

Apuntes de álgebra lineal

EDUARDO LIZ MARZÁN

Septiembre de 2022.

Índice general

1. Conceptos preliminares	7
1.1. Operaciones con números reales y números complejos	7
1.2. El espacio vectorial \mathbb{R}^n	11
2. Matrices y determinantes	15
2.1. Introducción	15
2.2. Definición y tipos de matrices	15
2.3. Operaciones con matrices	17
2.4. Trasposición de matrices	20
2.5. Matrices elementales	21
2.6. Forma escalonada y rango de una matriz	23
2.7. Cálculo de la inversa	25
2.8. Determinantes	26
2.9. Formas cuadráticas	28
3. Sistemas de ecuaciones lineales	31
3.1. Introducción	31
3.2. Expresión matricial	31
3.3. Existencia de soluciones	32
3.4. Sistemas compatibles determinados	33
3.5. Sistemas compatibles indeterminados	33
3.6. Soluciones de un sistema en el sentido de mínimos cuadrados	36
4. Espacios vectoriales y aplicaciones lineales	41
4.1. Introducción	41
4.2. Espacios y subespacios vectoriales	41
4.3. Independencia lineal	43
4.4. Bases y dimensión	43
4.5. Cambio de base en \mathbb{R}^n	44
4.6. Bases ortonormales	46
4.7. Aplicaciones lineales	48
4.8. Transformaciones ortogonales.	49
4.9. Proyección ortogonal	50

5. Diagonalización y funciones de matrices	53
5.1. Introducción.	53
5.2. Autovalores y autovectores.	53
5.3. Matrices diagonalizables	57
5.4. Diagonalización ortogonal	58
5.5. Descomposición espectral	60
5.6. Descomposición en valores singulares	62
5.7. Polinomios de matrices y el teorema de Cayley-Hamilton	63
5.8. Funciones de matrices	65
Referencias	69

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.

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, septiembre de 2022.

Capítulo 1

Conceptos preliminares

1.1. Operaciones con números reales y números complejos

En esta asignatura nos interesa el conjunto \mathbb{R} de los números reales como un conjunto de números con los que se puede operar (en la asignatura de Cálculo, la recta real juega un papel “más dinámico”, como un conjunto en el que se mueve una magnitud escalar, como el tiempo).

Como es bien sabido, en \mathbb{R} se consideran dos operaciones internas (es decir, que asignan a cada par de números reales un resultado que también es un número real): la suma (+) y el producto (\cdot). Estas operaciones tienen muy buenas propiedades, que en general usamos de modo inconsciente. Como en el Paraíso, solo hay un fruto prohibido: **¡no se puede dividir por cero!**

Recordamos las propiedades principales:

1. Propiedad asociativa: $(x + y) + z = x + (y + z)$, $\forall x, y, z \in \mathbb{R}$. Esta propiedad permite sumar más de dos elementos sin usar paréntesis. Escribiremos simplemente $x + y + z$. El producto de números reales tiene la misma propiedad.
2. Propiedad conmutativa.
 - Propiedad conmutativa de la suma: $x + y = y + x$, $\forall x, y \in \mathbb{R}$.
 - Propiedad conmutativa del producto: $x \cdot y = y \cdot x$, $\forall x, y \in \mathbb{R}$.

3. Elemento neutro:

En la suma el elemento neutro es el cero (0) y cumple que $0 + x = x$, $\forall x \in \mathbb{R}$.

En el producto el elemento neutro es el uno (1) y cumple que $1 \cdot x = x$, $\forall x \in \mathbb{R}$.

4. Elemento opuesto y elemento inverso:

En el caso de la suma, todo número real x tiene opuesto, que se denota por $-x$ y cumple que $x + (-x) = 0$. Hacer la resta $x - y$ consiste en sumar a x el opuesto de y .

En el caso del producto, todo número $x \neq 0$ tiene inverso, que se denota por $1/x$ y cumple que $x \cdot (1/x) = 1$. Cualquier número real x se puede dividir por otro número real $y \neq 0$:

$$\frac{x}{y} = x \cdot \frac{1}{y}, \quad \forall y \neq 0.$$

5. Propiedad distributiva (sacar factor común): $x \cdot (y + z) = x \cdot y + x \cdot z$, $\forall x, y, z \in \mathbb{R}$.

Con estas propiedades, se dice que \mathbb{R} es un cuerpo conmutativo con las operaciones de suma y producto.

El conjunto de los números reales tiene muy buenas propiedades pero no permite resolver algunas ecuaciones simples como $x^2 + 1 = 0$. Para resolver este problema necesitamos un conjunto de números más grande, que será el conjunto \mathbb{C} de los números complejos (por supuesto, esta no es la única utilidad de los números complejos).

Se define un **número complejo** como 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**.

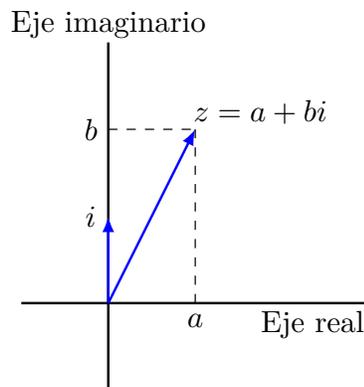


Figura 1.1: Representación de un número complejo $z = a + bi$ en el plano complejo. La unidad imaginaria i se sitúa en el eje imaginario.

Operaciones en \mathbb{C}

- **Suma.** Si $z_1 = a_1 + b_1i$ y $z_2 = a_2 + b_2i$ son dos números complejos, se define su suma como

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

Es decir, la parte real de la suma es la suma de las partes reales y lo mismo para las partes imaginarias.

- **Producto.** El producto de números complejos es un producto habitual de binomios, 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} 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

$$(a + bi)(a - bi) = a^2 - (bi)^2 = a^2 + b^2 \implies \frac{(a + bi)(a - bi)}{a^2 + b^2} = 1 \implies \frac{1}{a + bi} = \frac{a - bi}{a^2 + b^2},$$

que está bien definido para todo $z = a + bi \neq 0$.

Por ejemplo,

$$\frac{1}{1 - 2i} = \frac{1 + 2i}{1^2 + (-2)^2} = \frac{1 + 2i}{5} = \frac{1}{5} + \frac{2}{5}i.$$

Esto permite hacer la división de dos números complejos z_1/z_2 siempre que $z_2 \neq 0$. Se suele realizar multiplicando y dividiendo por el conjugado del denominador. Por ejemplo:

$$\frac{2 + 3i}{1 - 2i} = \frac{(2 + 3i)(1 + 2i)}{(1 - 2i)(1 + 2i)} = \frac{-4 + 7i}{5} = \frac{-4}{5} + \frac{7}{5}i.$$

Módulo y argumento

Si $z = a + bi \in \mathbb{C}$, se define el **módulo** de z como el número real $|z| = \sqrt{a^2 + b^2}$ (se considera siempre la raíz cuadrada positiva). El módulo de $z = a + bi$ representa la distancia entre los puntos (a, b) y $(0, 0)$ en el plano complejo. Usando el módulo, el inverso de un número complejo $z \neq 0$ se expresa como

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

Se define el **argumento** de $z = a + bi$ como el ángulo $\alpha \in (-\pi, \pi]$ que cumple las igualdades

$$|z| \cos(\alpha) = a \quad ; \quad |z| \operatorname{sen}(\alpha) = b.$$

De este modo,

$$z = a + bi = |z|(\cos(\alpha) + \operatorname{sen}(\alpha)i),$$

que es la llamada **forma trigonométrica** de z . El argumento representa el ángulo que forma el vector (a, b) en el plano complejo con la semirrecta positiva del eje real (ver la figura 1.2).

Forma exponencial, potencias, raíces n -ésimas

Usando desarrollos en serie de potencias de las funciones exponencial, seno y coseno (se verá en la asignatura de Cálculo), se obtiene la siguiente fórmula, conocida como **fórmula de Euler**:

$$e^{\alpha i} = \cos(\alpha) + \operatorname{sen}(\alpha)i, \quad \forall \alpha \in \mathbb{R}.$$

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

Las fórmulas y la interpretación para el producto y las potencias de números complejos resultan más sencillas cuando se utiliza la forma exponencial. En particular, el módulo del producto de z_1 y z_2 es el producto de los módulos y el argumento del producto es la suma de los argumentos.

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

El número complejo de módulo 1 y argumento α se escribe en forma exponencial como $e^{\alpha i}$ (ver la figura 1.2). Geométricamente, multiplicar por $e^{\alpha i}$ equivale a hacer una rotación de ángulo α en el plano complejo. Como curiosidad, de esta manera se puede interpretar la fórmula trigonométrica $\text{sen}^2(\alpha) + \text{cos}^2(\alpha) = 1$ del siguiente modo:

$$e^{\alpha i} e^{-\alpha i} = e^{\alpha i - \alpha i} = e^0 = 1 \iff (\cos(\alpha) + \text{sen}(\alpha)i)(\cos(\alpha) - \text{sen}(\alpha)i) = 1 \iff \text{cos}^2(\alpha) + \text{sen}^2(\alpha) = 1.$$

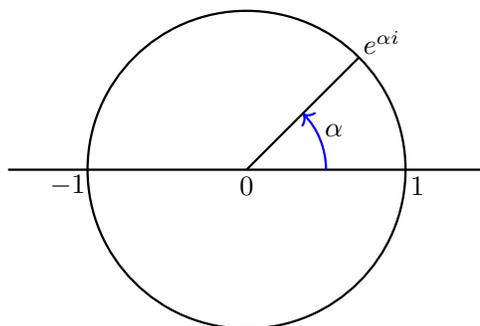


Figura 1.2: Representación de un número complejo $e^{\alpha i}$ de módulo 1 y argumento α .

Una propiedad importante es que todo número complejo distinto de cero tiene exactamente n raíces n -ésimas distintas en \mathbb{C} . Por ejemplo, $-1 = e^{i\pi}$, de modo que las tres raíces cúbicas de -1 son

$$\begin{aligned} z_1 &= (e^{\pi i})^{1/3} = e^{(\pi/3)i} = \cos(\pi/3) + \text{sen}(\pi/3)i = \frac{1}{2} + \frac{\sqrt{3}}{2}i \\ z_2 &= (e^{3\pi i})^{1/3} = e^{(\pi)i} = e^{\pi i} = \cos(\pi) + \text{sen}(\pi)i = -1 \\ z_3 &= (e^{5\pi i})^{1/3} = e^{(5\pi/3)i} = \cos(5\pi/3) + \text{sen}(5\pi/3)i = \frac{1}{2} - \frac{\sqrt{3}}{2}i. \end{aligned}$$

Hemos utilizado que el argumento de -1 es el ángulo π , pero también lo son $3\pi = \pi + 2\pi$ y $5\pi = \pi + 4\pi$. A partir de ahí, los argumentos que se forman sumando múltiplos de 2π no aportan nuevas raíces cúbicas.

1.2. El espacio vectorial \mathbb{R}^n

Los números (reales o complejos) se suelen definir en física como magnitudes escalares. Hay otras magnitudes llamadas vectoriales que “viven” en otro tipo de conjuntos: los espacios vectoriales. Los espacios vectoriales más conocidos son el plano y el espacio, pero veremos que hay muchos más. Por el momento introducimos el espacio vectorial \mathbb{R}^n por su utilidad en los siguientes temas del curso.

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; 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) . Trabajaremos fundamentalmente con dos tipos de operaciones con vectores: suma y producto por escalares.

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

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

La suma en \mathbb{R}^2 hereda las propiedades de la suma de números reales. El elemento neutro es el vector $(0, 0)$ (también llamado vector cero) y el opuesto de $v = (x, y)$ es $-v = (-x, -y)$.

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

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

Los escalares se suelen denotar por letras del alfabeto griego: $\alpha, \beta, \gamma, \lambda, \mu, \dots$

Con estas dos operaciones (suma de vectores y producto de vectores por escalares), se dice que el conjunto \mathbb{R}^2 es un **espacio vectorial**. Este concepto lo generalizaremos en el capítulo 4, donde también estableceremos las propiedades que relacionan la suma y el producto por escalares.

Tanto el espacio vectorial \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 el espacio vectorial

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

Por ejemplo, $x = (2, -1, 0, -2)$ es un vector de \mathbb{R}^4 y el cero de \mathbb{R}^5 es $(0, 0, 0, 0, 0)$.

Recordamos a continuación algunos de los conceptos fundamentales asociados al espacio vectorial \mathbb{R}^n , como son los de combinación lineal, independencia lineal, conjunto de generadores, base y dimensión.

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 vectorial 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 $\dim(U)$.

Hablando informalmente, un subespacio vectorial de \mathbb{R}^n de dimensión $k < n$ es un subconjunto de \mathbb{R}^n que se puede generar con un conjunto de k vectores linealmente independientes. Así, la dimensión de un espacio o un subespacio vectorial es una medida del número máximo de direcciones linealmente independientes que hay en él. Los subespacios vectoriales de \mathbb{R}^n de dimensión 1 son rectas y los de dimensión 2 son planos. Por ejemplo, en \mathbb{R}^3 el subespacio U generado por los vectores $(1, 0, 0)$ y $(0, 1, 0)$ es el plano horizontal, que tiene dimensión 2.

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 subconjunto $U = \{(x, y, z) \in \mathbb{R}^3 / y = 2x - z\}$. Para probar que U es un subespacio vectorial de \mathbb{R}^3 y calcular una base de U , observemos que un vector genérico de U tiene la forma $(x, y, z) = (x, 2x - z, z)$. Por tanto:

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

Por tanto, U es un subespacio vectorial de \mathbb{R}^3 , $\mathcal{B} = \{(1, 2, 0), (0, -1, 1)\}$ es una base de U y $\dim(U) = 2$ (así que U es un plano).

Producto escalar

El producto escalar se define como una operación externa, que asigna a cada par de vectores de \mathbb{R}^n un número real $x \cdot y$. En concreto, si $x = (x_1, x_2, \dots, x_n)$ e $y = (y_1, y_2, \dots, y_n)$ son dos vectores de \mathbb{R}^n , se define su producto escalar 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 su norma como la raíz cuadrada positiva del producto escalar del vector por sí mismo:

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

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 :

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

En general dos vectores no nulos en \mathbb{R}^n forman un ángulo y se cumple la misma fórmula. 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.

Un concepto importante asociado al producto escalar es el de ortogonalidad. Se dice que dos vectores x e y de \mathbb{R}^n son **ortogonales** si $x \cdot y = 0$ (es decir, forman un ángulo de $\pi/2$). 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 todos los vectores son unitarios, es decir $\|v_i\| = 1$, $\forall i = 1, 2, \dots, k$.

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 .

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 el elemento $a_{ij} \in \mathbb{R}$ ocupa el lugar correspondiente a la fila i y la columna j , con $i = 1, 2, \dots, p$ y $j = 1, 2, \dots, n$. 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.

Podemos dividir cada matriz cuadrada $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ en tres partes:

- la parte diagonal, formada por los elementos a_{ij} con $i = j$, es decir, $a_{11}, a_{22}, \dots, a_{nn}$;
- la parte triangular superior, formada por los elementos que están encima de la diagonal. Se caracterizan por ser de la forma a_{ij} con $i < j$;
- la parte triangular inferior, formada por los elementos que están debajo de la diagonal. Se caracterizan por ser de la forma a_{ij} con $i > j$;

La suma de los elementos diagonales de una matriz cuadrada $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ 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, que son aquellas en las que una de las partes triangulares solo tiene ceros.

- Una matriz cuadrada $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es **triangular superior** si todos los elementos de la parte triangular inferior son ceros. Por ejemplo,

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

- Análogamente, una matriz cuadrada $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es **triangular inferior** si todos los elementos de la parte triangular superior son ceros.

Si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, se define su **traspuesta** (y se denota A^t) como la matriz cuyas columnas son las filas de A .

En general, cuando hacemos productos de matrices que incluyan vectores, estos 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,

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

Es fácil comprobar que $(\mathcal{M}_{p \times n}(\mathbb{R}), +)$ hereda las propiedades de la suma de números reales. 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,

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

Producto de matrices

Dados dos vectores $u = (u_1, u_2, \dots, u_n)$ y $v = (v_1, v_2, \dots, v_n)$ de \mathbb{R}^n , su producto escalar se expresa en forma matricial como el producto de un vector fila por un vector columna. Usando notación de columnas como se indicó anteriormente:

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

En adelante usaremos esta expresión matricial para el producto escalar.

En el caso general, se define el producto de dos matrices A y B en función de los productos escalares de las filas de A por las columnas de B : dadas dos matrices $A = (a_{ij}) \in \mathcal{M}_{p \times n}(\mathbb{R})$ y

$B = (b_{ij}) \in \mathcal{M}_{n \times q}(\mathbb{R})$, podemos denotar las filas de A por $u_1^t, u_2^t, \dots, u_p^t$ y las columnas de B como v_1, v_2, \dots, v_q . De esta forma, podemos definir el producto AB del siguiente modo:

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

Si denotamos $C = AB = (c_{ij}) \in \mathcal{M}_{p \times q}(\mathbb{R})$, tenemos la siguiente expresión para cada una de las entradas c_{ij} de C :

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

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 . Por ejemplo, también es posible multiplicar un vector columna de \mathbb{R}^n por un vector fila, pero el resultado es muy diferente:

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

Observaciones:

El producto de una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ por un vector columna $b \in \mathbb{R}^n$ se puede interpretar como una combinación lineal de las columnas de A ; en concreto, si

$$A = (u_1 | u_2 | \dots | u_n) \in \mathcal{M}_{p \times n}(\mathbb{R}) \quad \text{y} \quad b = \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} \in \mathbb{R}^n,$$

entonces:

$$Ab = (u_1 | u_2 | \dots | u_n) \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{pmatrix} = b_1 u_1 + b_2 u_2 + \dots + b_n u_n \in \mathbb{R}^p.$$

Esta propiedad es bastante útil y tiene una generalización que también usaremos más adelante; si expresamos la matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ en notación de columnas y $B \in \mathcal{M}_{n \times q}(\mathbb{R})$ en notación de filas, entonces:

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

Propiedades del producto de matrices

- El producto de matrices cumple la propiedad asociativa, es decir si A , B y C se pueden multiplicar entonces $ABC = (AB)C = A(BC)$.
- El producto de matrices cumple 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 cuadradas no es conmutativo, es decir, si $A, B \in \mathcal{M}_{n \times n}(\mathbb{R})$, en general $AB \neq BA$.

Ejemplo:

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

- Recordemos que si el producto de dos números reales es cero entonces al menos uno de ellos debe ser cero. Esta propiedad no es cierta para matrices: si $A, B \in \mathcal{M}_{n \times n}(\mathbb{R})$, en general $AB = 0 \not\Rightarrow A = 0$ o $B = 0$. Esto tiene importantes consecuencias a la hora de resolver ecuaciones con matrices.

Ejemplo: Consideremos la ecuación $X^2 = X$, con $X \in \mathcal{M}_{2 \times 2}(\mathbb{R})$. Se tiene:

$$X^2 = X \iff X^2 - X = 0 \iff X(X - I) = 0.$$

En particular, $X = 0$ y $X = I$ son soluciones de la ecuación, pero no podemos deducir que sean las únicas. Por ejemplo, para

$$X = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/2 \end{pmatrix}$$

se cumple que $X^2 = X$ y por tanto X es solución de la ecuación (de hecho, todas las matrices cuadradas 2×2 de rango 1 y traza 1 son soluciones de la ecuación).¹

¹Esta propiedad se puede probar usando el teorema de Cayley-Hamilton, que se enuncia en el capítulo 5.

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 **inversible** 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:

Inversa del producto: Si A y $B \in \mathcal{M}_{n \times n}(\mathbb{R})$ son inversibles entonces AB también lo es y además $(AB)^{-1} = B^{-1}A^{-1}$.

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $k \geq 2$ es un número natural, 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: si A es diagonal entonces A^k también es diagonal y además

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

En el capítulo 5 se proporcionarán métodos para calcular la expresión de la potencia k -ésima de una matriz.

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 inversible 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.

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.

Una de las propiedades características de las matrices ortogonales es que no modifican la norma de los vectores, es decir, si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es ortogonal y $x \in \mathbb{R}^n$, entonces $\|Ax\| = \|x\|$. En efecto, como $A^t A = I$:

$$\|Ax\| = \sqrt{(Ax)^t(Ax)} = \sqrt{x^t A^t A x} = \sqrt{x^t x} = \|x\|.$$

En la sección 8 del capítulo 4 comentaremos más propiedades de estas matrices.

2.5. Matrices elementales

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

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

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. Nos centraremos en matrices elementales de filas; las matrices elementales de columnas se definen de manera análoga.

Tipos de matrices elementales de filas

Distinguiremos tres tipos de matrices elementales de filas según los tipos de operaciones elementales definidos anteriormente:

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 .

Ejemplos:

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

$$F_{23} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}, F_2(3) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 1 \end{pmatrix}, F_{13}(2) = \begin{pmatrix} 1 & 0 & 2 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

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:

Consideremos la matriz $A = \begin{pmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \end{pmatrix}$.

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,

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

2.6. Forma escalonada y rango de una matriz

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

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

Si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, se define el **rango** de A como el número de filas no nulas de la forma escalonada reducida de A . Denotaremos el rango de A por $\text{rg}(A)$.

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.

Como 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, tenemos la siguiente propiedad:

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

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

Rango y matrices inversibles

La siguiente propiedad será de utilidad en la caracterización de matrices inversibles y el cálculo de la inversa.

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz cuadrada y $\text{rref}(A)$ no tiene filas de ceros entonces $\text{rref}(A) = I$. Es decir, la forma escalonada reducida de las matrices $n \times n$ de rango n es la identidad.

La siguiente propiedad caracteriza las matrices inversibles en función de su rango, lo que permite saber de forma sencilla si una matriz cuadrada tiene inversa o no.

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz cuadrada entonces: A es inversible $\iff \text{rg}(A) = n$.

La prueba de esta propiedad es interesante porque proporciona la idea para calcular la inversa de una matriz haciendo operaciones elementales. 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. Así:

- Si A es inversible entonces $\text{rref}(A) = FA$ también es inversible. En particular, $\text{rref}(A)$ no tiene filas de ceros y el rango de A es n .
- Si el rango de A es n , entonces $\text{rref}(A) = I$. Esto quiere decir que existe una matriz F tal que $FA = \text{rref}(A) = I$. Por tanto, A es inversible y $F = A^{-1}$.

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:

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 propiedad permite calcular la inversa de A utilizando operaciones elementales del siguiente modo: si F_1, F_2, \dots, F_k son 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 operaciones elementales sobre las filas de la matriz $(A|I)$:

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

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.
- 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:

$$\begin{vmatrix} 1 & 1 & 2 \\ 1 & 1 & 0 \\ 2 & 1 & 2 \end{vmatrix} \begin{array}{l} F_{21}(-1) \\ = \\ F_{31}(-2) \end{array} \begin{vmatrix} 1 & 1 & 2 \\ 0 & 0 & -2 \\ 0 & -1 & -2 \end{vmatrix} \begin{array}{l} F_{23} \\ = \\ - \end{array} \begin{vmatrix} 1 & 0 & 2 \\ 0 & -1 & -2 \\ 0 & 0 & -2 \end{vmatrix} = -[1(-1)(-2)] = -2.$$

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

Consideremos una matriz $A = (a_{ij}) \in \mathcal{M}_{n \times n}(\mathbb{R})$. Si \widetilde{A}_{ij} es 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 cumple una fórmula análoga para cada columna de A . En particular, se tienen las siguientes consecuencias:

1. Si $n = 2$, entonces

$$\begin{vmatrix} a & b \\ c & d \end{vmatrix} = ad - bc.$$

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 $\det(A) = (-1)^{i+k} a_{ik} \det(\widetilde{A}_{ik})$.

Ejemplo: En este ejemplo desarrollamos el determinante de A por los elementos de la fila 2.

$$\begin{vmatrix} 1 & 1 & 2 \\ 0 & 3 & 0 \\ 2 & 1 & 2 \end{vmatrix} = 3 \begin{vmatrix} 1 & 2 \\ 2 & 2 \end{vmatrix} = 3(-2) = -6.$$

Otras propiedades de los determinantes

1. (Fórmula del producto): $|AB| = |A| |B|$, $\forall A, B \in \mathcal{M}_{n \times n}(\mathbb{R})$.
2. $|A^t| = |A|$, $\forall A \in \mathcal{M}_{n \times n}(\mathbb{R})$.
3. Si $\lambda \in \mathbb{R}$ entonces

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

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

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

$$|A^{-1}| = \frac{1}{|A|}.$$

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: La expresión matricial de 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 la siguiente:

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

Formas cuadráticas degeneradas y no degeneradas

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz simétrica y $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ es la forma cuadrática definida por $\omega(x) = x^t A x$, 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.

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 $\omega(x) = x^t A x$, 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 $\omega(x) = x^t Ax > 0, \forall x \neq 0$,
- (b) ω es **definida negativa** si $\omega(x) = x^t Ax < 0, \forall x \neq 0$,
- (c) ω es **indefinida** si no tiene signo constante, es decir, 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 $\omega_A(x) = x^t Ax$.

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,

$$\omega(1, 0, 0) = 1 > 0 \quad ; \quad \omega(0, 0, 1) = -1 < 0.$$

Además es no degenerada ya que

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

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

Si la matriz asociada es diagonal, es sencillo deducir que la forma cuadrática es definida positiva si todos los elementos de la diagonal son positivos, definida negativa si todos los elementos de la diagonal son negativos, e indefinida si hay elementos de signos distintos.

Sin embargo, en general es difícil determinar directamente la clasificación de ω si la matriz asociada no es diagonal. 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.

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

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

Caracterización de matrices definidas positivas: Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz simétrica entonces A es definida positiva si y solo 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 $\omega(x) = x^t Ax$, donde

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

Los menores principales de A son:

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

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 solo 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:

Caracterización de matrices definidas negativas: Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz simétrica entonces A es definida negativa si y solo 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:

Claasificación de formas cuadráticas no degeneradas: Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz simétrica con $|A| \neq 0$ entonces la forma cuadrática $\omega(x) = x^t Ax$ 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.

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 de la forma

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

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

Dado un sistema de p ecuaciones lineales con n incógnitas $Ax = b$, las soluciones son vectores de \mathbb{R}^n . 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. 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.

Recordemos que la expresión matricial del producto $Ax = b$ se puede escribir en términos de las columnas de A como

$$Ax = (u_1 | u_2 | \cdots | u_n) \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix} = x_1 u_1 + x_2 u_2 + \cdots + x_n u_n = b.$$

De aquí se obtiene que el sistema $Ax = b$ es compatible si y solo si b es combinación lineal de las columnas de A . Así, la compatibilidad del sistema se puede estudiar usando el rango de matrices:

Teorema de Rouché-Frobenius: El sistema de ecuaciones lineales $Ax = b$ es compatible si y solo si $\text{rg}(A) = \text{rg}(A|b)$.

Eliminación gaussiana

Las operaciones elementales por filas proporcionan el método más sencillo para resolver sistemas de ecuaciones lineales y determinar si son compatibles determinados o indeterminados.

Resolución de sistemas por el método de Gauss: Consideremos un sistema $Ax = b$ 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.

La prueba es un simple ejercicio usando matrices elementales: si F_1, F_2, \dots, F_k son las matrices elementales correspondientes a las operaciones por filas sobre $(A|b)$ y $F = F_k \dots F_2 F_1$, entonces

$(A'|b') = (FA|Fb)$ y el nuevo sistema es $FAx = Fb$, que es equivalente a $Ax = b$ ya que F es inversible.

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

- Si $\text{rg}(A) = \text{rg}(A|b) = n$ ($n = \text{número de incógnitas} = \text{número de columnas de } A$) 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 que no corresponden a entradas principales de $\text{rref}(A)$ (incógnitas libres).

3.4. Sistemas compatibles determinados

Cuando A es una matriz cuadrada, el criterio para determinar si el sistema $Ax = b$ es compatible determinado depende solo de la matriz A :

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $b \in \mathbb{R}^n$, entonces el sistema $Ax = b$ tiene solución única si y solo si $\text{rg}(A) = n$.

Este resultado es consecuencia de que la matriz $(A|b)$ tiene n filas. Por tanto, si $\text{rg}(A) = n$, necesariamente $\text{rg}(A|b) = n$.

Combinando varios resultados que hemos ido obteniendo, podemos enunciar la siguiente caracterización de las matrices inversibles:

Las siguientes propiedades son equivalentes para una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$:

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

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 es computacionalmente adecuado calcular la inversa de A para resolver el sistema.

3.5. Sistemas compatibles indeterminados

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,

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

Discusión de sistemas homogéneos:

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

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

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

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

Los vectores u_1, u_2, \dots, u_k de la base se determinan despejando las incógnitas correspondientes a las entradas principales de la forma escalonada reducida de A en función del resto.

Ejemplo: Consideremos el sistema

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

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

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

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

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

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 un plano de \mathbb{R}^4 . Más concretamente, 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 $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $b \in \mathbb{R}^p$.

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:

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

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

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

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:

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

En primer lugar, $\text{rg}(A|b) = \text{rg}(A'|b') = 2 < 3 =$ 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

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

3.6. Soluciones de un sistema en el sentido de mínimos cuadrados

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 . El enunciado del teorema de Rouché-Frobenius es equivalente a decir que el sistema $Ax = b$ es compatible si y solo si $b \in \text{Im}(A)$.

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 $x_0 \in \mathbb{R}^n$ tales que $Ax_0 = b'$ serán lo que llamaremos soluciones del sistema $Ax = b$ en el sentido de mínimos cuadrados.

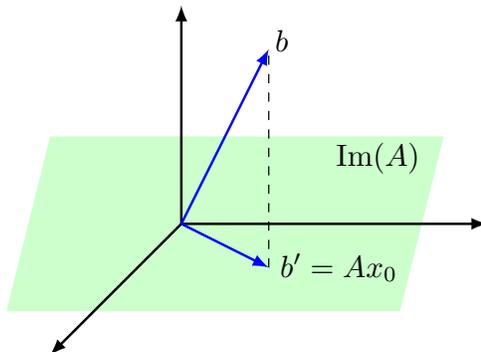


Figura 3.1: Interpretación geométrica del problema de mínimos cuadrados. El plano sombreado es la imagen de A . El vector x_0 es solución de $Ax = b$ en el sentido de mínimos cuadrados si $b' = Ax_0$ es la proyección ortogonal de b sobre $\text{Im}(A)$.

Dicho de otro modo, si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $b \in \mathbb{R}^p$, entonces $x_0 \in \mathbb{R}^n$ 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 solo 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:

Soluciones de mínimos cuadrados: Consideremos el sistema $Ax = b$, donde $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ 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 .

Si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ y $b \in \mathbb{R}^p$, entonces el sistema $A^tAx = A^tb$ es 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 solo si $\text{rg}(A) = n$.

Ajuste polinómico de datos mediante mínimos cuadrados

Una de las principales aplicaciones del método de mínimos cuadrados es el ajuste de datos. Supongamos que se calcula experimentalmente el valor de una cierta magnitud y que se supone que es función polinómica de otra magnitud 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 $A^tAx = A^tb$ proporciona los coeficientes del polinomio de grado n que mejor ajusta los datos en el sentido de mínimos cuadrados.

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 $y = \frac{5}{2} + \frac{2}{5}x$ (línea azul en la figura 3.2).

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 $A^tAx = A^tb$ 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 3.2 se representa la parábola en trazo de color negro.

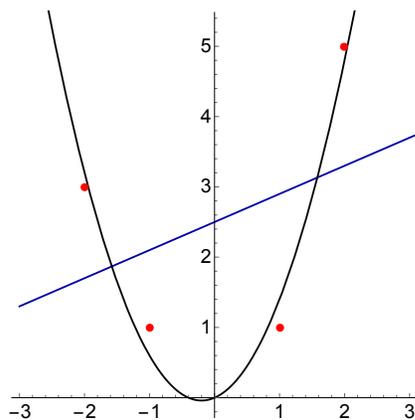


Figura 3.2: Aproximaciones lineal (en azul) y cuadrática (en negro) de los datos (puntos rojos).

Capítulo 4

Espacios vectoriales y aplicaciones lineales

4.1. Introducción

En este capítulo se introducen la definición de espacio vectorial y los principales conceptos relacionados (subespacio vectorial, 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*), con las propiedades habituales: asociativa, conmutativa, existencia de elemento neutro (cero) y existencia de elemento opuesto ($-v$) para cada $v \in V$.
- 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 2.
3. El conjunto $\mathcal{C}(\mathbb{R}) = \{f : \mathbb{R} \rightarrow \mathbb{R} / f \text{ es continua}\}$ 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

Un subconjunto U de un espacio vectorial 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 e $\text{Im}(A)$ es un subespacio vectorial de \mathbb{R}^p .
2. El conjunto $U = \{A \in \mathcal{M}_{n \times n}(\mathbb{R}) / A^t = A\}$ de las matrices simétricas de tamaño $n \times n$ 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 $|A_1| = |A_2| = 0$. Sin embargo,

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

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

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

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

Si U es un subespacio vectorial de un espacio vectorial V , diremos 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 .

Ejemplo: El subespacio vectorial de las matrices simétricas 2×2 es

$$U = \left\{ \begin{pmatrix} a & b \\ b & c \end{pmatrix} / a, b, c \in \mathbb{R} \right\} = \left\{ a \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix} + b \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} + c \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} / a, b, c \in \mathbb{R} \right\}.$$

Por tanto,

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

es un conjunto de generadores de U .

4.3. Independencia lineal

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

Si V es un espacio vectorial y S es un subconjunto de V , diremos 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 solo si su rango es n .

Si $V = \mathbb{R}^n$ entonces estudiar si un conjunto de vectores S es linealmente independiente se reduce a calcular el rango de la matriz que tiene como filas los vectores de S .

Ejemplo: Si $S = \{(1, 2, 1, 1), (-1, 1, 0, 0), (1, 5, 2, 2)\}$, entonces:

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

Por tanto, S no es linealmente independiente.

Observación: Si solo 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' .

En el ejemplo anterior,

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

4.4. Bases y dimensión

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

Ejemplo:

$$\text{El conjunto } \mathcal{B} = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\} \text{ es una base de } \mathcal{M}_{2 \times 2}(\mathbb{R}).$$

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 $\dim(V)$.

Ejemplo: $\dim(\mathcal{M}_{2 \times 2}(\mathbb{R})) = 4$ y, 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: Consideremos el subespacio $U = \langle \{(1, 2, 1, 1), (0, 1, -1, -1), (0, 0, 0, 1)\} \rangle$. 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 $U = \{A = (a_{ij}) \in \mathcal{M}_{3 \times 4}(\mathbb{R}) / a_{11} = a_{22} = a_{33} = 0\}$, entonces

$$\dim(U) = \dim(\mathcal{M}_{3 \times 4}(\mathbb{R})) - 3 = 12 - 3 = 9.$$

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

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

Coordenadas de un vector: Si $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ es una base de \mathbb{R}^n entonces cada vector $x \in \mathbb{R}^n$ 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:

Calcular las coordenadas del vector $x = (1, 0, 0) \in \mathbb{R}^3$ respecto de la base

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

Escribiendo $(1, 0, 0) = (\alpha, \beta, \gamma)_{\mathcal{B}}$, se tiene:

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

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

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

$$x_{\mathcal{B}} = \begin{pmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \end{pmatrix}.$$

Si consideramos la base canónica \mathcal{C} , entonces las coordenadas de un vector $x = (x_1, x_2, \dots, x_n)$ respecto de \mathcal{C} son precisamente (x_1, x_2, \dots, x_n) , es decir,

$$x_{\mathcal{C}} = x = \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{pmatrix}.$$

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

Si $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ es una base de \mathbb{R}^n , la matriz $P_{\mathcal{B}\mathcal{C}} \in \mathcal{M}_{n \times n}(\mathbb{R})$ es la **matriz de cambio de base** de \mathcal{B} a la base canónica \mathcal{C} si $P_{\mathcal{B}\mathcal{C}} x_{\mathcal{B}} = x_{\mathcal{C}}, \forall x \in \mathbb{R}^n$.

Observemos que si $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ y $(\lambda_1, \lambda_2, \dots, \lambda_n)$ es su vector de coordenadas respecto de \mathcal{B} , entonces:

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

Por tanto, tenemos la siguiente propiedad:

Si $\mathcal{B} = \{u_1, u_2, \dots, u_n\}$ es una base de \mathbb{R}^n entonces la matriz de cambio de base de \mathcal{B} a \mathcal{C} es la matriz cuyas columnas son los vectores de \mathcal{B} , es decir,

$$P_{\mathcal{BC}} = (u_1 | u_2 | \dots | u_n) \in \mathcal{M}_{n \times n}(\mathbb{R}).$$

Ejemplo: Si $\mathcal{B} = \{(1, 1, 1), (1, 2, 0), (0, 0, 1)\}$, entonces la matriz de cambio de base de \mathcal{B} a \mathcal{C} es

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

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:

Si \mathcal{B} y \mathcal{B}' son dos bases de \mathbb{R}^n entonces:

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

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

$$P_{\mathcal{CB}} = (P_{\mathcal{BC}})^{-1} = \left(\begin{array}{ccc} 1 & 1 & 0 \\ 1 & 2 & 0 \\ 1 & 0 & 1 \end{array} \right)^{-1} = \left(\begin{array}{ccc} 2 & -1 & 0 \\ -1 & 1 & 0 \\ -2 & 1 & 1 \end{array} \right).$$

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 $i = 1, 2, \dots, p$.

Una de las propiedades más interesantes de las bases ortonormales es su relación con las matrices ortogonales, como muestra el siguiente resultado:

Una matriz $P \in \mathcal{M}_{n \times n}(\mathbb{R})$ es ortogonal si y solo si sus columnas son una base ortonormal de \mathbb{R}^n . En particular, la matriz de cambio de coordenadas $P_{\mathcal{BC}}$ es una matriz ortogonal si y solo si \mathcal{B} es una base ortonormal.

En efecto, 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,

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

Construcción de bases ortonormales

El procedimiento de ortonormalización de Gram-Schmidt permite calcular una base ortonormal de un subespacio vectorial U de \mathbb{R}^n a partir de una base de U . Si $\mathcal{B} = \{v_1, v_2, \dots, v_p\}$ es una base de U se construye 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:

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

- (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 .

4.7. Aplicaciones lineales

Si V y W son dos espacios vectoriales, una aplicación $L : V \rightarrow W$ es una aplicación 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.

Ejemplo: La aplicación $D : \mathcal{C}^1(\mathbb{R}) \rightarrow \mathcal{C}(\mathbb{R})$, definida por $D(f) = f'$ para cada función continuamente diferenciable $f : \mathbb{R} \rightarrow \mathbb{R}$, es una aplicación lineal porque

$$D(f + g) = (f + g)' = f' + g' = D(f) + D(g) \quad \text{y} \quad D(\lambda f) = (\lambda f)' = \lambda f' = \lambda D(f).$$

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

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

$$L(x) = Ax, \forall x \in \mathbb{R}^n.$$

La prueba es de nuevo una aplicación del producto de matrices. Denotemos la base canónica de \mathbb{R}^n por $C = \{e_1, e_2, \dots, e_n\}$ y tomemos un vector $x = (x_1, x_2, \dots, x_n) = x_1 e_1 + x_2 e_2 + \cdots + x_n e_n$. 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 es habitual que la matriz asociada a una aplicación lineal se pueda obtener directamente.

Ejemplo: Consideremos la aplicación lineal $L : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ definida por $L(x, y, z) = (x + 2y - z, y + 4z)$. Escribiendo los vectores en forma de columna:

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

En consecuencia, 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. Transformaciones ortogonales.

Se dice que una aplicación lineal $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es una transformación ortogonal si el producto escalar de las imágenes de dos vectores x e y coincide con el producto escalar de x por y , es decir, si

$$(L(x))^t L(y) = x^t y, \forall x, y \in \mathbb{R}^n.$$

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:

Si $L : \mathbb{R}^n \rightarrow \mathbb{R}^n$ es una aplicación lineal y $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es su matriz asociada, entonces L es una transformación ortogonal si y solo si A es una matriz ortogonal.

Es fácil probar que las transformaciones ortogonales conservan la norma, la distancia y el ángulo:

- $\|L(x)\| = \sqrt{(L(x))^t L(x)} = \sqrt{x^t x} = \|x\|.$
- $d(L(x), L(y)) = \|L(x) - L(y)\| = \|L(x - y)\| = \|x - y\| = d(x, y).$
- $\cos(L(x), L(y)) = \frac{(L(x))^t L(y)}{\|L(x)\| \|L(y)\|} = \frac{x^t y}{\|x\| \|y\|} = \cos(x, y).$

Por esta razón, las transformaciones ortogonales no deforman las figuras geométricas y se suelen llamar **movimientos rígidos**. En \mathbb{R}^2 las únicas transformaciones ortogonales son giros o simetrías respecto a un eje.

Como las columnas de una matriz ortogonal $n \times n$ son una base ortonormal de \mathbb{R}^n , un cambio de coordenadas dado por una matriz $P_{\mathcal{B}\mathcal{C}}$ es un movimiento rígido si y solo si \mathcal{B} es una base ortonormal. En el caso de movimientos rígidos, el cambio de base no deforma las figuras geométricas.

Como ilustración, en la figura 4.1 se muestra cómo afectan dos cambios de base a la elipse de ecuación $5x^2 + 5y^2 + 6xy = 8$. En el gráfico de la izquierda se representa la elipse (en rojo) en coordenadas respecto de la base $\mathcal{B}' = \{(4, 0), (0, 1)\}$, que es ortogonal pero no ortonormal. Claramente

la elipse se deforma, en particular reduciendo su perímetro. En el gráfico de la derecha se representa la elipse (en negro) en coordenadas respecto de la base $\mathcal{B} = \{(1/\sqrt{2}, -1/\sqrt{2}), (1/\sqrt{2}, 1/\sqrt{2})\}$, que es ortonormal. La elipse no se deforma, de hecho el cambio de coordenadas corresponde a un giro de ángulo $\pi/4$ en el sentido positivo de giro.

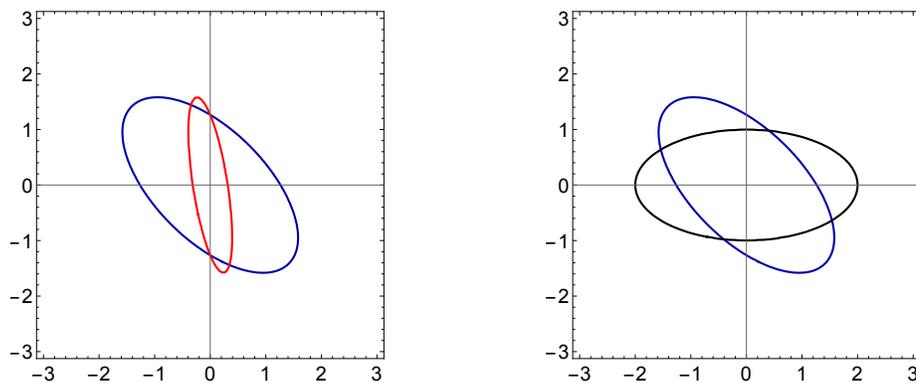


Figura 4.1: Efecto de dos cambios de coordenadas (ortonormal a la derecha) en la representación de la elipse $5x^2 + 5y^2 + 6xy = 8$ (en azul).

4.9. Proyección ortogonal

Recordemos que si $b \in \mathbb{R}^n$ y U es un subespacio de \mathbb{R}^n con $\dim(U) = p < n$, la proyección ortogonal de b sobre el subespacio U es el único vector $b' \in U$ tal que $(b - b')$ es ortogonal a U (ver la figura 3.1). 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.

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

$$P = AA^t = (u_1|u_2|\dots|u_p) \begin{pmatrix} u_1^t \\ u_2^t \\ \vdots \\ u_p^t \end{pmatrix} = u_1u_1^t + u_2u_2^t + \dots + u_pu_p^t.$$

Caso particular: $\dim(U) = 1$.

Si u es un vector unitario y $U = \langle \{u\} \rangle$ es la recta en la dirección de u , la matriz de proyección ortogonal sobre U es $P = uu^t$. En este caso, el rango de P es 1.

Ejemplos:

1. Construir la matriz de proyección ortogonal sobre la recta $W = \langle \{(2, 2, 1)\} \rangle$ de \mathbb{R}^3 .

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,

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

2. Hallar la matriz de proyección ortogonal sobre el plano

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

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:

$$\begin{aligned} P &= u_1 u_1^t + u_2 u_2^t = (u_1 | u_2) \begin{pmatrix} u_1^t \\ u_2^t \end{pmatrix} = \begin{pmatrix} 1/\sqrt{2} & -1/\sqrt{6} \\ 0 & 2/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{6} \end{pmatrix} \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ -1/\sqrt{6} & 2/\sqrt{6} & 1/\sqrt{6} \end{pmatrix} = \\ &= \begin{pmatrix} 1/2 + 1/6 & 0 - 2/6 & 1/2 - 1/6 \\ 0 - 2/6 & 0 + 4/6 & 0 + 2/6 \\ 1/2 - 1/6 & 0 + 2/6 & 1/2 + 1/6 \end{pmatrix} = \begin{pmatrix} 2/3 & -1/3 & 1/3 \\ -1/3 & 2/3 & 1/3 \\ 1/3 & 1/3 & 2/3 \end{pmatrix}. \end{aligned}$$

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.

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

Geoméricamente, si $\lambda \in \mathbb{R}$ es un autovalor no nulo de A , entonces un autovector $x \in \mathbb{R}^n$ asociado a λ se caracteriza por el hecho de que x y Ax apuntan en la misma dirección.

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 lo denotaremos por $\text{Sp}(A)$.

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. Es claro geoméricamente porque se trata de un giro de ángulo $\pi/2$, de modo que es imposible que x y Ax apunten en la misma dirección.

En este caso, $\lambda = i \in \text{Sp}(A)$ y el vector $(i, 1) \in \mathbb{C}^2$ es un autovector asociado:

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

Cálculo de autovalores: polinomio característico

Observemos que

$$Ax = \lambda x \iff Ax - \lambda x = 0 \iff (A - \lambda I)x = 0.$$

Por tanto, existe un vector $x \neq 0$ tal que $Ax = \lambda x$ si y solo si el sistema homogéneo $(A - \lambda I)x = 0$ tiene soluciones distintas de cero. Esto es equivalente a que $|A - \lambda I| = 0$. Este argumento proporciona la forma de calcular los autovalores de una matriz:

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y λ es un escalar, entonces $\lambda \in \text{Sp}(A) \iff |A - \lambda I| = 0$. En consecuencia,

$$\text{Sp}(A) = \{\lambda \in \mathbb{C} \mid |A - \lambda I| = 0\}.$$

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, se llama **polinomio característico** de A al polinomio de grado n 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: El polinomio característico de

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

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 $\lambda_1 = 3$ y $\lambda_2 = -1$.

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

Si $p(x)$ es un polinomio 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,

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

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

Por ejemplo, el polinomio $p(x) = x^3 - 3x + 2$ se puede factorizar como

$$p(x) = x^3 - 3x + 2 = (x - 1)^2(x + 2).$$

Por tanto sus raíces son $\lambda_1 = 1$ (con multiplicidad 2) y $\lambda_2 = -2$ (con multiplicidad 1).

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

De las observaciones anteriores se deduce que 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.

Ejemplo: Se considera la matriz

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

Calculamos el polinomio característico de A :

$$\begin{aligned} |A - xI| &= \begin{vmatrix} -x & 1 & 1 \\ 1 & -x & 1 \\ 1 & 1 & -x \end{vmatrix} \begin{array}{l} F_{12}(1) \\ = \\ F_{13}(1) \end{array} \begin{vmatrix} 2-x & 2-x & 2-x \\ 1 & -x & 1 \\ 1 & 1 & -x \end{vmatrix} = (2-x) \begin{vmatrix} 1 & 1 & 1 \\ 1 & -x & 1 \\ 1 & 1 & -x \end{vmatrix} = \\ & \begin{array}{l} F_{21}(-1) \\ = \\ F_{31}(-1) \end{array} (2-x) \begin{vmatrix} 1 & 1 & 1 \\ 0 & -1-x & 0 \\ 0 & 0 & -1-x \end{vmatrix} = (2-x)(-1-x)^2. \end{aligned}$$

Por tanto, $\text{Sp}(A) = \{2, -1\}$, con m.a.(2) = 1, m.a.(-1) = 2. Equivalentemente, podemos escribirlo como $\text{Sp}(A) = \{2, -1, -1\}$, repitiendo los autovalores tantas veces como indica su multiplicidad algebraica.

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:

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

La propiedad 2) es útil para comprobar si los autovalores se han calculado correctamente, ya que su suma debe coincidir con la traza de la matriz. En el ejemplo anterior:

$$\operatorname{tr}(A) = 0 + 0 + 0 = 0 = 2 + (-1) + (-1) = \lambda_1 + \lambda_2 + \lambda_3.$$

Cálculo de autovectores. Subespacios propios

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

$$V(\lambda) = \{x \in \mathbb{R}^n / Ax = \lambda x\} = \{x \in \mathbb{R}^n / (A - \lambda I)x = 0\} = \operatorname{Ker}(A - \lambda I).$$

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

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

La multiplicidad geométrica de λ proporciona la cantidad de autovectores independientes asociados a λ .

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

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

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

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $\lambda \in \operatorname{Sp}(A)$, entonces $1 \leq \operatorname{m.g.}(\lambda) \leq \operatorname{m.a.}(\lambda) \leq n$. En particular, si $\lambda \in \operatorname{Sp}(A)$ y $\operatorname{m.a.}(\lambda) = 1$ entonces $\operatorname{m.g.}(\lambda) = \operatorname{m.a.}(\lambda) = 1$.

Volviendo al ejemplo anterior, como $\operatorname{m.a.}(2) = 1$, se tiene que $\operatorname{m.g.}(2) = \operatorname{m.a.}(2) = 1$.

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

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

Los subespacios propios asociados a 2 y -1 son:

$$V(2) = \operatorname{Ker}(A - 2I) = \{(x, y, z) \in \mathbb{R}^3 / x = y = z\} = \langle \{(1, 1, 1)\} \rangle.$$

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

Las multiplicidades algebraica y geométrica no siempre coinciden. Por ejemplo, el polinomio característico de

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

es

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

Por tanto, $\operatorname{Sp}(A) = \{-1\}$, con $\operatorname{m.a.}(-1) = 2$. Sin embargo,

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

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

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

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

Caracterización de matrices diagonalizables: Para una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ se cumplen las siguientes propiedades:

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

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

- (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.

En términos de aplicaciones lineales, diagonalizar una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es equivalente a encontrar una base de \mathbb{R}^n en la que la expresión matricial de la aplicación lineal $L(x) = Ax$ sea diagonal. La base correspondiente está formada por las columnas de la matriz P .

La diagonalización se puede aplicar al cálculo de potencias de matrices: si $A = PDP^{-1}$ entonces $A^k = PD^k P^{-1}$, $\forall k \geq 1$.

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

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

En este caso $\text{Sp}(A) = \{6, 3\}$, con $\text{m.a.}(6) = \text{m.g.}(6) = 1$, $\text{m.a.}(3) = \text{m.g.}(3) = 2$. Por tanto, A es diagonalizable. Además,

$$\text{Ker}(A - 6I) = \langle \{(1, 1, 1)\} \rangle, \quad \text{Ker}(A - 3I) = \langle \{(1, 0, -1), (0, 1, -1)\} \rangle.$$

Por tanto, podemos tomar

$$D = \begin{pmatrix} 6 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & 3 \end{pmatrix}, \quad P = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix},$$

de tal forma que $A = PDP^{-1}$. Finalmente,

$$\begin{aligned} A^n = P D^n P^{-1} &= \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & -1 \end{pmatrix} \begin{pmatrix} 6^n & 0 & 0 \\ 0 & 3^n & 0 \\ 0 & 0 & 3^n \end{pmatrix} \begin{pmatrix} 1/3 & 1/3 & 1/3 \\ 2/3 & -1/3 & -1/3 \\ -1/3 & 2/3 & -1/3 \end{pmatrix} = \\ &= \frac{1}{3} \begin{pmatrix} 6^n + 3^n & 2 \cdot 6^n - 3^n & 6^n - 3^n \\ 6^n - 3^n & 6^n + 3^n & 2 \cdot 6^n - 3^n \\ 6^n - 3^n & 6^n - 3^n & 6^n + 3^n \end{pmatrix}. \end{aligned}$$

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 $P^t P = I$.

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$, 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 $A = P D P^t = P D P^{-1}$. En tal caso, se dice que la factorización $A = P D P^t$ es una **diagonalización ortogonal de A** .

La importancia de la diagonalización ortogonal desde un punto de vista geométrico es que la base en la cual la matriz asociada a la aplicación lineal $L(x) = Ax$ es diagonal es una base ortonormal y por tanto el cambio de base es un movimiento rígido. El siguiente resultado establece que la diagonalización ortogonal es posible únicamente para matrices simétricas.

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

Cálculo de la diagonalización ortogonal de una matriz simétrica

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

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:

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

La prueba de esta propiedad es un ejercicio interesante: consideremos dos autovalores $\lambda_1 \neq \lambda_2$ de A , $x_1 \in V(\lambda_1)$ y $x_2 \in V(\lambda_2)$. 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 $x_1^t x_2 = 0$.

Teniendo en cuenta las propiedades anteriores, los pasos para calcular una diagonalización ortogonal $A = PDP^t$ de una matriz simétrica A 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: Calculamos una diagonalización ortogonal de la matriz

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

Dado que A es una matriz simétrica real, es ortogonalmente diagonalizable, es decir, existen dos matrices $P, D \in \mathcal{M}_{3 \times 3}(\mathbb{R})$ 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 $q_A(x) = (-3 - x)^2(3 - x)$ (hágase como ejercicio).

Por tanto, $\text{Sp}(A) = \{-3, -3, 3\}$ y la matriz D es

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

Los vectores columna de la matriz ortogonal $P = (u_1|u_2|u_3)$ constituyen una base ortonormal de \mathbb{R}^3 formada por autovectores de A . Para determinarlos, aplicaremos el procedimiento de ortogonalización 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:

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

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

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

$$\tilde{u}_2 = v_2 - (v_2^t u_1)u_1 = (-1, 1, 1) \quad ; \quad u_2 = \frac{\tilde{u}_2}{\|\tilde{u}_2\|} = (-1/\sqrt{3}, 1/\sqrt{3}, 1/\sqrt{3}).$$

Por otra parte,

$$V(3) = \text{Ker}(A - 3I) = \langle \{(1, -1, 2)\} \rangle = \langle \{v_3\} \rangle ,$$

de modo que el vector columna u_3 de P viene dado por

$$u_3 = \frac{v_3}{\|v_3\|} = (1/\sqrt{6}, -1/\sqrt{6}, 2/\sqrt{6}).$$

Así, la matriz ortogonal

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

cumple que $A = PDP^t$.

5.5. Descomposición espectral

Consideremos una diagonalización ortogonal $A = PDP^t$ de una matriz simétrica A de rango r . Si $\lambda_1, \lambda_2, \dots, \lambda_r$ son sus autovalores no nulos, contados con su multiplicidad, y u_1, u_2, \dots, u_n son las columnas de P entonces, usando las expresiones del producto de matrices dadas en el capítulo 2, se tiene:

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

$$= \lambda_1 u_1 u_1^t + \lambda_2 u_2 u_2^t + \dots + \lambda_r u_r u_r^t.$$

De esta manera se expresa A como la suma de r matrices $A_i = \lambda_i u_i u_i^t$ de rango uno. Este desarrollo de A se llama **descomposición espectral de A** . Obsérvese que cada sumando es el producto de un autovalor de A por la matriz de proyección sobre el subespacio generado por el autovector correspondiente.

Formas cuadráticas y diagonalización ortogonal

La clasificación de formas cuadráticas se puede hacer utilizando la diagonalización ortogonal de su matriz asociada.

En el capítulo 2 ya habíamos visto la clasificación de formas cuadráticas no degeneradas usando menores principales. El uso de autovalores permite clasificar también las formas cuadráticas degeneradas.

Recordemos que una forma cuadrática $\omega(x) = x^t Ax$ es degenerada si $\det(A) = 0$. 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 $\omega(x) > 0, \omega(y) < 0$.

Consideremos una forma cuadrática $\omega : \mathbb{R}^n \rightarrow \mathbb{R}$ definida por $\omega(x) = x^t Ax$, donde A es una matriz simétrica. 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 $A = PDP^t$. Entonces, para cada $x \in \mathbb{R}^n$:

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

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:

Clasificación de matrices simétricas según sus autovalores: Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ es una matriz simétrica entonces:

1. A es *definida positiva* si y solo si $\lambda > 0, \forall \lambda \in \text{Sp}(A)$.
2. A es *definida negativa* si y solo si $\lambda < 0, \forall \lambda \in \text{Sp}(A)$.
3. A es *semidefinida positiva* si y solo si $|A| = 0$ y $\lambda \geq 0, \forall \lambda \in \text{Sp}(A)$.
4. A es *semidefinida negativa* si y solo si $|A| = 0$ y $\lambda \leq 0, \forall \lambda \in \text{Sp}(A)$.
5. A es *indefinida* si y solo si A tiene autovalores positivos y negativos.

Ejemplo: La matriz

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

es semidefinida negativa ya que $\text{Sp}(A) = \{0, -3, -3\}$.

5.6. Descomposición en valores singulares

Recordemos que si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ 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, todos los autovalores de $A^t A$ son mayores o iguales que cero, como se deduce del siguiente argumento: si $\lambda \in \text{Sp}(A^t A)$ y x es un autovector asociado, entonces $A^t A x = \lambda x$ y por tanto:

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

Si $A \in \mathcal{M}_{p \times n}(\mathbb{R})$, se llaman **valores singulares** de A a las raíces cuadradas positivas de los autovalores de $A^t A$, es decir, si $\text{Sp}(A^t A) = \{\lambda_1, \lambda_2, \dots, \lambda_n\}$ entonces los valores singulares de A son $\sqrt{\lambda_1}, \sqrt{\lambda_2}, \dots, \sqrt{\lambda_n}$. Se suelen denotar por $\sigma_1, \sigma_2, \dots, \sigma_n$ 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 $r = \text{rg}(A)$.

Descomposición en valores singulares: Consideremos una matriz $A \in \mathcal{M}_{p \times n}(\mathbb{R})$ de rango r , con 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, $A = U\Sigma V^t$, con

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

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:

Si $A = U\Sigma V^t$ es una descomposición en valores singulares de una matriz A de rango r , u_1, u_2, \dots, u_r son las r primeras columnas de U y v_1, v_2, \dots, v_r son las r primeras columnas de V , 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:

Si $A = \sigma_1 u_1 v_1^t + \sigma_2 u_2 v_2^t + \dots + \sigma_r u_r v_r^t$ es la descomposición en valores singulares de una matriz A de rango r y k es un 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 $\mathcal{M}_{p \times n}(\mathbb{R})$ de modo que

$$\|A - A_k\| = \text{mín} \{ \|A - B\| / B \in \mathcal{M}_{p \times n}(\mathbb{R}), \text{rg}(B) = k \}.$$

Como en el caso vectorial, la norma de $A - A_k$ es una medida de lo próximas que están A y A_k .

La descomposición en valores singulares de una matriz A se suele llamar $\text{SVD}(A)$ (las iniciales de la traducción al inglés *Singular Value Decomposition*).

La determinación de las matrices U y V de la descomposición en valores singulares no es complicada, pero es un procedimiento que conlleva muchas operaciones y se realiza usando programas de cálculo en el ordenador.

5.7. Polinomios de matrices y el 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.

Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y $p(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_k x^k$ es un polinomio, se define

$$p(A) = a_0 I + a_1 A + a_2 A^2 + \dots + a_k A^k \in \mathcal{M}_{n \times n}(\mathbb{R}).$$

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

Ejemplo: El polinomio $p(x) = x^2 - 4x + 3$ es un polinomio anulador de la matriz

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

En efecto,

$$p(A) = A^2 - 4A + 3I = \begin{pmatrix} 5 & 4 \\ 4 & 5 \end{pmatrix} - 4 \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix} + 3 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}.$$

El teorema de Cayley-Hamilton establece que el polinomio característico de una matriz A es un polinomio anulador.

Teorema de Cayley-Hamilton: Si $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ 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 . En efecto, 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:

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

Llegamos así al siguiente resultado:

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

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:

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

(Nótese que $p'(x) = 10x^9 - 2$ y $r'(x) = b + 2cx$). La única solución del sistema es $a = 1$, $b = -2$, $c = 1$ y por tanto $r(x) = 1 - 2x + x^2$. Es decir, $p(A) = r(A) = I - 2A + A^2$.

5.8. 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 .

Consideremos una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y una función analítica $f : D \rightarrow \mathbb{R}$ 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 .

Si para una matriz $A \in \mathcal{M}_{n \times n}(\mathbb{R})$ y una función f 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.

Sin entrar en detalles técnicos, una justificación de la existencia de un polinomio $r(x)$ de grado $n - 1$ tal que $e^A = r(A)$ es la siguiente (un razonamiento análogo valdría para otras funciones de matrices):

En primer lugar, la función $f(x) = e^x$ admite un desarrollo en serie de potencias

$$e^x = \sum_{k=0}^{\infty} \frac{1}{k!} x^k = \lim_{m \rightarrow \infty} \left(\sum_{k=0}^m \frac{1}{k!} x^k \right).$$

Por tanto, e^x es el límite de polinomios de grado m , que podemos denotar por $p_m(x)$, es decir,

$$e^x = \lim_{m \rightarrow \infty} \left(\sum_{k=0}^m \frac{1}{k!} x^k \right) = \lim_{m \rightarrow \infty} p_m(x).$$

Para cada m , sabemos que existe un polinomio $r_m(x)$ de grado menor que n tal que $r_m(A) = p_m(A)$. Así,

$$f(A) = e^A = \lim_{m \rightarrow \infty} \left(\sum_{k=0}^m \frac{1}{k!} A^k \right) = \lim_{m \rightarrow \infty} p_m(A) = \lim_{m \rightarrow \infty} r_m(A) = r(A),$$

donde $r(x)$ es un polinomio de grado menor que n para el que se cumple que $V_{f,A} = V_{r,A}$, es decir,

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

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 $\text{Sp}(A) = \{0, 0, 0\}$. 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 usar este método para calcular una raíz cuadrada de la matriz

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

En efecto, consideremos la función $f(x) = \sqrt{x} = x^{1/2}$, de modo que $f(A) = A^{1/2}$.

Como $\text{Sp}(A) = \{0, 0\}$, los valores de f sobre el espectro de A son $f(0)$ y $f'(0)$. Pero no existe $f'(0)$ ya que $f'(x) = 1/(2\sqrt{x})$.

De hecho, no es difícil demostrar en este caso que la matriz A no tiene raíces cuadradas.

Funciones de matrices usando la diagonalización

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

Si D es una matriz diagonal y f es una función definida sobre D entonces $f(D)$ también es diagonal. En concreto,

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

Ejemplo: Para $f(x) = e^x$,

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

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

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 $f(A)$

El siguiente resultado permite obtener propiedades de la matriz $f(A)$ sin calcularla:

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:

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

En particular, estas propiedades permiten 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:

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

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

Por ejemplo, para

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

se tiene que $\text{Sp}(A) = \{-1, 1\}$. Por tanto, tomando $f(x) = e^x$:

$$|e^A| = |f(A)| = f(-1)f(1) = e^{-1}e^1 = e^{-1+1} = e^0 = 1.$$

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” (3^a ed.), Cengage Learning, 2011.