (y_1, y_2)$ entonces

$$x + y = (x_1, x_2) + (y_1, y_2) = (x_1 + y_1, x_2 + y_2).$$

El producto de un escalar $\lambda \in \mathbb{R}$ por un vector (x_1, x_2) de \mathbb{R}^2 proporciona otro vector λx dado por

$$\lambda x = \lambda(x_1, x_2) = (\lambda x_1, \lambda x_2).$$

Tanto el conjunto \mathbb{R}^2 como las operaciones de suma y producto por escalares se generalizan a dimensiones mayores. Así,

$$\mathbb{R}^3 = \{(x_1, x_2, x_3) / x_1, x_2, x_3 \in \mathbb{R}\},$$

y, en general, para cada número natural $n \geq 2$, se define

$$\mathbb{R}^n = \{(x_1, x_2, \dots, x_n) / x_i \in \mathbb{R}, \forall i = 1, 2, \dots, n\}.$$

Por ejemplo, $x = (2, -1, 0, -2)$ es un vector de \mathbb{R}^4 .

Un vector $v \in \mathbb{R}^n$ es una **combinación lineal** de vectores v_1, v_2, \dots, v_k de \mathbb{R}^n si se obtiene de los anteriores mediante sumas y productos por escalares, es decir:

$$v = \lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_k v_k.$$

Por ejemplo,

$$(5, -2, 8) = 2(1, -1, 1) + 3(1, 0, 2),$$

de modo que $v = (5, -2, 8)$ es una combinación lineal de $v_1 = (1, -1, 1)$ y $v_2 = (1, 0, 2)$.

Se dice que k vectores v_1, v_2, \dots, v_k de \mathbb{R}^n son **linealmente independientes** si ninguno de ellos es combinación lineal del resto. Por ejemplo, $v_1 = (1, -1, 1)$ y $v_2 = (1, 0, 2)$ son vectores de \mathbb{R}^3 linealmente independientes.

El conjunto U de todas las combinaciones lineales de k vectores v_1, v_2, \dots, v_k de \mathbb{R}^n se llama **subespacio generado** por v_1, v_2, \dots, v_k y se denota por $U = \langle \{v_1, v_2, \dots, v_k\} \rangle$. El conjunto $\mathcal{B} = \{v_1, v_2, \dots, v_k\}$ se llama conjunto de **generadores** de U . Si \mathcal{B} es linealmente independiente se dice que \mathcal{B} es una **base** de U . El número de elementos de \mathcal{B} se llama **dimensión** de U y lo denotaremos por $\dim(U)$.

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**. En particular, $\dim(\mathbb{R}^n) = n$.

Ejemplo: Se considera en \mathbb{R}^3 el conjunto $U = \{(x, y, z) \in \mathbb{R}^3 / y = 2x - z\}$. Entonces:

$$\begin{aligned} U &= \{(x, 2x - z, z) / x, z \in \mathbb{R}\} = \{(x, 2x, 0) + (0, -z, z) / x, z \in \mathbb{R}\} = \\ &= \{x(1, 2, 0) + z(0, -1, 1) / x, z \in \mathbb{R}\} = \langle \{(1, 2, 0), (0, -1, 1)\} \rangle. \end{aligned}$$

Por tanto $\mathcal{B} = \{(1, 2, 0), (0, -1, 1)\}$ es una base de U y $\dim(U) = 2$.

La dimensión de un subespacio caracteriza su número máximo de direcciones linealmente independientes y proporciona una medida de su tamaño. El subespacio U del ejemplo anterior es un plano en \mathbb{R}^3 .

Producto escalar

Se define el **producto escalar usual** de dos vectores $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$ de \mathbb{R}^n como

$$x \cdot y = x_1 y_1 + x_2 y_2 + \dots + x_n y_n = \sum_{i=1}^n x_i y_i.$$

El producto escalar permite definir una **norma** (o módulo). Si $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, se define

$$\|x\| = +\sqrt{x \cdot x} = +\sqrt{x_1^2 + x_2^2 + \dots + x_n^2}.$$

Si x, y son dos vectores de \mathbb{R}^n entonces $\|x - y\|$ representa la distancia de x a y . En particular, la norma de x representa su distancia al origen de coordenadas.

En \mathbb{R}^2 el producto escalar usual de dos vectores x, y coincide con la definición clásica en función del ángulo ϕ que forman x e y :

$$x \cdot y = \|x\| \|y\| \cos(\phi).$$

El concepto de ángulo se extiende a \mathbb{R}^n usando el producto escalar. Si $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$ son dos vectores no nulos de \mathbb{R}^n entonces se define el ángulo que forman como el ángulo $\phi \in [0, \pi]$ que cumple la fórmula:

$$\cos(\phi) = \frac{x \cdot y}{\|x\| \|y\|}.$$

Un coseno próximo a 1 indica que las direcciones de x e y están próximas.

Por ejemplo, si $x = (1, 1, 1)$ e $y = (1, 0, -1)$ entonces $\cos(\phi) = 0$ y por tanto x e y forman un ángulo de $\pi/2$.

Se dice que dos vectores x e y de \mathbb{R}^n son **ortogonales** si $x \cdot y = 0$. Un conjunto de vectores $\{v_1, v_2, \dots, v_k\}$ de \mathbb{R}^n es **ortogonal** si $v_i \cdot v_j = 0$, $\forall i \neq j$. Un conjunto de vectores $\{v_1, v_2, \dots, v_k\}$ de \mathbb{R}^n es **ortonormal** si es ortogonal y $\|v_i\| = 1$, $\forall i = 1, 2, \dots, k$.

Por ejemplo, el conjunto

$$\left\{ \left(\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}} \right), \left(0, \frac{1}{\sqrt{2}}, -\frac{1}{\sqrt{2}} \right) \right\}$$

es un conjunto ortonormal de \mathbb{R}^3 .

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.

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. Como aplicación de los determinantes veremos la clasificación de formas cuadráticas no degeneradas.

2.2. Definición y tipos de matrices.

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

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

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 $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$.

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

Si $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz cuadrada, se llama **diagonal** de A al vector de \mathbb{R}^n que contiene los elementos a_{ij} con $i = j$, es decir, $\text{diag}(A) = (a_{11}, a_{22}, \dots, a_{nn})$. La suma de los elementos diagonales de A se llama **traza** de A y se denota por $\text{tr}(A)$. Es decir,

$$\text{tr}(A) = \sum_{i=1}^n a_{ii} = a_{11} + a_{22} + \cdots + a_{nn}.$$

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

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

También serán importantes las matrices triangulares.

- Una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ 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 . 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 comprobar 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:

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

es el producto escalar $(u \cdot v)$, mientras que

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

Expresiones 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. Consideraremos cuatro casos.

1. El producto de dos matrices $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $B \in \mathcal{M}_{n \times q}(\mathbb{R})$ se puede expresar en función de productos escalares de las filas de A por las columnas de B :

$$AB = \begin{pmatrix} \frac{u_1^t}{\vdots} \\ \frac{u_p^t}{\vdots} \end{pmatrix} (v_1 | v_2 | \cdots | v_q) = \begin{pmatrix} u_1^t v_1 & u_1^t v_2 & \cdots & u_1^t v_q \\ u_2^t v_1 & u_2^t v_2 & \cdots & u_2^t v_q \\ \vdots & \vdots & \vdots & \vdots \\ u_p^t v_1 & u_p^t v_2 & \cdots & u_p^t v_q \end{pmatrix} \in \mathcal{M}_{p \times q}(\mathbb{R}).$$

De este modo se suele calcular el producto en la práctica.

2. El producto AB también se puede obtener como suma de matrices que resultan de multiplicar las columnas de A por las filas de B . Esta fórmula será útil en varias partes del curso.

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

3. En el caso particular de que B sea un vector columna, el producto se puede interpretar como una combinación lineal de las columnas de A : sean

$$A = (u_1 | u_2 | \cdots | u_n) \in \mathcal{M}_{p \times n}(\mathbb{R}) \quad , \quad B = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in \mathcal{M}_{n \times 1}(\mathbb{R}).$$

Entonces:

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

4. Finalmente, si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $B = (u_1|u_2|\cdots|u_q) \in \mathcal{M}_{n \times q}(\mathbb{R})$, entonces:

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

Propiedades del producto de matrices

- El producto de matrices cumple la propiedad asociativa, es decir si A , B y C se pueden multiplicar entonces $(AB)C = A(BC)$.
- 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 $IB = B$, $\forall B \in \mathcal{M}_{n \times q}(\mathbb{R})$.

- El producto de matrices no es conmutativo, 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:

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

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.

Una matriz cuadrada $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ se dice **invertible** si existe una matriz, que llamaremos inversa de A y denotaremos por A^{-1} , tal que $AA^{-1} = A^{-1}A = I$, donde I es la matriz identidad.

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

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $k \in \mathbb{N}$. La **potencia** k -ésima de A es la matriz que resulta de multiplicar A por sí misma k veces. Se denota por A^k . Es decir,

$$A^k = \underbrace{A \cdot A \cdots A}_k.$$

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,

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

2.4. Trasposición de matrices.

Recordemos que si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ entonces $A^t \in \mathcal{M}_{n \times p}(\mathbb{R})$ 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 invertible entonces $(A^t)^{-1} = (A^{-1})^t$.

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

- Una matriz $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$ es **simétrica** si $A^t = A$, es decir, si

$$a_{ij} = a_{ji}, \forall i, j = 1, 2, \dots, n.$$

Ejemplo:

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

La siguiente propiedad permite construir una matriz simétrica a partir de cualquier matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y será importante en temas posteriores.

Propiedad: Si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ entonces $A^t A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es simétrica.

- Una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ 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 matriz de rotación de ángulo α

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

es ortogonal.

2.5. Matrices elementales.

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

- Permutar dos filas o dos columnas de A .
- Sumar a una fila (o columna) de A un múltiplo de otra fila (o columna) de A .
- Multiplicar una fila o columna de A por un escalar no nulo.

Las operaciones elementales no afectan a la independencia lineal. Si una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ tiene k filas linealmente independientes y se realizan operaciones elementales por filas en A entonces la matriz resultante también tiene k filas linealmente independientes. Además, el subespacio de \mathbb{R}^n que generan es el mismo.

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 la matriz identidad. 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 $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ equivale 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. Restar a la fila 2 de A la fila 1 multiplicada por 3 es equivalente a multiplicar A por la izquierda por $F_{21}(-3)$:

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

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

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

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:

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

2.6. Forma escalonada y rango de una matriz.

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 $a_{ij} \neq 0$ y $a_{ik} = 0$ para todo $k < j$.

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.

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.*

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. 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.

Proposición 2.1 *El rango de una matriz A coincide con el número de filas linealmente independientes de A .*

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

Observación: El rango de A también coincide con el número de columnas linealmente independientes de A . Esto es equivalente a decir que $\text{rg}(A) = \text{rg}(A^t)$.

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

Proposición 2.2 *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

2.7. 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 2.3 *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.

2.8. 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.

- c) 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:

1. Si $n=2$,

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

2. Si A tiene una fila o una columna de ceros entonces $|A| = 0$.

3. 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:

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

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

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

$$\left| \begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ \lambda a_{i1} & \lambda a_{i2} & \cdots & \lambda a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{array} \right| = \lambda \left| \begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & \vdots & \vdots \\ a_{i1} & a_{i2} & \cdots & a_{in} \\ \vdots & \vdots & \vdots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{array} \right|$$

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 invertible si y sólo si $|A| \neq 0$. Además, en ese caso, $|A^{-1}| = 1/|A|$.

Prueba de la propiedad 5.

Si A es invertible, 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 invertible 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 invertible porque $\text{rg}(A) = n$. \square

2.9. Formas cuadráticas.

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

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

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

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

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

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 $\omega(x) = x^t A x$, $\forall x \in \mathbb{R}^n$. Esta matriz se llama matriz asociada a la forma cuadrática.

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:

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

Formas cuadráticas degeneradas y no degeneradas

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 $\omega(x) = x^t A x$, $\forall x \in \mathbb{R}^n$. Se dice que ω es **no degenerada** si $\text{rg}(A) = n$, es decir, si $|A| \neq 0$. Si

el determinante de A es cero entonces se dice que la forma cuadrática ω es degenerada.

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 no degenerada porque $\omega(x) = x^tAx$, con

$$|A| = \begin{vmatrix} 2 & -2 & 1 \\ -2 & 3 & -1 \\ 1 & -1 & 1 \end{vmatrix} = 1 \neq 0.$$

Clasificación de formas cuadráticas no degeneradas.

Las formas cuadráticas no degeneradas $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ pueden ser de tres tipos.

- (a) ω es **definida positiva** si $\omega(x) = x^tAx > 0, \forall x \neq 0$,
- (b) ω es **definida negativa** si $\omega(x) = x^tAx < 0, \forall x \neq 0$,
- (c) ω es **indefinida** si existen dos vectores $x, y \in \mathbb{R}^n$ tales que $\omega(x) > 0, \omega(y) < 0$.

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

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

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

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 los menores principales.

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

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:

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

Teorema 2.2 *Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica. Entonces A es definida positiva si y sólo si todos los menores principales de A son mayores que cero.*

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

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

Los menores principales de A son:

$$\Delta_1 = 2 > 0 ; \quad \Delta_2 = \begin{vmatrix} 2 & -2 \\ -2 & 3 \end{vmatrix} = 2 > 0 ; \quad \Delta_3 = \begin{vmatrix} 2 & -2 & 1 \\ -2 & 3 & -1 \\ 1 & -1 & 1 \end{vmatrix} = 1 > 0.$$

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 2.4 *Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica. A es definida negativa si y sólo si los menores principales de orden impar son menores que cero y los de orden par son mayores que cero, es decir, $\Delta_1 < 0$, $\Delta_2 > 0$, $\Delta_3 < 0$, \dots*

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

Teorema 2.3 *Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ una matriz simétrica tal que $|A| \neq 0$. Entonces la forma cuadrática $\omega(x) = x^t A x$ se clasifica en función de los menores principales del siguiente modo:*

- (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.*

Formas cuadráticas degeneradas.

Las formas cuadráticas degeneradas $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ pueden ser de tres tipos.

1. ω es **semidefinida positiva** si $\omega(x) = x^t Ax \geq 0, \forall x \in \mathbb{R}^n$,
2. ω es **semidefinida negativa** si $\omega(x) = x^t Ax \leq 0, \forall x \in \mathbb{R}^n$,
3. ω es **indefinida** si existen dos vectores $x, y \in \mathbb{R}^n$ tales que $\omega(x) > 0, \omega(y) < 0$.

En este caso la clasificación no se puede deducir directamente de los menores principales. Volveremos sobre esta cuestión en el tema 5.

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:

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

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.

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.

Teorema 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 (incógnitas libres). 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 las $(n - r)$ incógnitas libres.

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,

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

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

En otras palabras, el núcleo de A es el subespacio de \mathbb{R}^n generado por los vectores u_1, u_2, \dots, u_k :

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

y $\boxed{\dim(\text{Ker}(A)) = n - \text{rg}(A)}$.

Resolver el sistema homogéneo $Ax = 0$ en el caso compatible indeterminado equivale a calcular una base del núcleo de A .

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:

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

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

Como $\text{rg}(A) = \text{rg}(A') = 2 < 4 =$ 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

$$\begin{cases} \boxed{x} + 2z + 2t = 0 \\ \boxed{y} - z - t = 0 \end{cases}$$

Despejando las incógnitas x e y correspondientes a las entradas principales en función de las incógnitas libres 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 es el subespacio de \mathbb{R}^4 de dimensión 2 formado por todas 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:

Teorema 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:

$$\begin{aligned}
 (A|b) &= \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 1 & 2 & 0 & 1 \\ 1 & 0 & 2 & 1 \end{array} \right) \xrightarrow{F_{21}(-1)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 1 & 0 & 2 & 1 \end{array} \right) \xrightarrow{F_{31}(-1)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & -1 & 1 & 0 \end{array} \right) \\
 &\xrightarrow{F_{32}(1)} \left(\begin{array}{ccc|c} 1 & 1 & 1 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) \xrightarrow{F_{12}(-1)} \left(\begin{array}{ccc|c} 1 & 0 & 2 & 1 \\ 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 0 \end{array} \right) = (A'|b').
 \end{aligned}$$

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 es

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

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 3.1 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 $\text{rg}(A|b) = n$, ya que la matriz $(A|b)$ tiene n filas. \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. $\det(A) \neq 0$.

Observación: Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es inversible, 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 .

Proposición 3.2 *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,

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

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:

$$\begin{aligned} & \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \xrightarrow{\substack{F_{43}(2), F_{42}(-2) \\ F_{41}(-1)}}} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ -1 & -2 & 2 & 1 \end{pmatrix} \\ & \xrightarrow{F_{31}(-1)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & -2 & 2 & 1 \end{pmatrix} \xrightarrow{F_{21}(2)} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 2 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & -2 & 2 & 1 \end{pmatrix} = L. \end{aligned}$$

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 $Ux = z$:

$$\begin{aligned} 4x_4 &= -4 \implies x_4 = -1 \\ -x_3 + x_4 &= -4 \implies x_3 = 3 \\ -2x_2 + x_3 + 3x_4 &= -4 \implies x_2 = 2 \\ 2x_1 - x_2 + x_4 &= -5 \implies x_1 = -1 \end{aligned}$$

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

3.6. Mínimos cuadrados. Ajuste.

Consideremos un sistema de ecuaciones lineales $Ax = b$, donde $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $b \in \mathbb{R}^p$. Se define la **imagen** de A , y se denota por $\text{Im}(A)$, como el subespacio generado por las columnas de A . La compatibilidad del sistema $Ax = b$ se caracteriza en términos de la imagen de A de forma sencilla.

Proposición 3.3 *El sistema $Ax = b$ es compatible si y sólo si $b \in \text{Im}(A)$.*

La proposición anterior dice que $Ax = b$ es compatible sólo cuando b es combinación lineal de las columnas de A . En el caso de que el sistema sea incompatible, se puede buscar una “solución aproximada”. Una posibilidad es determinar el vector $b' \in \text{Im}(A)$ cuya distancia al término independiente b sea la menor posible. Los vectores $x \in \mathbb{R}^n$ tales que $Ax = b'$ serán lo que llamaremos soluciones del sistema $Ax = b$ en el sentido de mínimos cuadrados.

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 cumple la siguiente igualdad:

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

La distancia mínima de b a la imagen de A es la distancia de b a la proyección ortogonal de b sobre $\text{Im}(A)$, es decir, al único vector $b' \in \text{Im}(A)$ tal que $(b - b')$ es ortogonal a todos los vectores de la imagen de A . 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 a las columnas de A . Esto es equivalente a la relación

$$A^t(Ax_0 - b) = 0.$$

Por lo tanto, se cumple el siguiente resultado:

Teorema 3.3 *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*

$$A^tAx_0 = A^tb.$$

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

Teorema 3.4 *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.*
- (3) *El sistema $A^tAx = A^tb$ tiene solución única si y sólo si $\text{rg}(A) = n$.*

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

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

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

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

El sistema de mínimos cuadrados $A^tAx = A^tb$ 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 $y = \frac{5}{2} + \frac{2}{5}x$.

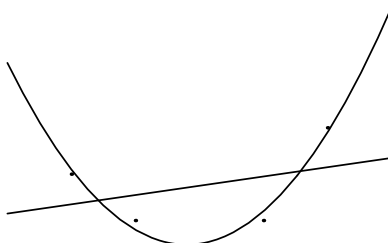


Figura 3.1: Aproximaciones lineal y cuadrática de los datos.

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 3.1 se representan los puntos y las aproximaciones lineal y cuadrática. Se observa que ésta última es mucho más precisa.

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, que generalizan a los ya conocidos para \mathbb{R}^n . También se interpretan las matrices como aplicaciones lineales.

4.2. Espacios y subespacios vectoriales.

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.

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.

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

4. El conjunto $\mathcal{C}_0(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R} / f \text{ es continua}\}$ es un espacio vectorial real con las operaciones habituales de suma de funciones y producto de un escalar por una función.

Muchos de los conceptos definidos para \mathbb{R}^n se extienden a otros espacios vectoriales. A continuación repasamos algunos.

Subespacios vectoriales.

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) $\lambda u \in U, \forall \lambda \in \mathbb{R}, \forall u \in U$.

Ejemplos:

1. 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 .
2. El conjunto $U = \{A \in \mathcal{M}_{n \times n}(\mathbb{R}) / A^t = A\}$ es un subespacio vectorial de $\mathcal{M}_{n \times n}(\mathbb{R})$.
3. 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,

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

Al igual que en \mathbb{R}^n , 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

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

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

Tenemos la siguiente caracterización de los subespacios vectoriales:

Propiedad: 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 .

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 . Si S es un conjunto de generadores de U , diremos que U es el subespacio generado por S .

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.

Los conceptos de dependencia e independencia lineal se extienden de manera natural a cualquier espacio vectorial.

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

Un conjunto de vectores es **linealmente independiente** si ninguno de ellos es combinación lineal del resto. Se llama **rango** de un conjunto de vectores al número de vectores linealmente independientes que contiene. Por tanto, un conjunto de n vectores es linealmente independiente si y sólo si su rango es n .

Si $V = \mathbb{R}^n$ entonces estudiar si un conjunto de vectores S es libre se reduce a calcular el rango de la matriz que tiene como filas los vectores de S : un conjunto $S = \{v_1, v_2, \dots, v_p\}$ de vectores de \mathbb{R}^n es libre si y sólo si

$$\text{rg} \begin{pmatrix} \frac{v_1^t}{v_2^t} \\ \vdots \\ v_p^t \end{pmatrix} = p.$$

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

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

Por tanto, S no es libre.

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,

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

Para otros espacios vectoriales, resulta útil la siguiente caracterización de la independencia lineal:

Proposición 4.1 *Un conjunto $S = \{v_1, v_2, \dots, v_n\}$ de vectores es linealmente independiente si y sólo si se cumple la siguiente propiedad:*

“Si $\lambda_1, \dots, \lambda_n$ son números reales tales que $\lambda_1 v_1 + \lambda_2 v_2 + \dots + \lambda_n v_n = 0$ entonces necesariamente $\lambda_1 = \lambda_2 = \dots = \lambda_n = 0$.”

Por ejemplo, el conjunto

$$S = \left\{ \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 1 \\ 4 & 0 \end{pmatrix} \right\}$$

es libre porque

$$\alpha \begin{pmatrix} 2 & 1 \\ 3 & 1 \end{pmatrix} + \beta \begin{pmatrix} 0 & 1 \\ 2 & 1 \end{pmatrix} + \gamma \begin{pmatrix} 1 & 1 \\ 4 & 0 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} \iff \begin{pmatrix} 2 & 0 & 1 \\ 1 & 1 & 1 \\ 3 & 2 & 4 \\ 1 & 1 & 0 \end{pmatrix} \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

y la única solución del sistema es $(\alpha, \beta, \gamma) = (0, 0, 0)$ porque el rango de la matriz de coeficientes coincide con el número de incógnitas.

4.4. Bases y dimensión.

Un conjunto linealmente independiente de generadores de un espacio vectorial V se llama **base** de V .

Ejemplos:

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

2. 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.

Todas las bases de un espacio vectorial V tienen el mismo número de vectores. El número de vectores de cualquier base de V se llama **dimensión** de V y se denota por $\dim(V)$.

Ejemplos:

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

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

En general, $\dim(\mathcal{M}_{p \times n}(\mathbb{R})) = p \cdot n$.

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

Cálculo de la dimensión de un subespacio vectorial

- 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

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

- Ya sabemos que si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, entonces $\text{Ker}(A)$ es un subespacio de \mathbb{R}^n y

$$\dim(\text{Ker}(A)) = n - \text{rg}(A).$$

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 4.2 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 \mathbb{R}^n 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 $P_{\mathcal{B}\mathcal{C}} \in \mathcal{M}_{n \times n}(\mathbb{R})$ cuyas columnas son los vectores de \mathcal{B} , es decir,

$$P_{\mathcal{B}\mathcal{C}} = (u_1 | u_2 | \dots | u_n).$$

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

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

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

Proposición 4.3 Si $P_{\mathcal{B}\mathcal{C}}$ 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:

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

De modo análogo, si \mathcal{B} y \mathcal{B}' son dos bases de \mathbb{R}^n se define la matriz de cambio de base $P_{\mathcal{B}'\mathcal{B}}$ de \mathcal{B}' a \mathcal{B} como la que tiene la siguiente propiedad:

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

El cambio de base de \mathcal{B}' a \mathcal{B} se puede hacer utilizando las siguientes propiedades:

Proposición 4.4 Sean \mathcal{B} y \mathcal{B}' dos bases de \mathbb{R}^n . Entonces:

1. $P_{\mathcal{B}\mathcal{C}}$ es inversible y $(P_{\mathcal{B}\mathcal{C}})^{-1} = P_{\mathcal{C}\mathcal{B}}$.
2. $P_{\mathcal{B}'\mathcal{B}} = P_{\mathcal{C}\mathcal{B}} P_{\mathcal{B}'\mathcal{C}} = (P_{\mathcal{B}\mathcal{C}})^{-1} P_{\mathcal{B}'\mathcal{C}}$.

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. Bases ortonormales.

Una base $\mathcal{B} = \{u_1, u_2, \dots, u_p\}$ de un subespacio vectorial U de \mathbb{R}^n es una base ortonormal si todos los vectores son unitarios y ortogonales entre sí, es decir, $u_i^t u_j = 0$ si $i \neq j$ y $u_i^t u_i = 1$ para todo $i = 1, 2, \dots, p$.

El **procedimiento de ortonormalización de Gram-Schmidt** permite calcular una base ortonormal a partir de una base de U . Sea $\mathcal{B} = \{v_1, v_2, \dots, v_p\}$ una base de un subespacio vectorial U de \mathbb{R}^n . Es posible construir una base ortonormal $T = \{u_1, u_2, \dots, u_p\}$ de U a partir de \mathcal{B} 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:

$$\tilde{u}_i = v_i - \sum_{j=1}^{i-1} (v_i^t u_j) u_j = v_i - (v_i^t u_1) u_1 - \cdots - (v_i^t u_{i-1}) u_{i-1}.$$

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

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

Ejemplo: Vamos a calcular una base ortonormal del subespacio $U = \langle \{(1, 0, 1), (1, 1, 1)\} \rangle$.

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

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

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

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

El conjunto $T = \{u_1, u_2\} = \left\{ \left(\frac{1}{\sqrt{2}}, 0, \frac{1}{\sqrt{2}} \right), (0, 1, 0) \right\}$ es una base ortonormal de U .

El siguiente resultado relaciona las bases ortonormales con las matrices ortogonales y será de utilidad en el Capítulo 5.

Proposición 4.5 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 . En particular, si \mathcal{B} es una base ortonormal, la matriz de cambio de coordenadas $P_{\mathcal{B}\mathcal{C}}$ es una matriz ortogonal.

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,

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

□

4.7. Definición de aplicación lineal y matriz asociada.

Sean V y W dos espacios vectoriales. Una aplicación $L : V \rightarrow W$ es lineal si cumple las siguientes propiedades:

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

De estas propiedades se obtiene por inducción que

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

donde $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{R}, v_1, v_2, \dots, v_n \in V.$ En otras palabras, si $L : V \rightarrow W$ es una aplicación lineal entonces la imagen de la combinación lineal de n vectores de V es igual a la combinación lineal de sus imágenes.

Matriz asociada a una aplicación lineal.

Una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ define una aplicación lineal $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ dada por $L(x) = Ax$, donde $x \in \mathbb{R}^n$ es un vector columna. Recíprocamente, el siguiente resultado prueba 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 $A \in \mathcal{M}_{p \times n}(\mathbb{R})$.

Proposición 4.6 *Dada una aplicación lineal $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$, 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 se llama **matriz asociada** a L y 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

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

4.8. Núcleo e imagen de una aplicación lineal.

Sea $L : V \rightarrow W$ una aplicación lineal. Se define el **núcleo** de L como

$$\text{Ker}(L) = \{x \in V / L(x) = 0\}.$$

Si $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ y A es su matriz asociada a L entonces es claro que

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

La **imagen** de L se define como el subespacio formado por todos los vectores de W que son imagen de algún vector de V por la aplicación L :

$$\text{Im}(L) = \{L(x) / x \in V\}.$$

Proposición 4.7 Sea $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ una aplicación lineal y $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ su matriz asociada. Entonces $\text{Im}(L) = \text{Im}(A)$, es decir, la imagen de L está generada por las columnas de A .

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

Teniendo en cuenta que $A = (L(e_1) | L(e_2) | \dots | L(e_n))$ y que L es una aplicación lineal:

$$\begin{aligned} b \in \text{Im}(L) &\iff \exists x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n / b = L(x) \iff \\ &\iff b = L(x_1 e_1 + x_2 e_2 + \dots + x_n e_n) = x_1 L(e_1) + x_2 L(e_2) + \dots + x_n L(e_n) \iff \\ &\iff b \in \langle \{L(e_1), L(e_2), \dots, L(e_n)\} \rangle \iff b \in \text{Im}(A). \quad \square \end{aligned}$$

La fórmula de las dimensiones $\dim(\text{Ker}(A)) + \text{rg}(A) = n$ para una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ se reescribe para una aplicación lineal $L : \mathbb{R}^n \rightarrow \mathbb{R}^p$ como

$$\dim(\text{Ker}(L)) + \dim(\text{Im}(L)) = n = \dim(\mathbb{R}^n).$$

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 :

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

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

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

4.9. Inversas de aplicaciones lineales.

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

Propiedad: 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

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

4.10. Transformaciones ortogonales.

Se dice que una aplicación lineal $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es una transformación ortogonal si conserva el producto escalar, es decir, si para cada par de vectores x e y de \mathbb{R}^n se cumple que

$$(L(x))^t L(y) = x^t y.$$

Observemos que si A es la matriz asociada a la aplicación L entonces

$$(L(x))^t L(y) = (Ax)^t Ay = x^t A^t Ay.$$

De esta relación se obtiene el siguiente resultado que caracteriza las transformaciones ortogonales:

Proposición 4.8 *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 una transformación ortogonal si y sólo si A es una matriz ortogonal.*

Es fácil probar que las transformaciones ortogonales conservan la norma, la distancia y el ángulo. Por esta razón se suelen llamar **movimientos rígidos**. En \mathbb{R}^2 las únicas transformaciones ortogonales son giros o simetrías respecto a un eje.

4.11. Proyección ortogonal.

Sea $b \in \mathbb{R}^n$ y sea U un subespacio de \mathbb{R}^n con $\dim(U) = p < n$. Se llama **proyección ortogonal** de b sobre el subespacio U al único vector $b' \in U$ tal que $(b - b')$ es ortogonal a U . La norma del vector $b - b'$ representa la **mínima distancia** de b al subespacio U , es decir, $d(b, U) = \|b - b'\|$.

La proyección ortogonal se puede considerar como una aplicación lineal de \mathbb{R}^n en \mathbb{R}^n cuya matriz asociada se llama matriz de proyección ortogonal. El siguiente resultado permite construir la matriz de proyección ortogonal sobre un subespacio U a partir de una base ortonormal.

Proposición 4.9 *Sea U un subespacio vectorial de \mathbb{R}^n de dimensión p y sea $\mathcal{B} = \{u_1, \dots, u_p\}$ una base ortonormal de U . Si $A = (u_1 | u_2 | \dots | u_p)$, entonces la matriz de proyección ortogonal sobre U es*

$$P = AA^t = (u_1 | u_2 | \dots | u_p) \begin{pmatrix} \frac{u_1^t}{|u_1|} \\ \frac{u_2^t}{|u_2|} \\ \vdots \\ \frac{u_p^t}{|u_p|} \end{pmatrix} = u_1 u_1^t + u_2 u_2^t + \dots + u_p u_p^t.$$

Demostración. Tenemos que probar que Pb es la proyección ortogonal de b sobre U para cada vector $b \in \mathbb{R}^n$. En primer lugar, $Pb = u_1(u_1^t b) + u_2(u_2^t b) + \dots + u_p(u_p^t b) \in U$ por ser combinación lineal de vectores de una base de U .

Por otra parte, $(b - Pb)$ 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:

$$u_1^t(Pb) = u_1^t(u_1 u_1^t b + u_2 u_2^t b + \dots + u_p u_p^t b) = (u_1^t u_1) u_1^t b + (u_1^t u_2) u_2^t b + \dots + (u_1^t u_p) u_p^t b = u_1^t b.$$

Por tanto, $u_1^t(b - Pb) = u_1^t b - u_1^t(Pb) = 0$.

Del mismo modo se prueba para u_2, \dots, u_p . \square

Nótese que $\text{rg}(P) = \dim(U) = p$ ya que U es la imagen de la aplicación de proyección ortogonal.

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 :

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

$$\tilde{u}_2 = v_2 - (v_2^t u_1) u_1 = \begin{pmatrix} 0 \\ 1 \\ 1 \end{pmatrix} - \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \end{pmatrix} = \begin{pmatrix} -1/2 \\ 1 \\ 1/2 \end{pmatrix};$$

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

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: $\dim(U) = 1$.

Sea u un vector unitario y sea $U = \langle \{u\} \rangle$. La matriz de proyección ortogonal sobre U es $P = uu^t$. En este caso P tiene rango 1.

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,

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

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. También se introduce el concepto de valor singular y su aplicación en la obtención de la mejor aproximación de rango k de una matriz.

5.2. Autovalores y autovectores.

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 $x \in \mathbb{C}^n$.

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 $\text{Sp}(A)$.

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

$$x^2 - 2x - 3 = 0 \iff x = \frac{2 \pm \sqrt{16}}{2} = \frac{2 \pm 4}{2},$$

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.

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 $p_1(\lambda) \neq 0$.

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 $\alpha_1 + \alpha_2 + \cdots + \alpha_r = n$.

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y sea $\lambda \in \text{Sp}(A)$. 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. (λ) .

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.

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

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

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

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

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

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

Si $\lambda \in \text{Sp}(A)$, 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 $\lambda \in \text{Sp}(A)$. Entonces $1 \leq \text{m.g.}(\lambda) \leq \text{m.a.}(\lambda) \leq n$.*

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:

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 última 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.

Una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ 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

$$A = PDP^{-1} \iff AP = PD \iff (Au_1|Au_2|\dots|Au_n) = (\lambda_1u_1|\lambda_2u_2|\dots|\lambda_nu_n).$$

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 de matrices.

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

Ejemplo: Hallar la expresión de A^k para la matriz

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

En este caso $Sp(A) = \{0, 3\}$, con $\text{m.a.}(0) = \text{m.g.}(0) = 2$. 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 $A = PDP^{-1}$. Por tanto,

$$\begin{aligned} A^k &= PD^kP^{-1} = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 3^k \end{pmatrix} \begin{pmatrix} 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \\ 1/3 & 1/3 & 1/3 \end{pmatrix} = \\ &= 3^{k-1} \begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix}. \end{aligned}$$

5.4. Diagonalización ortogonal.

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

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 5.3 (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.*

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 , pero necesitamos además que P sea ortogonal. En virtud de la propiedad probada en la sección 4.6, las columnas de A deben ser una base ortonormal. La siguiente propiedad hace que sea posible calcular una base ortonormal formada por autovectores:

Lema 5.1 *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 $\lambda_1, \lambda_2 \in \mathbb{R}$:

$$\lambda_1 x_1^t x_2 = (\lambda_1 x_1)^t x_2 = (Ax_1)^t x_2 = x_1^t A^t x_2 = x_1^t A x_2 = x_1^t \lambda_2 x_2 = \lambda_2 x_1^t x_2.$$

Por tanto, $\lambda_1 x_1^t x_2 = \lambda_2 x_1^t x_2$. Como $\lambda_1 \neq \lambda_2$, necesariamente $x_1^t x_2 = 0$. \square

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 $A = PDP^t$.

5.5. Clasificación de formas cuadráticas usando la diagonalización ortogonal.

Sea $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ una forma cuadrática definida por $\omega(x) = x^t Ax$, donde A es una matriz simétrica. La clasificación de la forma cuadrática se puede hacer utilizando la diagonalización ortogonal de A .

Como A es ortogonalmente diagonalizable, 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:

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

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

$$\omega(y) = y^t Dy = \lambda_1 y_1^2 + \lambda_2 y_2^2 + \cdots + \lambda_n y_n^2 = \sum_{i=1}^n \lambda_i y_i^2,$$

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 5.4 *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 $\lambda \leq 0, \forall \lambda \in \text{Sp}(A)$.

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 semidefinida negativa ya que $\text{Sp}(A) = \{0, -3\}$, con m.a.(0) = 1, m.a.(-3) = 2.

5.6. 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:

$$\begin{aligned} A = PDP^t &= (u_1|u_2|\dots|u_n) \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix} \begin{pmatrix} \frac{u_1^t}{\|u_1\|^2} \\ \frac{u_2^t}{\|u_2\|^2} \\ \vdots \\ \frac{u_n^t}{\|u_n\|^2} \end{pmatrix} = \\ &= \lambda_1 u_1 u_1^t + \lambda_2 u_2 u_2^t + \dots + \lambda_n u_n u_n^t = \lambda_1 u_1 u_1^t + \lambda_2 u_2 u_2^t + \dots + \lambda_r u_r u_r^t, \end{aligned}$$

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

De esta manera se descompone A en la suma de r matrices $A_i = \lambda_i u_i u_i^t$ 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.

5.7. 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 5.3 *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 $A^t A x = \lambda x$ y por tanto:

$$\|Ax\|^2 = (Ax)^t (Ax) = x^t A^t A x = \lambda x^t x = \lambda \|x\|^2 \implies \lambda = \frac{\|Ax\|^2}{\|x\|^2} \geq 0.$$

□

Sean $A \in \mathcal{M}_{p \times n}(\mathbb{R})$. Se llaman **valores singulares** de A a las raíces cuadradas positivas de los autovalores 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 5.5 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*

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

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 5.6 *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*

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

Esta expresión resulta útil para definir la aproximación de rango k de una matriz. Sea $A = \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \dots + \sigma_r u_r v_r^t$ 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,

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

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 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 $\text{rg}(A) = r$. El cálculo de la SVD de A parte de la diagonalización ortogonal de la matriz simétrica $A^t A$. Si denotamos $A^t A = V D V^t$ a esta diagonalización ortogonal y factorizamos $D = \Sigma^t \Sigma$, entonces, para cualquier matriz ortogonal $U \in \mathcal{M}_{p \times p}(\mathbb{R})$:

$$A^t A = V D V^t = V \Sigma^t \Sigma V^t = V \Sigma^t U^t U \Sigma V^t = (U \Sigma V^t)^t (U \Sigma V^t).$$

Teniendo esto en cuenta, la SVD se calcula en tres etapas:

- (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 :

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

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

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:

$$u_1 = \frac{1}{\sigma_1} Av_1 = \frac{1}{3} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} 1/3 \\ 0 \\ 2/3 \\ 2/3 \end{pmatrix};$$

$$u_2 = \frac{1}{\sigma_2} Av_2 = \frac{1}{2} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ -1/\sqrt{2} \\ 1/\sqrt{2} \end{pmatrix};$$

$$u_3 = \frac{1}{\sigma_3} Av_3 = \frac{1}{\sqrt{2}} \begin{pmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ -1 & 1 & 2 \\ 1 & -1 & 2 \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 1 \\ 0 \\ 0 \end{pmatrix}.$$

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:

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

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

5.8. Teorema de Cayley-Hamilton.

El objetivo de esta sección es definir algunas funciones reales sobre matrices y dar un método para calcularlas. Comenzamos definiendo **polinomios de matrices**.

Sea $A \in \mathcal{M}_{n \times n}(\mathbb{R})$. Sea $p(x) = a_0 + a_1x + a_2x^2 + \cdots + a_kx^k$ un polinomio. 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 5.7 (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:

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

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

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

Esto quiere decir que los autovalores múltiples proporcionan tantas ecuaciones como indica su multiplicidad algebraica. Esta propiedad permite calcular $r(x)$ resolviendo un sistema de n ecuaciones lineales cuyas n incógnitas son los coeficientes del polinomio

$$r(x) = a_0 + a_1x + \dots + a_{n-1}x^{n-1}.$$

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

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

Como los autovalores de A son $\lambda_1 = \lambda_2 = 0$, $\lambda_3 = 1$, el polinomio $r(x) = a + bx + cx^2$ de grado 2 debe cumplir las relaciones:

$$\begin{aligned} r(0) &= a = p(0) = 1 \\ r'(0) &= b = p'(0) = -2 \\ r(1) &= a + b + c = p(1) = 0. \end{aligned}$$

La única solución del sistema es $a = 1$, $b = -2$, $c = 1$ y por tanto $r(x) = 1 - 2x + x^2$. Es decir, $p(A) = r(A) = I - 2A + A^2$.

5.9. 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. Estas funciones son límites de polinomios y eso permite calcular las 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

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

El conjunto

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

se llama conjunto de valores de f sobre el espectro de A .

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,

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

Como antes, 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 $\text{m.a.}(0) = 2$, para calcular $f(A) = A^{1/2}$ necesitamos determinar los valores de $f(0)$ y $f'(0)$.

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

Funciones de matrices usando la diagonalización.

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

Proposición 5.4 *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 5.5 Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es diagonalizable, es decir, $A = PDP^{-1}$ con D diagonal, entonces $f(A) = Pf(D)P^{-1}$.

Autovalores de $f(A)$.

Para terminar el tema damos una propiedad que resulta de utilidad.

Proposición 5.6 Si $\lambda_1, \lambda_2, \dots, \lambda_n$ son los autovalores de A (contados con su multiplicidad) entonces los autovalores de $f(A)$ son $f(\lambda_1), f(\lambda_2), \dots, f(\lambda_n)$.

Por ejemplo, si $\text{Sp}(A) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ entonces $\text{Sp}(A^k) = \{\lambda_1^k, \lambda_2^k, \dots, \lambda_n^k\}$, $\forall k \in \mathbb{N}$ y, si A es inversible, $\text{Sp}(A^{-1}) = \{1/\lambda_1, 1/\lambda_2, \dots, 1/\lambda_n\}$.

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

$$\det(f(A)) = f(\lambda_1)f(\lambda_2) \cdots f(\lambda_n)$$

$$\text{tr}(f(A)) = f(\lambda_1) + f(\lambda_2) + \cdots + f(\lambda_n).$$

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” (4^a ed.), Pearson Educación, 2012.
- G. NAKOS Y D. JOYNER, “Álgebra Lineal con aplicaciones”, Thomson, 1999.
- D. POOLE, “Álgebra Lineal con aplicaciones” (2^a ed.), Thomson, 2007.