

# Álgebra Lineal

Francisco Javier Mercader Martínez

## Índice

1	Número reales y complejos	1
2	Espacios vectoriales de dimensión finita	13
3	Factorizaciones LU y Cholesky	35
4	Subespacios Vectoriales	49

## 1) Número reales y complejos

### Números reales y su representación en el ordenador

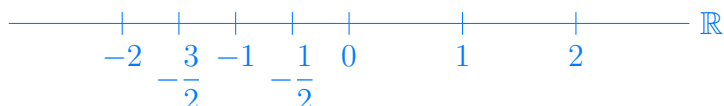
Recordaremos los conjuntos de números más habituales:

- Números naturales:  $\mathbb{N} = \{0, 1, 2, 3, \dots\}$
- Números enteros:  $\mathbb{Z} = \{\dots, -2, -1, 0, 1, 2, \dots\}$
- Números racionales:  $\mathbb{Q} = \left\{ \frac{n}{m} : n, m \in \mathbb{Z}, m \neq 0 \right\}$

Los números racionales se expresan en forma decimal número finito de cifras decimales (por ejemplo:  $\frac{1}{2} = 0.5$ )

o bien con un número infinito pero periódicas (e.g.,  $\frac{1}{3} = 0.333\dots$ )  
exempli gratia

- Números reales: incluye a todos los anteriores mas números racionales como  $\pi, e, \sqrt{2}$ , etc. Se denotan por  $\mathbb{R}$  y forman un continua, una "recta real"



Los números irracionales contienen un número infinito de cifras decimales no periódicas. Los ordenadores no pueden almacenar un número infinito de cifras decimales lo cual implica los llamados *"errores de representación"*, que se pueden ir propagando a medida que se hacen operaciones pudiendo dar resultados incorrectos.

## ¿Cómo se representan los número reales en un ordenador?

Hay varias posibilidades o formas. Todas ellas hacen uso del concepto de "bit".

Bit es el acrónimo de binary digit y representa un dígito en un sistema de representación binario, es decir, un bit es 0 ó 1.

En Computación, un bit es la unidad mínima de información. Se utiliza para representar la contraposición de dos valores (e.g. apagado y encendido, falso y verdadero, abierto y cerrado). Para obtener una mayor precisión (menor error de representación) lo más habitual es usar el formato de punto flotante según el estándar IEEE-754 de doble precisión (Python).

Se usan en total 64 bits para representar un número real  $x$  según

$$x = (-1)^s 2^{e-1023} (1 + f)$$

$s$  = signo. Se usa 1 bit. Por tanto  $s = 0$  ó  $1$ .

$e$  = exponente. Se usan 11 bits

$f$  = fracción. Se usan 52 bits

Ejemplo:

$$\underbrace{1}_s \quad \underbrace{10000000010}_e \quad \overbrace{100 \dots 0}^{52 \text{ dígitos}}_f$$

$$s = 1, \quad (-1)^s = -1 \text{ (número negativo)}$$

$$e = 1 \cdot 2^1 + 1 \cdot 2^0 = 1026$$

$$f = 1 \cdot \frac{1}{2^1} + 0 \cdot \frac{1}{2^2} + \dots + 0 \cdot \frac{1}{2^{52}} = 0.5$$

$$x = (-1)^1 \cdot 2^3 \cdot (1 + 0.5) = -12.0$$

Debido a este sistema de representación, los número representables están aproximadamente  $[-10^{200}, 10^{300}]$  con una precisión de 15 dígitos.

Además, estos números no están uniformemente distribuidos. Están más juntos los número pequeños y

más separados los grandes.

Se llama "*epsilon de la máquina*" al número  $\varepsilon$  representable más pequeño tal que  $1 + \varepsilon > 1$ .

$\varepsilon \approx 2.22 \cdot 10^{-16}$  en Python

- **Números complejos**

Origen histórico: En 1545 Cardano estudió ecuaciones algebraicas del tipo.

$$x^2 + 1 = 0$$

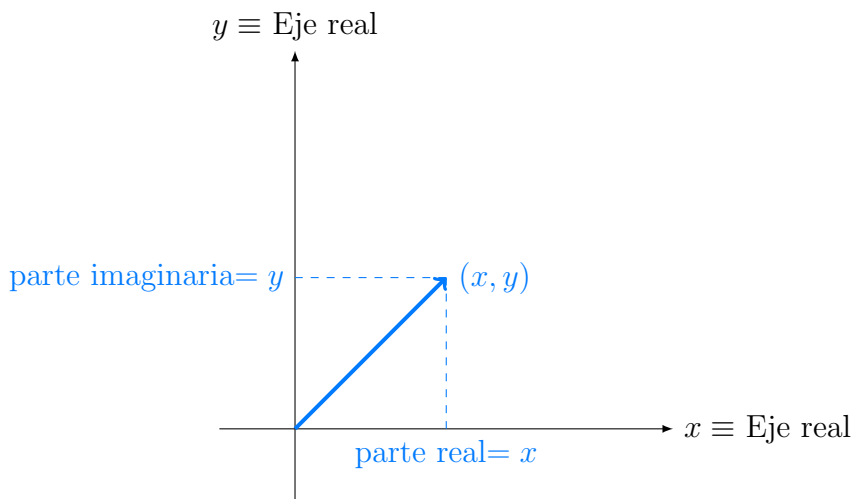
$$x^2 = -1 \longrightarrow x = \pm\sqrt{-1} \equiv i \equiv \text{imaginarios}$$

Su uso en Ciencia de Datos: Entre otros campos se usa en el tratamiento de señal mediante series y transformadas de Fourier.

- **Definición**

Se llama número complejo a todo par ordenado de números reales,  $z = (x, y)$ , con  $x, y \in \mathbb{R}$ . Por tanto, el conjunto de todos los números complejos

$$\mathbb{C} = \mathbb{R} \times \mathbb{R} = \{z = (x, y), \quad x, y \in \mathbb{R}\}$$



En  $\mathbb{C}$  tenemos definidas dos operaciones:

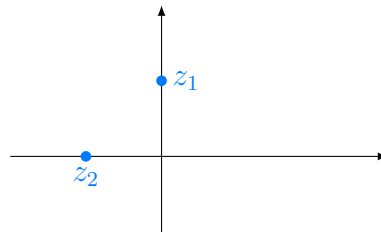
a) Suma:  $z_1 = (x_1, y_1), \quad z_2 = (x_2, y_2) \in \mathbb{C}$

$$z_1 + z_2 = (x_1 + x_2, \quad y_1 + y_2) \in \mathbb{C}$$

b) Producto:  $z_1 = (x_1, y_1), \quad z_2 = (x_2, y_2) \in \mathbb{C} \quad z_1 \cdot z_2 = (x_1 \cdot x_2 - y_1 \cdot y_2, \quad x_1 \cdot y_2 + y_1 \cdot x_2) \in \mathbb{C}$

Ejemplo:  $z_1 = (0, 1)$

$$z_2 = (0, 1) = z_1$$

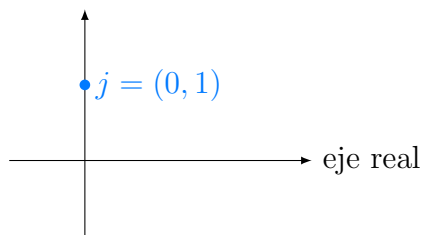


$$z_1^2 = z_1 \cdot z_2 = (0, 1) \cdot (0, 1) = (0 \cdot 0 - 1 \cdot 1, 0 \cdot 1 + 1 \cdot 0) = (-1, 0)$$

$$z_1^2 = -1 \rightarrow z_1 = \sqrt{-1} = i = j = (0, 1) \rightarrow \text{unidad imaginaria pura}$$

### Algunos números complejos destacados

eje imaginario



$$j = (0, 1) = \text{unidad imaginaria pura}$$

$\{(x, 0) : x \in \mathbb{R}\}$  los identificadores con los números reales  $(x, 0) \equiv x$ .

$\{(0, y) : y \in \mathbb{R}\}$  imaginarios puros.

Vamos a calcular:

$$\underbrace{(x, 0)}_x + \underbrace{(0, 1)}_j \cdot \underbrace{(0, y)}_y = (x, 0) + (0 \cdot y - 1 \cdot 0, 0 \cdot 0 + 1 \cdot y) = (x, 0) + (0, y) = (x, y)$$

$x \equiv$  parte real

$$y \equiv \text{parte imaginaria. Si } z = x + j \cdot y \begin{cases} x = \text{Re}z \\ y = \text{Im}z \end{cases}$$

$(x, y) \equiv$  forma cartesiana  
 $x + j \cdot y \equiv$  forma binómica

### Las operaciones suma y producto en forma binómica

$$\underbrace{(x_1 + j \cdot y_1)}_{(x_1, j_1)} + \underbrace{(x_2 + j \cdot y_2)}_{(x_2, j_2)} = (x_1 + x_2) + j \cdot (y_1 + y_2) = (x_1 + x_2, y_1 + y_2)$$

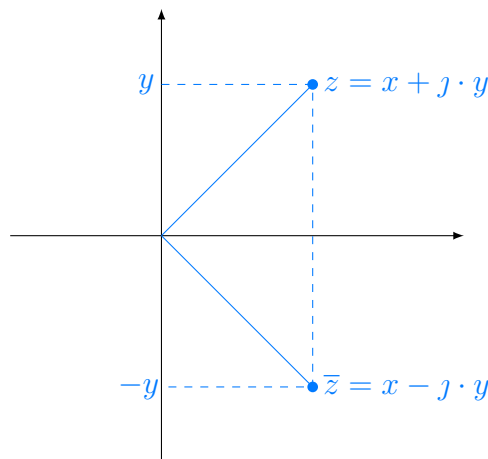
$$\underbrace{(x_1 + j \cdot y_1)}_{(x_1, j_1)} \cdot \underbrace{(x_2 + j \cdot y_2)}_{(x_2, j_2)} = (x_1 \cdot x_2 - y_1 \cdot y_2) + j \cdot (x_1 \cdot y_2 + y_1 \cdot x_2)$$

$\nwarrow j^2 = -1$

### Definición (conjugado)

Dado  $z = x + j \cdot y$ , se llama conjugado de  $z$ , denotado  $\bar{z}$ , al número complejo

$$\bar{z} = x - j \cdot y$$



### • Propiedades del conjugado

- a)  $\overline{\bar{z}} = z \quad \forall z \in \mathbb{C}$
- b)  $\overline{z + w} = \bar{z} + \bar{w} \quad \forall z, w \in \mathbb{C}$
- c)  $\overline{z \cdot w} = \bar{z} \cdot \bar{w} \quad \forall z, w \in \mathbb{C}$
- d)  $z \cdot \bar{z} = (\text{Re}z)^2 + (\text{Im}z)^2$

Veamos, por ejemplo, la última de estas propiedades

$$z = x + j \cdot y, \bar{z} = x - j \cdot y \longrightarrow z \cdot \bar{z} = (x + j \cdot y) \cdot (x - j \cdot y) = x^2 - jxy + jxy + y^2 = x^2 + y^2$$

### Definición (Inverso)

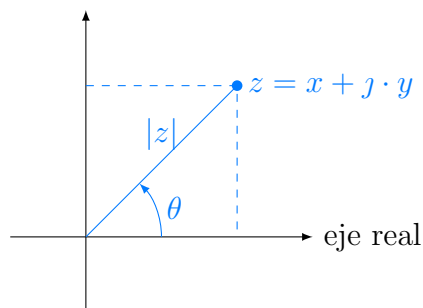
Dado  $z \in \mathbb{C}, z \neq 0$ , se llama inverso de  $z$ , denotado  $z^{-1}$  ó  $\frac{1}{z}$ , al número complejo  $z^{-1} = \frac{1}{z \cdot \bar{z}} \cdot z$

Vamos a calcular  $z \cdot z^{-1}$ :

$$z \cdot z^{-1} = z^{-1} = \frac{1}{z \cdot \bar{z}} \cdot z \cdot \bar{z} = \frac{\text{Re}z^2 + \text{Im}z^2}{\text{Re}z^2 + \text{Im}z^2} = 1$$

### Forma exponencial y polar de un número complejo

eje imaginario



- Se llama módulo de  $z$ , denotado  $|z|$ , al número real.  $|z| = +\sqrt{x^2 + y^2} = +\sqrt{z \cdot \bar{z}}$
- Dado  $z \in \mathbb{C}, z \neq 0$ , se define el argumento de  $z$ , denotado  $\arg z$ , como el conjunto

$$\arg z = \{\theta \in \mathbb{R} : \text{Re}z = |z| \cos \theta, \text{Im}z = |z| \sin \theta\}$$

Si  $\theta \in \arg z$ , entonces  $\theta + 2k\pi \in \arg z \forall k \in \mathbb{Z}$ .

Al único  $\theta \in \arg z$  tal que  $\theta \in [0, 2\pi[$  se le llama **argumento principal de  $z$** .

Por tanto,

$$z = \operatorname{Re} z + j \cdot \operatorname{Im} z = |z| \cos \theta + j \cdot |z| \sin \theta = |z|(\cos \theta + j \cdot \sin \theta) = |z|e^{j\theta}$$

donde  $e^{j\theta} = \cos \theta + j \cdot \sin \theta$

En resumen, podemos expresar  $z = \operatorname{Re} z + j \cdot \operatorname{Im} z$  como:

- a) Forma polar  $z = |z|_{\theta}$ .  
 $|z| = \text{módulo}$   
 $\theta = \text{argumento principal}$
- b) Forma exponencial  $z = |z| \cdot e^{j\theta}$
- c) Forma trigonométrica  $z = |z| \cdot (\cos \theta + j \cdot \sin \theta)$

Las formas exponencial y polar están especialmente diseñadas para realizar las operaciones producto y cociente de números complejos.

En efecto:

$$z_1 \cdot z_2 = |z_1| \cdot e^{j\theta_1} \cdot |z_2| \cdot e^{j\theta_2} = |z_1| \cdot |z_2| e^{j(\theta_1 + \theta_2)} = |z_1|_{\theta_1} \cdot |z_2|_{\theta_2} = |z_1| |z_2|_{\theta_1 + \theta_2}$$

es decir, se multiplican los módulos y se suman los argumentos.

La división o cociente de números complejos es multiplicar por el inverso. Nótese que si  $|z| \cdot e^{j\theta}$ , entonces

$$z^{-1} = \frac{1}{|z|} e^{-j\theta}. \text{ En efecto: } z \cdot z^{-1} = |z| e^{j\theta} = \frac{|z|}{|z|} e^{j(\theta - \theta)} = 1$$

Por tanto,

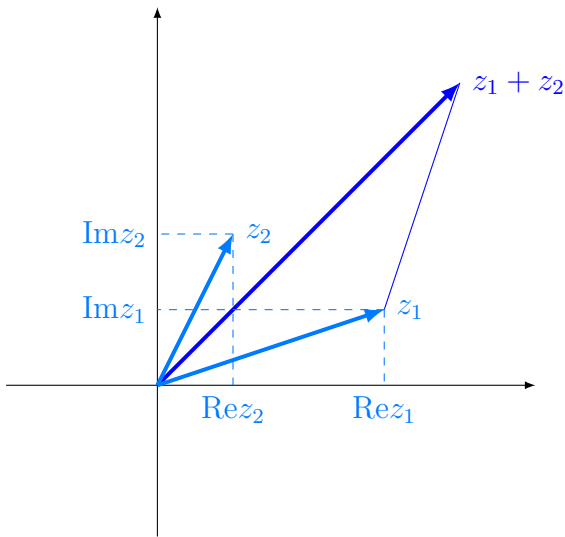
$$\frac{z_1}{z_2} = \frac{|z_1| e^{j\theta_1}}{|z_2| e^{j\theta_2}} = \frac{|z_1|}{|z_2|} e^{j(\theta_1 - \theta_2)},$$

es decir, se dividen los módulos y se restan los argumentos.

### Interpretación geométrica de la suma y el producto de números complejos

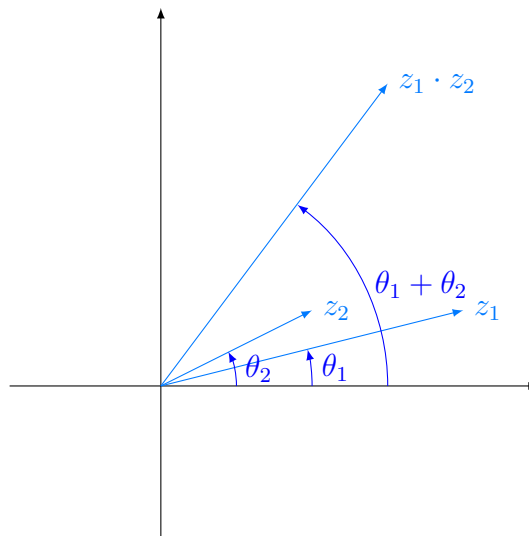
- **Suma (Forma binómica)**

$$\left. \begin{array}{l} z_1 = \operatorname{Re} z_1 + j \cdot \operatorname{Im} z_1 \\ z_2 = \operatorname{Re} z_2 + j \cdot \operatorname{Im} z_2 \end{array} \right\} z_1 + z_2 = (\operatorname{Re} z_1 + \operatorname{Re} z_2) + j \cdot (\operatorname{Im} z_1 + \operatorname{Im} z_2)$$



• Producto

$$\left. \begin{aligned} z_1 &= \operatorname{Re} z_1 + j \cdot \operatorname{Im} z_1 \\ z_2 &= \operatorname{Re} z_2 + j \cdot \operatorname{Im} z_2 \end{aligned} \right\} z_1 \cdot z_2 = |z_1| |z_2| \angle \theta_1 + \theta_2$$

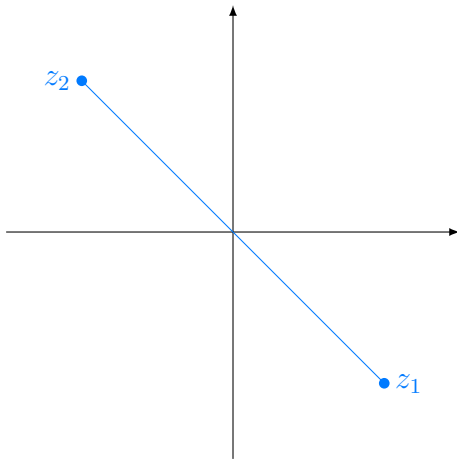


¿Cómo pasar los números complejos de una forma a otra?

Binómica  $\longrightarrow$  Polar/exponencial

$$z = x + j \cdot y \longrightarrow \begin{cases} |z| = +\sqrt{x^2 + y^2} \\ \theta = \arg z = \arctan \frac{y}{x} \end{cases}$$

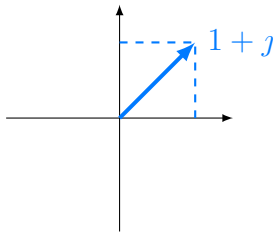
Las calculadoras solo devuelven valores del arcotangente en  $-90^\circ$  y  $90^\circ$ . Dependiendo de la localización del número complejo en cuestión hay que sumar  $180^\circ$ .



$z_1$  y  $z_2$  tienen el mismo arcotangente, según la calculadora.

Veamos algunos ejemplos:

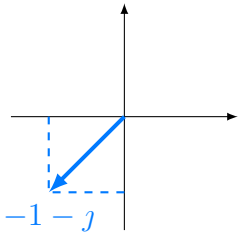
- $1 + j \rightarrow |1 + j| = +\sqrt{1^2 + 1^2} = \sqrt{2}$



$$\theta = \arctan \frac{1}{1} = 45^\circ = \frac{\pi}{4} \text{ radianes}$$

$$1 + j = \sqrt{2}_{45^\circ} = \sqrt{2}e^{j\frac{\pi}{4}}$$

- $-1 - j \rightarrow |-1 - j| = +\sqrt{(-1)^2 + (-1)^2} = +\sqrt{2}$



$$\theta = \arctan \frac{-1}{-1} = 45^\circ = 45^\circ + 180^\circ = 225^\circ = \frac{\pi}{4} + \pi = \frac{5\pi}{4} \text{ rad.}$$

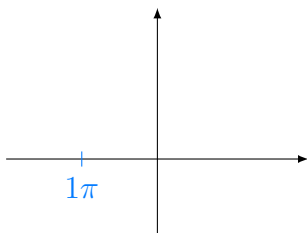
$$-1 - j = \sqrt{2}_{225^\circ} = +\sqrt{2}e^{j\frac{5\pi}{4}}$$

Polar—Exponencial—→Binómica

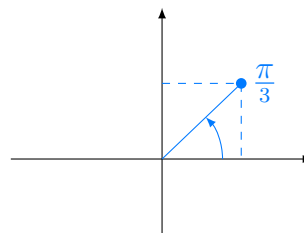
$$z = |z|_\theta = |z|e^{j\theta} \rightarrow z = |z|(\cos \theta + j \cdot \sin \theta) = |z| \cos \theta + j|z| \sin \theta$$

### Ejemplos

- $1\pi = 1 \cdot e^{j\pi} = \cos \pi + j \cdot \sin \pi = -1$



- $2\frac{\pi}{3} = 2 \cdot e^{j\frac{\pi}{3}} = 2(\cos \frac{\pi}{3} + j \cdot \sin \frac{\pi}{3}) = 2 \cdot (0.5 + j \cdot 0.866) = 1 + j \cdot 1.732$





En algunos campos de la Ciencia de Datos y la Inteligencia Artificial, e.g. reconocimiento del hablante o asistentes virtuales (Siri, en Apple), es necesario analizar y procesar señales, las cuales desde los tiempos de Fourier (siglo XIX), se procesan mediante una superposición de funciones trigonométricas con frecuencia discreta (series de Fourier) o continua (transformada de Fourier). En concreto, las señales periódicas se representan mediante series de Fourier y las no periódicas mediante transformadas de Fourier. Como veremos a continuación, los dos parámetros fundamentales de una onda son su frecuencia. Esto nos recuerda al módulo y argumento de un número complejo. Efectivamente, esto es así y es precisamente la forma exponencial de los números complejos la que facilita el procesamiento de estas señales.

Antes de nada, conviene señalar la conexión entre las funciones trigonométricas seno y coseno, y las exponenciales complejas.

Se tiene:

$$\cos \theta = \frac{e^{j\theta} + e^{-j\theta}}{2}$$

En efecto:

$$e^{j\theta} = \cos \theta + j \cdot \sin \theta$$

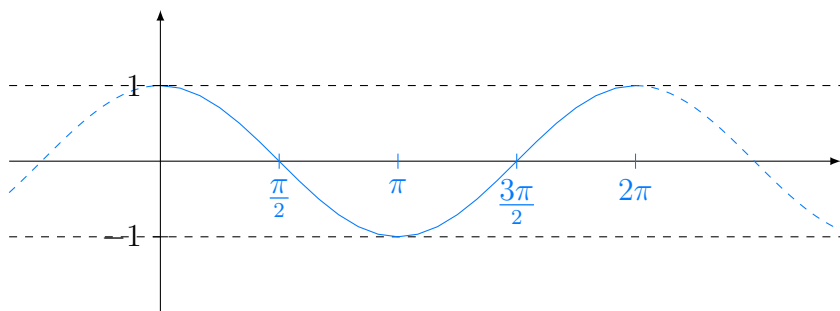
$$e^{-j\theta} = \cos(-\theta) + j \cdot \sin(-\theta) = \cos \theta - j \cdot \sin \theta$$

Por tanto:

$$e^{j\theta} + e^{-j\theta} = 2 \cos \theta$$

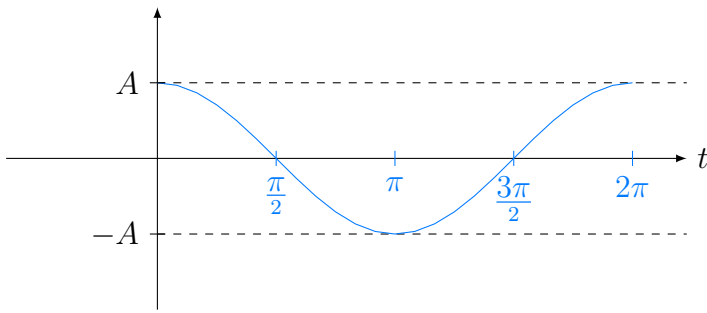
De igual forma,  $\sin \theta = \frac{e^{j\theta} - e^{-j\theta}}{2j}$  [Representación matemática de una honda elemental](#)

- La función  $\cos t = \frac{e^{jt} + e^{-jt}}{2}$ ,  $t \equiv$  tiempo.



Periodo  $T = 2\pi$

- La función  $A \cos t = \frac{Ae^{jt} + Ae^{-jt}}{2}$

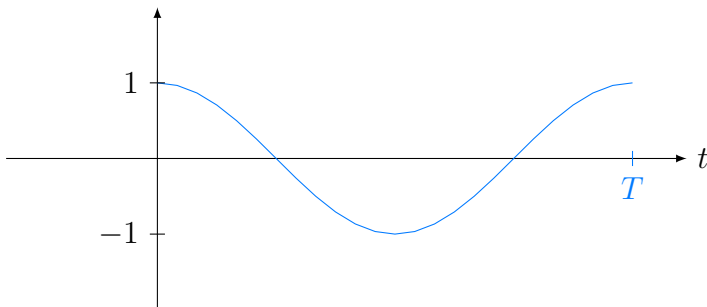


Periodo  $A \equiv$  Amplitud de la onda

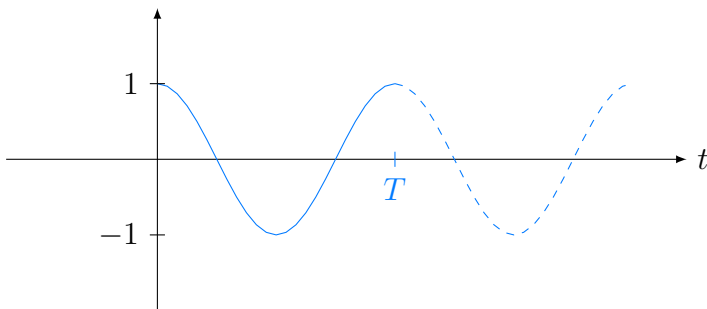
- La función  $\cos(wt) = \frac{e^{jwt} + e^{-jwt}}{2}$

Periodo de  $T$  es tal que  $wT = 2\pi \rightarrow T = \frac{2\pi}{w}$

Si  $w$  es pequeño  $\rightarrow T$  es grande (pocas oscilaciones)



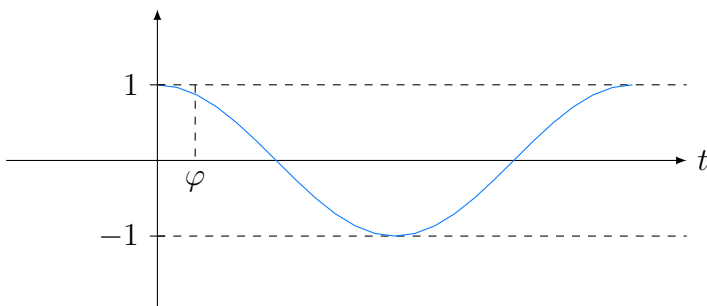
Si  $w$  es grande  $\rightarrow T$  es pequeño (muchas oscilaciones)



frecuencia  $f = \frac{1}{T} = \frac{w}{2\pi}$

velocidad angular  $\frac{2\pi}{T} = w$

- La función  $\cos(t - \varphi) = \frac{e^{j(t-\varphi)} + e^{-j(t-\varphi)}}{2}$



Si  $t = \varphi \longrightarrow \cos(t - \varphi) = \cos 0 = 1$

$\varphi \equiv$  fase inicial

$\varphi$  traslada la honda hacia la derecha ( $\varphi$  positiva) o hacia la izquierda ( $\varphi$  negativa).

### En resumen

$$f(t) = A \cos(\omega t - \varphi) = \frac{A \cdot (e^{j(\omega t - \varphi)} + e^{-j(\omega t - \varphi)})}{2}$$

Amplitud
Fase

Velocidad angular
(frecuencia y periodo)

### Raíces $n$ -ésimas y raíces de polinomios

Consideremos la ecuación  $x^2 + 1 = 0$ .

Como sabemos, no existe ningún número real que sea solución de dicha ecuación.

Veamos si hay algún número complejo.

Buscamos  $z = |z|e^{j\theta}$  tal que

$$z^2 = |z|^2 e^{2j\theta} = -1 = e^{j\pi}.$$

Para que dos números complejos sean iguales sus módulos han de ser iguales y sus argumentos diferir en un múltiplo de  $2\pi$ . Por tanto:

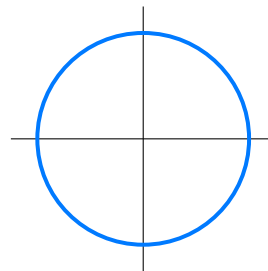
$$\left. \begin{array}{l} |z|^2 = 1 \\ 2\theta = \pi + 2k\pi \end{array} \right\} \quad \begin{array}{l} |z| = +\sqrt{1} = 1 \\ \theta = \frac{\pi}{2} + k\pi \end{array}$$

$$k = 0 \longrightarrow \theta_1 = \frac{\pi}{2}$$

$$k = 1 \longrightarrow \theta_2 = \frac{\pi}{2} + \pi = \frac{3\pi}{2}$$

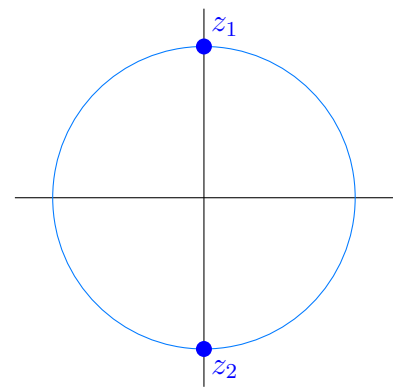
$$k = 2 \longrightarrow \theta_3 = \frac{\pi}{2} + 3\pi = \frac{7\pi}{2} = (\text{Como argumentos}) = \frac{3\pi}{2}$$

$$k = -1 \longrightarrow \theta_4 = \frac{\pi}{2} - \pi = -\frac{\pi}{2} = (\text{Como argumentos}) = \frac{3\pi}{2}$$



Por tanto, la ecuación  $x^2 + 1 = 0$  tiene dos soluciones en

$$\mathbb{C} : \begin{cases} z_1 = e^{j\frac{\pi}{2}} \\ z_2 = e^{j\frac{3\pi}{2}} \end{cases}$$



Veamos que los argumentos se van repartiendo.

Solo hay dos argumentos en  $[0, 2\pi[$ .

Se las llamas "raíces cuadráticas de la unidad".

Para la ecuación  $x^3 + 1 = 0$  procedemos de igual modo: buscamos  $z = |z|e^{j\theta}$  tal que  $z^3 = |z|^3 e^{3j\theta} = -1 = e^{j\pi}$ .

Por tanto:

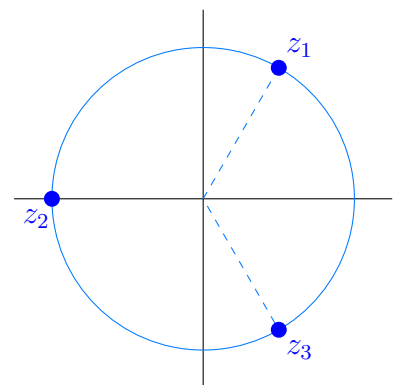
$$|z|^3 = 1 \longrightarrow |z| = 1$$

$$3\theta = \pi + 2k\pi \longrightarrow \theta = \frac{\pi}{3} + \frac{2k\pi}{3}$$

$$k = 0 \longrightarrow \theta_1 = \frac{\pi}{3}$$

$$k = 1 \longrightarrow \theta_2 = \frac{\pi}{3} + \frac{2\pi}{3}$$

$$k = 2 \longrightarrow \theta_3 = \frac{\pi}{3} + \frac{4\pi}{3}$$



Las raíces cúbicas de la unidad son  $\left\{ e^{j\frac{\pi}{3}}, e^{j\pi}, e^{j\left(\frac{\pi}{3} + \frac{4\pi}{3}\right)} \right\}$

En general, la ecuación  $x^n + 1 = 0$  tiene n-raíces en  $\mathbb{C}$ , llamadas raíces  $n$ -ésimas de la unidad y son:

$$\left\{ z = |z|e^{j\theta} : |z| = 1 = \frac{\pi}{n} + \frac{2k\pi}{n}, k = 0, 1, \dots, n-1 \right\}$$

Para completar, se puede demostrar (lo hizo Gauss en su tesis doctoral) que todo polinomio de grado  $n$  tiene directamente n-raíces (reales y/o complejas). Es el llamado Teorema Fundamental del Álgebra.

## 2) Espacios vectoriales de dimensión finita

### Vectores

Sea  $\mathbb{K} = \mathbb{R}$  ó  $\mathbb{C}$ . A los elementos de  $\mathbb{K}$  les llamamos **escalares**.

- Definición (Espacio vectorial de dimensión finita)

Sea  $n \in \mathbb{N}$ . Definimos

$$\mathbb{K}^n = \{(x_1, x_2, \dots, x_n) : x_i \in \mathbb{K}, 1 \leq i \leq n\}$$

A los elementos  $v = (x_1, \dots, x_n) \in \mathbb{K}^n$  se les llama vectores de componentes en  $\mathbb{K}$ .

En Python se les llaman 1d-array.

En  $\mathbb{K}^n$  tenemos definidas dos operaciones:

- Suma de vectores:

$$u + v = (x_1, x_2, \dots, x_n) + (y_1, y_2, \dots, y_n) = (x_1 + y_1, x_2 + y_2, \dots, x_n + y_n)$$

- Producto por escalares:  $\lambda \in \mathbb{K}, u \in \mathbb{K}^n$

$$\lambda \cdot u = \lambda \cdot (x_1, \dots, x_n) = (\lambda x_1, \lambda x_2, \dots, \lambda x_n) \in \mathbb{K}^n$$

Al conjunto de vectores  $\mathbb{K}^n$  con las dos operaciones anteriores se le llama **espacio vectorial sobre  $\mathbb{K}$** .

### Dependencia lineal

- Definición (Combinación lineal)

Dados vectores  $u, u_1, u_2, \dots, u_m \in \mathbb{K}^n$ , se dice que  $u$  es **combinación lineal** de  $u_1, u_2, \dots, u_m$  si existen escalares  $\lambda_1, \lambda_2, \dots, \lambda_m \in \mathbb{K}$  de modo que  $u$  puede expresarse como

$$\lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_m u_m = 0 = (0, \dots, 0)$$

se cumple que  $\lambda_1 = \lambda_2 = \dots = \lambda_m = 0$ .

Es decir, la única forma de obtener una combinación lineal nula es como todos los coeficientes nulos.

- Demostración

→ Supongamos que  $\{u_1, u_2, \dots, u_m\}$  es linealmente independiente y que

$$\lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_m u_m = 0$$

Si algún  $\lambda_i$  fuese nulo, por ejemplo  $\lambda_1 \neq 0$ , entonces

$$u_1 = \frac{\lambda_2}{\lambda_1}u_2 + \cdots + \frac{\lambda_m}{\lambda_1}u_m$$

y así  $u_1$  es combinación lineal de los demás.

—→ Recíprocamente, supongamos que si

$$\lambda_1 u_1 + \lambda_2 u_2 + \cdots + \lambda_m u_m = 0$$

entonces  $\lambda_1 = \lambda_2 = \cdots = \lambda_m = 0$

El conjunto  $\{u_1, u_2, \dots, u_m\}$  ha de ser linealmente independiente pues si no lo fuera entonces al giro de sus vectores, por ejemplo  $u_2$ , se podría expresar

$$u_1 = \lambda_2 u_2 + \cdots + \lambda_m u_m$$

y entonces

$$u_1 - \lambda_2 u_2 - \cdots - \lambda_m u_m = 0$$

sería una combinación lineal nula con al menos un coeficiente, el 1 que acompaña a  $u_1$ , no nulo.

- Definición (Rotación de dependencia)

Dados unos vectores  $u_1, u_2, \dots, u_m$  linealmente dependientes, se llama **relación de dependencia** a cualquier combinación lineal nula

$$\lambda_1 u_1, \lambda_2 u_2, \dots, \lambda_m u_m = 0$$

con no todos los coeficientes nulos.

- Definición (Base canónica)

El conjunto de vectores

$$C = \{e_1 = (1, 0, 0, \dots, 0), e_2 = (0, 1, 0, \dots, 0), \dots, e_n = (0, 0, \dots, 0, 1) \subset \mathbb{K}^n\}$$

se llama **base canónica** de  $\mathbb{K}^n$ .

Se trata de un conjunto de vectores linealmente independiente y cada vector  $u = (a_1, a_2, \dots, a_n) \in \mathbb{K}^n$  se escribe como combinación lineal en  $C$  en la forma

$$u = a_1 e_1 + a_2 e_2 + \cdots + a_n e_n$$

- Proposición

Sean  $S = \{u_1, u_2, \dots, u_m\}$  un conjunto de vectores linealmente independientes y  $u \in \mathbb{K}^n$ . Entonces:

1. Todo subconjunto de  $S$  no vacía es linealmente independiente.
2. El conjunto  $S \cap \{u\}$  es linealmente independiente si y sólo si  $u$  no es combinación lineal de  $u_1, u_2, \dots, u_m$ .

### Norma y producto escalar

#### • Definición (Norma euclídea)

Sea  $u = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$ . Se llama norma euclídea o norma 2, escalar.

$$\|u\|_2 = +\sqrt{x_1^2 + \dots + x_n^2}$$

Si  $\|u\|_2 = 1$ ,  $u$  se dice unitario.

Otras normas utilizadas en Ciencia de Datos son:

$$\|u\|_1 = |x_1| + |x_2| + \dots + |x_n| = \sum_{j=1}^n |x_j|$$

$$\|u\|_p = (|u_1|^p + |u_2|^p + \dots + |u_n|^p)^{\frac{1}{p}}, \quad 1 < p < +\infty$$

$$\|u\|_\infty = \max\{|x_1| + |x_2| + \dots + |x_n|\} \quad 1 \leq j \leq n \longrightarrow \text{Norma del máximo o norma del infinito.}$$

#### • Ejemplo

$$u = (1, -2, -1)$$

$$\|u\|_1 = |1| + |-2| + |-1| = 4$$

$$\|u\|_2 = \sqrt{1^2 + (-2)^2 + (-1)^2} = +\sqrt{6}$$

$$\|u\|_\infty = \max\{|1|, |-2|, |-1|\} = 2$$

Todas las normas comparten las siguientes propiedades:

- Sean  $\|\cdot\|$  cualquiera de las normas anteriores,  $u \in \mathbb{R}^n$  y  $\alpha \in \mathbb{R}$ . Entonces:

a)  $\|u\| \geq 0$  y  $\|u\| = 0$  si y sólo si  $u = 0$ .

b)  $\|\alpha u\| = |\alpha| \cdot \|u\|$

c)  $\|u + v\| \leq \|u\| + \|v\|$  (desigualdad triangular)

$\longrightarrow$  A partir de ahora, salvo que se indique lo contrario,  $\|\cdot\| = \|\cdot\|_2$  la norma euclídea.

#### • Definición (Producto escalar euclídeo)

Dados  $u = (x_1, x_2, \dots, x_n), v = (y_1, y_2, \dots, y_n) \in \mathbb{R}^n$ , se define el producto escalar euclídeo de  $u$  por  $v$  como

$$u \cdot v = x_1 \cdot y_1 + x_2 \cdot y_2 + \dots + x_n \cdot y_n$$

Nótese que entonces  $\|u\|_2 = \sqrt{u \cdot u}$

- **Propiedades del producto escalar**

a)  $(u_1 + u_2) \cdot v = u_1 \cdot v + u_2 \cdot v.$

b)  $u \cdot (v_1 + v_2) = u \cdot v_1 + u \cdot v_2.$

c)  $(\alpha \cdot u) \cdot v = u \cdot (\alpha \cdot v) = \alpha(u \cdot v).$

d)  $u \cdot v = v \cdot u.$

e)  $u \cdot u \geq 0$  y  $u \cdot u = 0$  si y sólo si  $u = 0$ .

- **Propiedades de la norma euclídea.**

1.  $\|u + v\|^2 = \|u\|^2 + \|v\|^2 + 2(u \cdot v)$

2.  $\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2(u \cdot v)$

3.  $|u \cdot v| \leq \|u\| \cdot \|v\| \longrightarrow$  Desigualdad de Cauchy-Schwarz

- **Demostración**

Veamos el punto (3).

Si  $u = 0$  los dos miembros son nulos y se cumple la desigualdad. Supongamos  $v \neq 0$ . Entonces:

$$\begin{aligned} 0 &\leq \left(u - \frac{u \cdot v}{\|v\|^2} v\right) \cdot \left(u - \frac{u \cdot v}{\|v\|^2} v\right) = u \cdot \left(u - \frac{u \cdot v}{\|v\|^2} v\right) - \frac{u \cdot v}{\|v\|^2} \cdot v \cdot \left(u - \frac{u \cdot v}{\|v\|^2} v\right) = \\ &u \cdot u - \frac{u \cdot v}{\|v\|^2} \cdot u \cdot v - \frac{u \cdot v}{\|v\|^2} \cdot v \cdot u + \frac{u \cdot v}{\|v\|^2} \cdot \frac{u \cdot v}{\|v\|^2} \cdot v \cdot v = \|u\|^2 - \frac{2(u \cdot v)^2}{\|v\|^2} + \frac{(u \cdot v)^2}{\|v\|^2} = \|u\|^2 - \frac{|u \cdot v|^2}{\|v\|^2} \end{aligned}$$

Nótese que como consecuencia de la desigualdad de Cauchy-Schwarz, si  $u, v$  son vectores no nulos, entonces

$$-1 \leq \frac{u \cdot v}{\|u\| \cdot \|v\|} \leq 1.$$

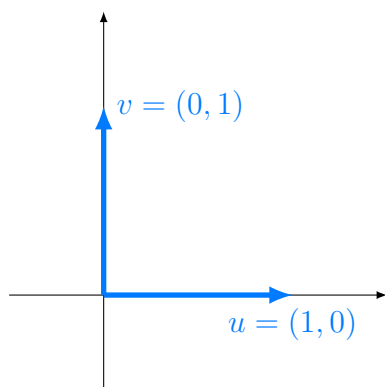
Por tanto, existe un único ángulo  $0 \leq \varphi \leq \pi$  tal que  $\cos \varphi = \frac{u \cdot v}{\|u\| \cdot \|v\|}$ .

A este ángulo se le llama ángulo formado por los vectores  $u$  y  $v$  y su coseno se denota  $\cos(u, v)$ . Así,

$$u \cdot v = \|u\| \cdot \|v\| \cos(u, v).$$

- **Ejemplo:**





$$u \cdot v = (1, 0) \cdot (0, 1) = 1 \cdot 0 + 0 \cdot 1 = 0$$

$$\cos(u, v) = 0 \longrightarrow \varphi = 90^\circ = \frac{\pi}{2}$$

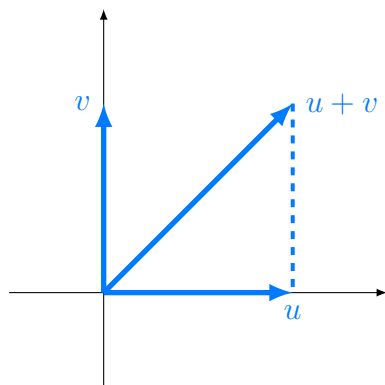
$$\|u\| = \|v\| = 1$$

$u \cdot v$  son perpendiculares (ortogonales).

- Definición (ortogonalidad)

Dos vectores  $u, v \in \mathbb{K}^n$  se llaman ortogonales o perpendiculares si  $u \cdot v = 0$ .

Como consecuencia  $\|u \pm v\|^2 = \|u\|^2 + \|v\|^2$  si  $u$  y  $v$  son perpendiculares (Teorema de Pitágoras).



- Definición

Un conjunto de vectores  $\{u_1, \dots, u_n\} \subset \mathbb{K}^n$  se dice **ortogonal** si  $u_i \cdot u_j = 0, i \neq j$ . Si además los vectores son unitarios, el conjunto se dice ortonormal.

- Ejemplo:

La base canónica  $C = \{e_1, \dots, e_n\}$  es ortonormal.

- Propiedades

Todo conjunto ortogonal es linealmente independiente.

- Demostración

Sea  $\{u_1, \dots, u_m\}$  ortogonal.  $\lambda_1 u_1 + \lambda_2 u_2 + \dots + \lambda_m u_m = 0$ . Multiplicando por  $u_j$ :

$$\lambda_1 u_1 u_j + \lambda_2 u_2 u_j + \dots + \lambda_j u_j \cdot u_j + \dots + \lambda_m u_m \cdot u_j = 0 \longrightarrow \lambda_j \|u_j\|^2 = 0 \longrightarrow \lambda_j = 0$$

## Matrices

- Definición: (Matriz de tamaño  $m \times n$ )

Sean  $m, n \in \mathbb{N}$ . Se llama matriz de tamaño  $m \times n$  al conjunto formado por  $m$ -vectores de  $n$ -componentes apilados verticalmente. En Python, a las matriz se les llama 2d-array.

- Ejemplo (Matriz de Hilbert)

Matriz  $H$  con entradas  $H_{ij} = \frac{1}{i+j-1}$

$$H_4 = \begin{bmatrix} 1 & \frac{1}{2} & \frac{1}{3} & \frac{1}{4} \\ \frac{1}{2} & \frac{1}{3} & \frac{1}{4} & \frac{1}{5} \\ \frac{1}{3} & \frac{1}{4} & \frac{1}{5} & \frac{1}{6} \\ \frac{1}{4} & \frac{1}{5} & \frac{1}{6} & \frac{1}{7} \end{bmatrix} \begin{matrix} \leftarrow \text{filas} \\ \\ \\ \leftarrow \text{columnas} \end{matrix}$$

Denotaremos por  $M_{m \times n}(\mathbb{K})$  al conjunto de matrices de  $m$ -filas y  $n$ -columnas con entradas en  $\mathbb{K} = \mathbb{R}$  ó  $\mathbb{C}$ .

Si  $m = n$ , es decir, la matriz cuadrada (tiene el mismo número de filas que de columnas), entonces escribimos  $M_n(\mathbb{K})$ .

- Algunos tipos importantes de matrices

- Diagonal:  $A = (a_{ij})_{m \times n}$ ,  $a_{ij} = 0$  si  $i \neq j$ .  $A = \begin{pmatrix} a_{11} & & 0 \\ & a_{22} & \\ 0 & & \ddots \\ & & & a_{mn} \end{pmatrix}$

- Escalar:  $A \in M_n(\mathbb{K})$ ,  $A$  diagonal  $a_{ii} = \alpha$ ,  $1 \leq i \leq n$ .

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

- Matriz identidad: matriz escalar con  $\alpha = 1$

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

- Triangular superior (o inferior):  $A \in M_{m \times n}(\mathbb{K})$

$$A = \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ & a_{22} & \cdots & a_{2n} \\ & & \ddots & \\ 0 & & & a_{mn} \end{pmatrix} \leftarrow \text{Triangular superior}$$

- Matriz a banda (banded matrix)

$$A = \begin{pmatrix} & & & 0 \\ \text{en} & \diagdown & & \\ \text{general} & & & \\ & \diagup & \text{no} & \\ & & \text{nulos} & \\ 0 & \diagup & & \end{pmatrix}$$

## • Operaciones con matrices

→ Suma:  $A = (a_{ij}), B = (b_{ij}) \in M_{m \times n}(\mathbb{K})$

$$A + B = C, C = (C_{ij}) \in M_{m \times n}(\mathbb{K}) \longrightarrow C_{ij} = a_{ij} + b_{ij}$$

Ejemplo:

$$A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix}, B = \begin{bmatrix} 5 & 6 \\ 7 & 8 \end{bmatrix} \longrightarrow A + B = \begin{bmatrix} 6 & 8 \\ 10 & 12 \end{bmatrix}$$

→ Producto de un escalar por una matriz.

$$\alpha \in \mathbb{K}, A = (a_{ij}) \in M_{m \times n}(\mathbb{K}) \quad (\alpha A) = (\alpha a_{ij})$$

$$t_j : A = \begin{bmatrix} 1 & 2 \\ 3 & 4 \end{bmatrix} \quad \alpha = 2 \longrightarrow (\alpha A) = \begin{bmatrix} 2 & 6 \\ 6 & 8 \end{bmatrix}$$

Propiedades de la suma y el producto por escalares

1.  $A + B = B + A$
2.  $(A + B) + C = A + (B + C)$
3.  $(\alpha + \beta) \cdot A = \alpha A + \beta A$
4.  $\alpha(A + B) = \alpha A + \alpha B$
5.  $\alpha(\beta A) = (\alpha \cdot \beta)A$

→ Producto de matrices

$$A = (a_{ij}) \in M_{m \times n}(\mathbb{K}), B = (b_{ij}) \in M_{m \times n}(\mathbb{K})$$

$$A \cdot B = (C_{ij}) \in M_{m \times k}$$

$$C_{ij} = \sum_{k=1}^n a_{ik} \cdot b_{kj}$$

Ejemplo:

$$A = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix}_{1 \times 3}, \quad B = \begin{bmatrix} 1 \\ 0 \\ 1 \end{bmatrix}_{3 \times 1}$$

$\nearrow (A \cdot B)_{1 \times 1} = (2)$   
 $\searrow (B \cdot A)_{3 \times 3} = \begin{pmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{pmatrix}$

- En Python, este producto de matrices se llama **dotproduct**.
- En Ciencia de Datos también se trabaja con otros productos diferentes de matrices.

Propiedades del producto de matrices

1.  $(A \cdot B) \cdot C = A(B \cdot C)$
2.  $A \cdot 0 = 0 \cdot A = 0$
3.  $A(B + C) = A \cdot B + A \cdot C$
4.  $\alpha(A \cdot B) = (\alpha \cdot A) \cdot B$
5.  $A \cdot I = I \cdot A = A$

→ Como hemos visto antes, el producto de matrices no es conmutativo, es decir, en general  $A \cdot B \neq B \cdot A$ .

#### • Definición (Traspuesta)

Dada  $A = (a_{ij}) \in M_{m \times n}(\mathbb{K})$ , se llama traspuesta de  $A$  a la matriz.

$$A^T = (a_{ji}) \in M_{n \times m}(\mathbb{K}),$$

es decir, la traspuesta consiste en cambiar filas por columnas.

Una matriz se dice simétrica si  $A = A^T$  y antisimétrica si  $A = -A^T$ .

Ejemplos:

$$A = \begin{bmatrix} a & x & y \\ x & b & z \\ y & z & c \end{bmatrix} \quad A^T = \begin{bmatrix} a & x & y \\ x & b & z \\ y & z & c \end{bmatrix}$$

$$A = A^T \longrightarrow A \text{ simétrica}$$

$$B = \begin{bmatrix} 0 & j & -2 \\ -j & 0 & 5 \\ 2 & -5 & 0 \end{bmatrix} \quad B^T = \begin{bmatrix} 0 & -j & 2 \\ j & 0 & -5 \\ -2 & 5 & 0 \end{bmatrix}$$

$$B = -B^T \longrightarrow B \text{ antisimétrica}$$

### Propiedades de la traspuesta

1.  $(A^T)^T = A$
2.  $(A + B)^T = A^T + B^T$
3.  $(\alpha A)^T = \alpha A^T$
4.  $(A \cdot B)^T = B^T \cdot A^T$

### • Definición (Inversa)

Dada  $A = (a_{ij}) \in M_n(\mathbb{K})$ , se dice que  $A$  es invertible si existe una matriz  $B \in M_n(\mathbb{K})$  tal que

$$A \cdot B = B \cdot A = I.$$

Dicha matriz  $B$ , si existe es única y se denota  $A^{-1}$ .

### Propiedades de la inversa

Sean  $A, B \in M_n(\mathbb{K})$  invertibles. Se tiene:

1.  $A \cdot B$  es invertible y  $(A \cdot B)^{-1} = B^{-1} \cdot A^{-1}$
2.  $A^T$  es invertible y  $(A^T)^{-1} = (A^{-1})^T$ .
3. Una matriz triangular (superior o inferior) es invertible si y solo si los elementos de la diagonal son todos no nulos.
4. Sean  $u_1, u_2, \dots, u_m$  vectores de  $\mathbb{K}^n$  que representamos como vectores columna, y sea  $A \in M_n(\mathbb{K})$  invertible. Entonces  $u_1, u_2, \dots, u_m$  son linealmente independientes si y sólo si  $Au_1, Au_2, \dots, Au_m$  son linealmente independientes.

### Veamos la demostración de 4

→ Supongamos que  $u_1, u_2, \dots, u_m$  y veamos que  $Au_1, Au_2, \dots, Au_m$  son linealmente independientes.

Tomamos una combinación lineal nula

$$\lambda_1 Au_1, \lambda_2 Au_2, \dots, \lambda_m Au_m = 0$$

Por las propiedades del producto de matrices:

$$A(\lambda_1 u_1, \dots, \lambda_m u_m) = 0$$

Como  $A$  es invertible:

$$A^{-1}A(\lambda_1 u_1, \dots, \lambda_m u_m) = 0 \longrightarrow \lambda_1 u_1, \dots, \lambda_m u_m = 0$$

Y como  $u_1, u_2, \dots, u_m$  son linealmente independientes.

$$\lambda_1 = \lambda_2 = \dots = \lambda_m = 0$$

### • Definición (Matriz ortogonal)

Sea  $Q \in M_n(\mathbb{R})$ . Se dice que  $Q$  es ortogonal si es invertible (o no singular) y si su inversa coincide con su traspuesta, es decir,  $Q^T \cdot Q = Q \cdot Q^T = I$ .

Como consecuencia, si  $Q$  es ortogonal, entonces las columnas de  $Q$  forman un conjunto ortogonal de vectores.

### Ejemplo

$$Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \quad Q^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$Q^T \cdot Q = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \quad Q \cdot Q^T = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

$$u_1 = \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), u_2 = \left( -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \longrightarrow \|u_1\| = \|u_2\| = 1 \longrightarrow u_1 \cdot u_2 = 0$$

### • Definición (Traza de una matriz)

Dada  $A = (a_{ij}) \in M_n(\mathbb{K})$ , se define la traza de  $A$ , denotada  $tr(A) = \sum_{i=1}^n a_{ii}$ , es decir, la traza escalar que se obtiene sumando los elementos de la diagonal.

### Propiedades de la traza

1.  $tr(A + B) = tr(A) + tr(B)$
2.  $tr(\alpha A) = \alpha \cdot tr(A)$
3.  $tr(A) = tr(A^T)$
4.  $tr(A \cdot B) = tr(B \cdot A)$
5.  $tr(P^{-1} \cdot A \cdot P) = tr(A)$  para cualquier matriz invertible.

### Matrices por bloques

$$\text{Sea } A = \left[ \begin{array}{cccc} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{array} \right] \begin{array}{l} \longleftarrow u_1^T \\ \longleftarrow u_2^T \\ \cdots \\ \longleftarrow u_m^T \end{array}$$

$m \times n$

En algunas ocasiones conviene escribir  $A$  en la forma

$$A = [v_1, v_2, \dots, v_n]$$

donde  $v_i$  son las columnas de  $A$  o también

$$A = \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix}$$

Con  $u_j^T$  las filas de  $A$ .

Con esta notación, el producto de una matriz por un vector columna es:

$$\overbrace{\begin{bmatrix} -1 & 1 \\ 2 & 3 \end{bmatrix}}^{\text{Bloques}} \cdot \overbrace{\begin{bmatrix} a \\ b \end{bmatrix}}^x = a \cdot \begin{bmatrix} -1 \\ 2 \end{bmatrix} + b \cdot \begin{bmatrix} 1 \\ 3 \end{bmatrix}$$

$\begin{array}{ccc} \uparrow & \uparrow & \uparrow \\ v_1 & v_2 & \text{coeficiente} \end{array}$ 
 $a \cdot v_1 + b \cdot v_2$

es decir, el producto por la derecha de una matriz  $B$  por un vector columna  $x$  es una combinación lineal de las columnas de  $A$  con coeficientes los elementos de  $x$ .

De manera similar, el producto de un vector fila por una matriz:

$$[a, b] \cdot \begin{bmatrix} -1 & 1 \\ 2 & 3 \end{bmatrix} = a \cdot [-1, 1] + b \cdot [2, 3]$$

es decir, una combinación lineal de las filas de la matriz con coeficientes los componentes del vector.

Combinando ambas tenemos:

$$\begin{bmatrix} 1 & 4 \\ 2 & 5 \\ 3 & 6 \end{bmatrix} \cdot \begin{bmatrix} 7 & 8 & 9 \\ 10 & 11 & 12 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix} \cdot \begin{bmatrix} 7 & 8 & 9 \end{bmatrix} + \begin{bmatrix} 4 \\ 5 \\ 6 \end{bmatrix} \cdot \begin{bmatrix} 10 & 11 & 12 \end{bmatrix}$$

que podemos expresar:

$$A = [v_1, \dots, v_n], \quad B = \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} \quad A \cdot B = v_1 u_1^T + \dots + v_n u_m^T$$

Es habitual en Machine Learning tener que separar los datos contenidos en una matriz en varios bloques. Por ejemplo, en primer lugar los datos de la matriz se dividen en dos bloques: uno de ellos se utiliza para "entrenar" un modelo de predicción, y el otro para "testear" cuánto de bueno es dicho modelo. Es lo que se llama descomponer un [dataset](#) en [data training](#) y [data test](#).

$$\text{Sea } A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix}$$

training  
test

Además en el llamado "[supervised learning](#)" todas las columnas excepto la última contienen un tipo de datos llamados "[features](#)" y la última columna los llamados "[labels](#)". Tenemos así:

$$\text{Sea } A = \begin{bmatrix} a_{11} & \cdots & a_{1,n-1} & a_{1n} \\ a_{21} & \cdots & a_{2,n-1} & a_{2n} \\ \vdots & & \vdots & \vdots \\ a_{m1} & \cdots & a_{m,n-1} & a_{mn} \end{bmatrix}$$

features       lable



Así,  $A$  queda descompuesto en 4 bloques.

$$\text{Sea } A = \left[ \begin{array}{c|c} \text{features} & \text{labels} \\ \text{training} & \text{training} \\ \hline \text{features} & \text{labels} \\ \text{test} & \text{test} \end{array} \right] = \left[ \begin{array}{c|c} & \\ \hline A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array} \right]$$

El producto de matrices opera por bloques de igual manera a como se hace por entradas.

### Operaciones y matrices elementales

Muchos de los cálculos habituales del Álgebra Lineal requieren realizar las siguientes "operaciones elementales" sobre las filas y/o columnas de una matriz:

1. Intercambiar dos filas (columnas)
2. Multiplicar una fila (columna) por un escalar.
3. Añadir a cada fila (columna) otra fila (columna) multiplicada por un escalar.

Estas operaciones se pueden realizar de manera sistemática multiplicando la matriz en cuestión por una determinada matriz, que llamaremos **matriz elemental**.

Hay tres tipos de matrices elementales:

1. Matriz de **permutación simple** ( $P_{st}$ ) obtenida al intercambiar las filas (o columnas)  $st$ .

Ejemplo:

$$A = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}; P_{12} = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \quad \begin{array}{l} \leftarrow 1, \quad 1 \rightarrow 2 \\ \leftarrow 1, \quad 2 \rightarrow 1 \end{array}$$

$$P_{12} \cdot A = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{bmatrix}$$

$$\begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \xrightarrow{F_1 \leftrightarrow F_2} \begin{bmatrix} a_{21} & a_{22} \\ a_{11} & a_{12} \end{bmatrix};$$

2. Matriz de **dilatación**  $D_s(r)$

$$D_2(5) = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}; \quad D_2(5) \cdot A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ 5a_{21} & 5a_{22} \end{bmatrix}$$

3. Matriz de **adición**:

$S_{st}(r)$  = a la fila  $s$  se le suma la fila  $t$  multiplicada por  $r$ .

$$S_{12}(2) = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix}; \quad S_{12}(2) \cdot A = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} = \begin{bmatrix} a_{11} + 2a_{21} & a_{12} + 2a_{22} \\ a_{21} & a_{22} \end{bmatrix};$$

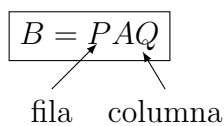
• **Propiedades:**

Las matrices elementales son invertibles. Además:

- 1)  $P_{st}^{-1} = P_{st}$
- 2)  $D_s(r)^{-1} = D_s \frac{1}{r}$
- 3)  $S_{st}(r)^{-1} = S_{st}(-r)$

En algunas ocasiones es necesario combinar operaciones elementales fila y también columna sobre una misma matriz  $A$ . De esta forma obtenemos una nueva matriz  $B$ , equivalente a  $A$ , y que se expresan en la forma.

$$\boxed{B = PAQ}$$



• **Definición (Matriz de permutación)**

Llamaremos matriz de permutación  $P$  a aquella que se expresa como producto de matrices de permutación elementales.

Matrices escalonadas. Rango

• **Definición**

Dada una matriz se llama **pivote** de una fila no nula a la primera entrada (contando izquierda a derecha) no nula a esa fila. Se dice que la **matriz** es **escalonada** por filas si:

- 1) El pivote de cada fila no nula está estrictamente a la derecha del pivote de la fila anterior.
- 2) Las filas nulas, si las hay, son las últimas.

Se dice que la matriz es **escalonada reducida** por filas si además:

1º) El pivote de cada fila no nula vale 1.

2º) Cada pivote es el único no nulo de su columna.

De manera análoga se puede definir el concepto de matriz escalonada (reducida) por columnas.

Ejemplo

$$A = \underbrace{\begin{bmatrix} 2 & 0 & 0 & 5 \\ 0 & 1 & 0 & 2 \\ 0 & 3 & 1 & 4 \end{bmatrix}}_{\text{No escalonada}} \quad B = \underbrace{\begin{bmatrix} 1 & 2 & 0 & 4 & 5 \\ 0 & 0 & -1 & 1 & 5 \\ 0 & 0 & 0 & 2 & 2 \end{bmatrix}}_{\text{Escalonada pero no reducida}} \quad C = \underbrace{\begin{bmatrix} 1 & 0 & 2 & 0 & 5 \\ 0 & 1 & 6 & 0 & 2 \\ 0 & 0 & 0 & 1 & 4 \end{bmatrix}}_{\text{Escalonada reducida}}$$

• Teorema (Factorización  $PAQ$ )

Toda matriz  $A$  es equivalente a una matriz de la forma

$$B = \left[ \begin{array}{c|c} Ir & 0 \\ \hline 0 & 0 \end{array} \right],$$

es decir, existen matrices invertibles  $P$  y  $Q$  tales que

$$B = PAQ$$

→ La forma de llegar de  $A$  a  $B$  es realizando operaciones elementales. Este proceso se llama **eliminación gaussiana**

Al número  $r$ , es decir, al número de filas no nulas que resultan después de escalonar una matriz se le llama **rango** de la matriz y se denota por  $r = \text{rg}(A)$

Ejemplo

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

Hacemos operaciones elementales fila con la matriz.

$$(A|I_3) = \left[ \begin{array}{cccc|cccc} 1 & 2 & -1 & 2 & 1 & 0 & 0 \\ 2 & 4 & 1 & 1 & 0 & 1 & 0 \\ 3 & 6 & 0 & 3 & 0 & 0 & 1 \end{array} \right] \xrightarrow[F_3 \rightarrow F_3 - 3F_1]{F_2 \rightarrow F_2 - 2F_1} \left[ \begin{array}{cccc|cccc} 1 & 2 & -1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 3 & -3 & -2 & 1 & 0 \\ 0 & 0 & 3 & -3 & -3 & 0 & 1 \end{array} \right] \xrightarrow{F_3 \rightarrow F_3 - F_2}$$

$$\left[ \begin{array}{cccc|ccc} 1 & 2 & -1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 3 & -3 & -2 & 1 & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 \end{array} \right]$$

El rango de  $A$  es  $r = \text{rg}(A) = 2$

Seguimos haciendo operaciones fila para conseguir la forma reducida, donde los pivotes son 1s y cada pivote es el único elemento no nulo de su columna.

$$\xrightarrow{F_2 \rightarrow \frac{1}{3}F_2} \left[ \begin{array}{cccc|ccc} 1 & 2 & -1 & 2 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & -\frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 \end{array} \right] \xrightarrow{F_1 \rightarrow F_1 + F_2} \left[ \begin{array}{cccc|ccc} 1 & 2 & 0 & 1 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 1 & -1 & -\frac{2}{3} & \frac{1}{3} & 0 \\ 0 & 0 & 0 & 0 & -1 & -1 & 1 \end{array} \right]$$

Se tiene que

$$\left[ \begin{array}{cccc} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{array} \right] = PA, \text{ con } \left[ \begin{array}{ccc} \frac{1}{3} & \frac{1}{3} & 0 \\ -\frac{2}{3} & \frac{1}{3} & 0 \\ -1 & -1 & 1 \end{array} \right]$$

Ahora realizamos operaciones columna para obtener la forma  $\left[ \begin{array}{c|c} Ir & 0 \\ \hline 0 & 0 \end{array} \right]$

$$\left[ \begin{array}{cccc} 1 & 2 & 0 & 1 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{C_2 \leftrightarrow C_3} \left[ \begin{array}{cccc} 1 & 0 & 2 & 1 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 0 \\ \hline 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow[\substack{C_3 \rightarrow C_3 - 2C_1 \\ C_4 \rightarrow C_4 - C_1}]{C_3 \rightarrow C_3 - 2C_1} \left[ \begin{array}{cccc} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ \hline 1 & 0 & -2 & -1 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{array} \right]$$

$$Q = \left[ \begin{array}{cccc} 1 & 0 & -2 & -1 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{array} \right] \quad \left[ \begin{array}{c|c} Ir_2 & 0 \\ \hline 0 & 0 \end{array} \right] = PAQ$$

## [Sistemas de ecuaciones](#)

### • Definición

Un sistema lineal de  $m$ -ecuaciones con  $n$ -incógnitas es un sistema del tipo

$$(*) \left\{ \begin{array}{l} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n = b_2 \\ \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n = b_m \end{array} \right.$$

que en forma matricial escribimos como

$$(*) \quad Ax = b$$

donde

- $A = (a_{ij})$  es la matriz del sistema
- $b = (b_{ij})$  es el término independiente
- $x = (x_{ij})$  es el vector incógnita

El sistema (\*) se dice:

- Homogéneo: si  $b = 0$
- Incompatible: si no tiene solución
- Compatible determinado: si tiene una única solución.
- Compatible indeterminado: infinitas soluciones.

- Teorema (Rouché-Frobenius)

Consideremos el sistema (\*) y sea  $(A|b)$  la llamada matriz ampliada. Entonces:

- 1) Si  $r(A) = r(A|b) = \text{número de incógnitas}$ , entonces el sistema  $(*)$  es compatible determinado (CD).
- 2) Si  $r(A) = r(A|b) < \text{número de incógnitas}$ , entonces el sistema  $(*)$  es compatible indeterminado (CI) y la solución depende del número de incógnitas  $- r(A)$  parámetros.
- 3) Si  $r(A) \neq r(A|b)$ , el sistema  $(*)$  es incompatible.

- Resolución de sistemas lineales mediante el método de Gauss

Consiste en realizar operaciones elementales fila sobre la matriz ampliada  $(A|b)$  hasta conseguir una matriz escalonada, cuyo sistema asociado tiene la misma solución que el sistema original, pero que se resuelve muy fácilmente.

### Ejemplo

$$\left. \begin{array}{l} 2x + y + 2z + 3t = 6 \\ x + 2y - 2z - t = 3 \\ 4x + 5y - 2z + t = 12 \end{array} \right\} (A|b) = \left[ \begin{array}{cccc|c} 2 & 1 & 2 & 3 & 6 \\ 1 & 2 & 2 & -1 & 3 \\ 4 & 5 & -2 & 1 & 12 \end{array} \right] \xrightarrow{F_1 \leftrightarrow F_2} \left[ \begin{array}{cccc|c} 1 & 2 & 2 & -1 & 3 \\ 2 & 1 & 2 & 3 & 6 \\ 4 & 5 & -2 & 1 & 12 \end{array} \right] \xrightarrow{\begin{array}{l} F_2 \rightarrow F_2 - 2F_1 \\ F_3 \rightarrow F_3 - 4F_1 \end{array}} \left[ \begin{array}{cccc|c} 1 & 2 & 2 & -1 & 3 \\ 0 & -3 & 6 & 5 & 0 \\ 0 & -3 & 6 & 5 & 0 \end{array} \right] \xrightarrow{F_3 \rightarrow F_3 - F_2} \left[ \begin{array}{cccc|c} 1 & 2 & 2 & -1 & 3 \\ 0 & -3 & 6 & 5 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]$$

$\text{rg}(A) = 2 = \text{rg}(A|b) < 4 = n^0$  de incógnitas  $\longrightarrow$  S.C.I tomamos 4-2

parámetros  $t = \lambda$ ,  $z = \mu$

$$\left. \begin{array}{l} x + 2y - 2\mu - \lambda = 3 \\ -3y + 6\mu + 5\lambda = 0 \end{array} \right\}$$

$$y = 2\mu + \frac{5}{3}\lambda$$

$$x = 3 - 2\left(2\mu + \frac{5}{3}\lambda\right) + 2\mu + \lambda = 3 - 4\mu - \frac{10}{3}\lambda + 2\mu + \lambda = 3 - 6\mu - \frac{7}{3}\lambda$$

$$\begin{array}{l} x = 3 - \frac{7}{3}\lambda - 6\mu \\ y = \frac{5}{3}\lambda + 2\mu \\ z = \mu \\ t = \lambda \end{array}$$

### • Cálculo de la inversa de una matriz mediante eliminación gaussiana

$\longrightarrow$  Método: realizar operaciones elementales fila sobre la matriz  $(A|I)$  hasta conseguir que a la izquierda aparezca la identidad. La matriz que resulta a la derecha es  $A^{-1}$

Ejemplo

$$A = \left[ \begin{array}{ccc} 1 & 2 & -1 \\ 2 & 4 & 1 \\ 3 & 7 & 0 \end{array} \right]$$

$$\left[ \begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 2 & 4 & 1 & 0 & 1 & 0 \\ 3 & 7 & 0 & 0 & 0 & 1 \end{array} \right] \xrightarrow{\begin{array}{l} F_2 \rightarrow F_2 - 2F_1 \\ F_3 \rightarrow F_3 - 3F_1 \end{array}} \left[ \begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 0 & 3 & -2 & 1 & 0 \\ 0 & 1 & 3 & -3 & 0 & 1 \end{array} \right] \xrightarrow{F_2 \leftrightarrow F_3} \left[ \begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & 3 & -3 & 0 & 1 \\ 0 & 0 & 3 & -2 & 1 & 0 \end{array} \right]$$

$$\xrightarrow{\begin{array}{l} F_2 \rightarrow F_2 - F_3 \\ F_3 \rightarrow \frac{1}{3}F_3 \end{array}} \left[ \begin{array}{ccc|ccc} 1 & 2 & -1 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & -\frac{2}{3} & \frac{1}{3} & 0 \end{array} \right] \xrightarrow{F_1 \rightarrow F_1 + F_3} \left[ \begin{array}{ccc|ccc} 1 & 2 & 0 & \frac{1}{3} & \frac{1}{3} & 0 \\ 0 & 1 & 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & -\frac{2}{3} & \frac{1}{3} & 0 \end{array} \right] \xrightarrow{F_1 \rightarrow F_1 - 2F_2}$$

$$\left[ \begin{array}{ccc|ccc} 1 & 0 & 0 & \frac{7}{3} & \frac{7}{3} & -2 \\ 0 & 1 & 0 & -1 & -1 & 1 \\ 0 & 0 & 1 & -\frac{2}{3} & \frac{1}{3} & 0 \end{array} \right] \quad A^{-1} = \left[ \begin{array}{ccc} \frac{7}{3} & \frac{7}{3} & -2 \\ -1 & -1 & 1 \\ -\frac{2}{3} & \frac{1}{3} & 0 \end{array} \right]$$

Determinantes

Sea  $A \in M_2(\mathbb{K})$ . Se define el determinante de  $A$ , denotado  $|A|$  o  $\det(A)$ , como

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

Sea  $A \in M_3(\mathbb{K})$ . Se define el determinante de  $A$  mediante la llamada regla de Sarrus como

$$\begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} = a_{11}a_{22}a_{33} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - (a_{13}a_{22}a_{31} + a_{12}a_{21}a_{33} + a_{11}a_{23}a_{32})$$

Para matrices de orden superior, el determinante se define de manera recurrente.

• **Definición**

Sea  $A = (a_{ij}) \in M_n(\mathbb{K})$ .

Se llama **menor complementario** el elemento  $a_{ij}$  al determinante  $(|A_{ij}|)$  de la submatriz  $A_{ij}$  que se obtiene la fila  $i$  y la columna  $j$  de la matriz original  $A$ .

Se llama **adjunto** de  $a_{ij}$ , denotado  $\Delta_{ij}$ , al escalar  $\Delta_{ij} = (-1)^{i+j} \cdot |A_{ij}|$ .

Ejemplo

$$A = \begin{bmatrix} 1 & 2 & 3 & 4 \\ 2 & 3 & 4 & 5 \\ 2 & 0 & 1 & 1 \\ 4 & 3 & 1 & 2 \end{bmatrix} \quad A_{23} = \begin{vmatrix} 1 & 2 & 4 \\ 2 & 0 & 1 \\ 4 & 3 & 2 \end{vmatrix} = 21 \longrightarrow \Delta_{23} = (-1)^{2+3} \cdot A_{23} = -21$$

• **Definición**

Sea  $A \in M_n(\mathbb{K})$ . El determinante de  $A$  se aplica de forma recurrente como

- Si  $n = 1$ ,  $|(a)| = a$
- Si  $n > 1$ ,  $|A| = a_{i1}\Delta_{i1} + a_{i2}\Delta_{i2} + \cdots + a_{in}\Delta_{in}$

Ejemplo

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

Desarrollado por la primera fila:

$$|A| = 1 \cdot \begin{vmatrix} 0 & 1 \\ 3 & 2 \end{vmatrix} - 2 \cdot \begin{vmatrix} 2 & 1 \\ 4 & 2 \end{vmatrix} + 4 \cdot \begin{vmatrix} 2 & 0 \\ 4 & 3 \end{vmatrix} = -3 - 2(4 - 4) + 4 \cdot 6 = 21$$

Y si desarrollamos por la segunda columna:

$$|A| = -2 \cdot \begin{vmatrix} 2 & 1 \\ 4 & 2 \end{vmatrix} + 0 \cdot \begin{vmatrix} 2 & 1 \\ 4 & 2 \end{vmatrix} - 3 \cdot \begin{vmatrix} 1 & 4 \\ 2 & 1 \end{vmatrix} = -3 \cdot (-7) = 21$$

### Propiedades básicas

- 1)  $|A| = |A^T|$
- 2)  $|A \cdot B| = |A| \cdot |B|$
- 3)  $A$  no es singular si y sólo si  $|A| \neq 0$ . En este caso  $|A^{-1}| = \frac{1}{|A|}$
- 4) Si  $Q$  es ortogonal, entonces  $|Q| = I \cdot 1$ .

→ Veamos 4)

Si  $Q$  es ortogonal, entonces  $Q^{-1} = Q^T$ .

$$I = Q \cdot Q^{-1} \longrightarrow 1 = |I| = |Q \cdot Q^{-1}| = |Q| \cdot |Q^{-1}| = |Q| \cdot |Q^T| = |Q| \cdot |Q| = |Q|^2$$

Por tanto  $Q = \pm 1$

### Propiedades de los determinantes y las operaciones elementales

- 1)  $|[u_1, u_2, \dots, u_j + u'_j, \dots, u_n]| = |[u_1, \dots, u_j, \dots, u_n]| + |[u_1, \dots, u'_j, \dots, u_n]|$
- 2)  $|[u_1, u_2, \dots, \alpha u_j, \dots, u_n]| = \alpha |[u_1, u_2, \dots, u_n]|$
- 3)  $|[u_1, \dots, u_i, \dots, u_j, \dots, u_n]| = -|[u_1, \dots, u_j, \dots, u_i, \dots, u_n]|$
- 4) Si a una columna (o fila) se le suma otra columna (o fila) multiplicada por un escalar, entonces el determinante no cambia.

### Ejemplo

$$\begin{vmatrix} a & a+1 & a+2 \\ a+3 & a+4 & a+5 \\ a+6 & a+7 & a+8 \end{vmatrix} \xrightarrow[F_3 \rightarrow F_3 - F_1]{F_2 \rightarrow F_2 - F_1} \begin{vmatrix} a & a+1 & a+2 \\ 3 & 3 & 3 \\ 6 & 6 & 6 \end{vmatrix} \longrightarrow 2 \cdot \begin{vmatrix} a & a+1 & a+2 \\ 3 & 3 & 3 \\ 3 & 3 & 3 \end{vmatrix} \xrightarrow{F_3 \rightarrow F_3 - F_2} \begin{vmatrix} a & a+1 & a+2 \\ 3 & 3 & 3 \\ 0 & 0 & 0 \end{vmatrix} = 0$$



## Resolución de sistemas lineales usando determinantes

Consideramos el sistema  $Ax = b$  y denotamos por  $\Delta_i$  el determinante de la matriz  $A$  donde se ha sustituido la columna  $i$ -ésima por  $b$ , es decir,

$$\Delta_i = [u_1, \dots, \underset{\substack{\nearrow \\ i\text{-ésimo}}}{b}, \dots, u_n]$$

Entonces, si  $|A| \neq 0$  se cumple  $x_i = \frac{a_i}{|A|}$  (regla de Cramer)

### Ejemplo

$$\left. \begin{array}{l} x + y + z = 1 \\ -x \quad + z = 1 \\ x + y - z = 1 \end{array} \right\} \longrightarrow |A| = \begin{vmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & -1 \end{vmatrix} = 1 - 1 - (1 + 1) = -2$$

$$\Delta_1 = \begin{vmatrix} 1 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & -1 \end{vmatrix} = 1 + 1 - (1 - 1) = 2$$

$$\Delta_2 = \begin{vmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 1 & 1 & -1 \end{vmatrix} = -1 + 1 - 1 - (1 + 1 + 1) = -4$$

$$\Delta_3 = \begin{vmatrix} 1 & 1 & 1 \\ -1 & 0 & 1 \\ 1 & 1 & 1 \end{vmatrix} = 0$$

$$x_1 = \frac{\Delta_1}{|A|} = \frac{2}{-2} = -1, \quad x_2 = \frac{\Delta_2}{|A|} = \frac{-4}{-2} = 2, \quad x_3 = \frac{\Delta_3}{|A|} = 0.$$

### Cálculo de la inversa mediante determinantes

#### • Definición (Matriz adjunta)

Sea  $A \in M_n(\mathbb{K})$ . Se llama matriz adjunta de  $A$ , denotada  $\tilde{A}$ , a la matriz cuyas entradas son los adjuntos de las entradas de  $A$ , es decir, si  $A = (a_{ij})$ , entonces  $\tilde{A} = (\Delta_{ij})$ .

#### • Proposición

Si  $A \in M_n(\mathbb{K})$  es invertible, entonces  $A^{-1} = \frac{1}{|A|} \cdot \tilde{A}^T$

### Ejemplo

Sea  $Q = \begin{bmatrix} a & c \\ b & d \end{bmatrix}$  una matriz ortogonal con determinante  $|Q| = +1$

Vamos a calcular su inversa:

$$\Delta_{11} = |d| = d, \quad \Delta_{12} = -b$$

$$\Delta_{21} = -c \quad \Delta_{22} = a$$

Por tanto,

$$Q^{-1} = \underbrace{\frac{1}{|Q|}}_1 \cdot \begin{bmatrix} d & -c \\ -b & a \end{bmatrix} = \begin{bmatrix} d & -c \\ -b & a \end{bmatrix}$$

y como  $Q^T = Q^{-1}$ , al ser  $Q$  ortogonal,  $d = a$ ,  $-c = b$ .

En resumen,

$$Q^{-1} = \begin{bmatrix} a & b \\ -b & a \end{bmatrix} \text{ con } a^2 + b^2 = 1$$

### 3) Factorizaciones LU y Cholesky

#### Álgebra lineal computacional

Problemas básicos:

- Resolución de sistemas lineales
- Cálculo de la inversa
- Cálculo de determinantes
- Cálculo de vectores propios

Todo ello para [matrices de gran tamaño](#).

Ejemplo: Cálculo del determinante de una matriz cuadrada de tamaño  $n = 100$ .

Usando el método estudiado hasta ahora, el número de operaciones necesarias es aproximadamente.

$$100! \approx 9.3 \times 10^{157}$$

El Marenostrum 4 de Barcelona Supercomputing Center (uno de los más potentes a nivel mundial) tiene una capacidad de cálculo de

$$11.15 \text{ Petaflops (11,500 billones de operaciones por segundo).}$$

Por tanto, para calcular el determinante anterior tardaría  $\frac{9.3 \cdot 10^{157}}{1.115 \cdot 10^{16}}$  segundos que son  $2.6 \cdot 10^{134}$  años.

En cambio, usando la factorización  $LU$  que veremos a continuación, este cálculo es casi instantáneo.

#### Factorización $LU$

Dada una matriz  $A \in M_{n \times m}(\mathbb{K})$ , se llama factorización  $LU$  a una factorización de la forma

$$A = LU$$

donde  $L$  es una matriz triangular inferior con 1s en la diagonal principal, y  $U$  es una matriz triangular superior. ( $L \equiv \text{Lower}$ ,  $U \equiv \text{Upper}$ ). Si  $A$  es cuadrada:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ l_{21} & 1 & 0 & \cdots & 0 \\ l_{31} & l_{32} & 1 & \cdots & 0 \\ \vdots & & & \ddots & \\ l_{n1} & \cdots & l_{n, n-1} & & 1 \end{bmatrix} = \begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ & u_{22} & \cdots & u_{2n} \\ & & \ddots & \\ 0 & & & u_{nn} \end{bmatrix}$$

De esta forma, el sistema  $Ax = b$  se reescribe como  $LUx = b$  y llamando  $z = Ux$  se llega a los sistemas

$$\begin{cases} Lz = b \\ Ux = z \end{cases}$$

que son dos sistemas fácilmente resolubles al ser  $L$  y  $U$  triangulares inferior y superior, respectivamente. Para el cálculo del determinante de  $A$ , se tiene

$$|A| = |LU| = |L| \cdot |U| = 1 \cdot |U| = \prod_{j=1}^n u_{jj}$$

$$\text{Volviendo al sistema } Ax = b \longleftrightarrow \begin{cases} Lz = b \\ Ux = z \end{cases}$$

Veamos cómo se resuelve un sistema triangular y el coste computacional que lleva conflicto

$$\begin{bmatrix} u_{11} & u_{12} & \cdots & u_{1n} \\ 0 & u_{22} & \cdots & u_{2n} \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & u_{nn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

Se resuelven en cascada, de abajo hacia arriba.

$$u_{nn}x_n \longrightarrow x_n = \frac{z_n}{u_{nn}}$$

$$u_{n-1, n-1}x_{n-1} + u_{n-1, n}x_n = z_{n-1} \longrightarrow x_{n-1} = \frac{z_{n-1} - u_{n-1, n} \cdot x_n}{u_{n-1, n-1}}$$

En general,

$$x_i = \frac{z_i - \sum_{j=i+1}^n u_{ij}x_j}{u_{ii}}$$

Por tanto, el algoritmo es:

$$\begin{cases} x_n = \frac{z_n}{u_{nn}} \\ z_i - \sum_{j=i+1}^n u_{ij}x_j \\ x_i = \frac{\quad}{u_{ii}} \end{cases}$$

## Análisis del coste computacional

Consideremos el caso número 4.

$$\left. \begin{aligned} u_{11}x_1 + u_{12}x_2 + u_{13}x_3 + u_{14}x_4 &= z_1 \\ u_{22}x_2 + u_{23}x_3 + u_{24}x_4 &= z_2 \\ u_{33}x_3 + u_{34}x_4 &= z_3 \\ u_{44}x_4 &= z_4 \end{aligned} \right\}$$

$$x_4 = \frac{z_4}{u_{44}} \quad 1 \text{ Operación}$$

$$x_3 = \frac{z_3 - u_{34}x_4}{u_{33}} \quad 3 \text{ Operaciones}$$

$$x_2 = \frac{z_2 - u_{23}x_3 - u_{24}x_4}{u_{22}} \quad 5 \text{ Operaciones}$$

$$x_1 = \frac{z_1 - u_{12}x_2 - u_{13}x_3 - u_{14}x_4}{u_{11}} \quad 7 \text{ Operaciones}$$

En total,  $16 = 4^2$  operaciones

Se observa que para calcular  $x_i$ ; necesitamos  $2(n-i)+1$  operaciones.

Por tanto, el número total de operaciones es

$$\begin{aligned} \sum_{i=1}^n [2(n-i)+1] &= 2 \cdot \sum_{i=1}^n (n-i) + n = 2 \cdot [(n-1) + (n-2) + \dots + (n-n)] + n = 2(n^2 - \underbrace{(1+2+\dots+n)}_S) + n \\ n &= 2 \left( n^2 - \frac{n(n+1)}{2} \right) + n = \frac{2}{2}(2n^2 - n^2 - n) + n = n^2 = O(n^2) \end{aligned}$$

↑ notación de Landau

$$S = 1 + 2 + \dots + n$$

$$S = n + n - 1 + \dots + 1$$

---


$$2S = n \cdot (n+1) \longrightarrow S = \frac{n(n+1)}{2}$$

- ¿Cómo se calculan  $L$  y  $U$  tales que  $A = LU$ ?

Haciendo eliminación Gaussiana. Veamos un ejemplo:

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \text{ matriz de Toeplitz en dimensi3n 3.}$$

$$\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow{F_2 \rightarrow F_2 + \frac{1}{2}F_1} \begin{bmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & -1 & 2 \end{bmatrix} \xrightarrow{F_3 \rightarrow F_2 + \frac{2}{3}F_2} \begin{bmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix}$$

La matriz  $U$  es

$$U = \begin{bmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix}$$

• ¿Donde est3a  $L$ ?

Adem3as de los pivotes, en la eliminaci3n gaussiana juegan un papel los llamados **multiplicadores**.

Una vez conocemos el pivote de la fila  $j$ , y la entrada que se desea eliminar en la fila  $i$ , el multiplicador  $l_{ij}$  se define como

$$l_{ij} = \frac{\text{entrada a eliminar en la fila } i\text{-3sima}}{\text{pivote en la fila } j}$$

As3, en el ejemplo anterior:

$$l_{21} = \frac{-1}{2}, \quad l_{31} = \frac{0}{2} = 0, \quad l_{32} = \frac{-1}{\frac{3}{2}} = -\frac{2}{3}$$

La matriz  $L$  es justo la matriz de multiplicadores en la eliminaci3n gaussiana. En nuestro caso:

$$L = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix}$$

En efecto:

$$LU = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix} \cdot \begin{bmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix} = A$$

Veamos ahora una forma algor3tmica de calcular la factorizaci3n  $LU$ .

Sean  $A = (a_{ij}) \in M_{m \times n}$ ,  $L = (l_{ij}) \in M_{m \times n}$

$U = (u_{ij}) \in M_{m \times n}$

Entonces, si  $A = LU$ , se tiene:

$$a_{ij} = \sum_{k=1}^n l_{ik} u_{kj}$$

Como  $l_{ik} = 0$  si  $k > i$  y  $u_{kj} = 0$  si  $k > j$ , entonces podemos distinguir dos casos:

**Caso 1:**  $i > j \rightarrow \min\{i, j\} = j$

$$a_{ij} = \sum_{k=1}^j l_{ik} u_{kj} = l_{i1} u_{1j} + l_{i2} u_{2j} + \cdots + l_{ij} u_{jj} \text{ y así: } l_{ij} = \frac{a_{ij} - \sum_{k=1}^{j-1} l_{ik} u_{kj}}{u_{jj}}$$

**Caso 2:**  $i \leq j \rightarrow \min\{i, j\} = i$

$$a_{ij} = \sum_{k=1}^i l_{ik} u_{kj} = l_{i1} u_{1j} + l_{i2} u_{2j} + \cdots + l_{ii} u_{ij} \text{ y así: } u_{ij} = a_{ij} - \sum_{k=1}^{i-1} l_{ik} u_{kj} \text{ ya que } l_{ii} = 1$$

Veamos cómo funciona este algoritmo para

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}, \quad L = \begin{bmatrix} 1 & 0 & 0 \\ -\frac{1}{2} & 1 & 0 \\ 0 & -\frac{2}{3} & 1 \end{bmatrix}, \quad U = \begin{bmatrix} 2 & -1 & 0 \\ 0 & \frac{3}{2} & -1 \\ 0 & 0 & \frac{4}{3} \end{bmatrix}$$

Primera fila de  $U$ :

$$u_{11} = a_{11} = 2, \quad u_{12} = a_{12} = -1, \quad u_{13} = a_{13} = 0$$

Primera columna de  $L$ :

$$l_{11} = 1, \quad l_{21} = \frac{a_{21}}{u_{11}} = -\frac{1}{2}, \quad l_{31} = \frac{a_{31}}{u_{11}} = \frac{0}{2} = 0$$

Segunda fila de  $U$ :

$$u_{22} = a_{22} - \sum_{k=1}^1 l_{2k} u_{k2} = a_{22} - l_{21} u_{12} = 2 - \frac{1}{2}(-1) = \frac{3}{2}$$

$$u_{23} = a_{23} - \sum_{k=1}^1 l_{2k} u_{k3} = a_{23} - l_{21} u_{13} = -1 - \left(-\frac{1}{2}\right) \cdot 0 = -1$$

Segunda columna de  $L$ :

$$l_{32} = \frac{a_{32} - \sum_{k=1}^2 l_{3k} u_{k2}}{u_{22}} = \frac{-1 - (0)}{\frac{3}{2}} = -\frac{2}{3}$$

Tercera fila de  $U$

$$u_{33} = a_{33} - \sum_{k=1}^2 l_{2k}u_{k3} = 2 - (l_{31}u_{13} + l_{32}u_{23}) = 2 - \left(0 + 0 - \frac{2}{3} \cdot (-1)\right) = 2 - \frac{2}{3} = \frac{4}{3}$$

• Nota 1:

Se puede probar que el coste computacional de resolver un sistema lineal  $Ax = b$  mediante descomposición  $LU$  es  $\frac{n^3 + 3n^2 - 1}{3}$  siendo  $n$  el tamaño de la matriz  $A$ .

• Nota 2:

Realizando, si fuera necesario, permutaciones de filas (para llevar filas nulas al final) es fácil observar que toda matriz  $A \in M_{m \times n}(\mathbb{K})$  admite una factorización  $PLU$ , es decir,

$$PA = LU$$

donde:

- $P \in M_{m \times n}(\mathbb{K})$  es una matriz de permutación.
- $L \in M_{m \times n}(\mathbb{K})$  triangular inferior con 1s en la diagonal.
- $U \in M_{m \times n}(\mathbb{K})$  triangular superior.

Volviendo a la factorización:

$$\underbrace{\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}}_A = \underbrace{\begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ 0 & -\frac{2}{3} & 1 \end{bmatrix}}_L \cdot \underbrace{\begin{bmatrix} 2 & -1 & 0 \\ & \frac{3}{2} & -1 \\ & & \frac{4}{3} \end{bmatrix}}_U$$

se observa que el carácter simétrico de la matriz  $A$  se pierde en la factorización  $LU$ . Esta simetría se puede superar separando los pivotes (entradas en diagonal de  $U$ ) en una matriz diagonal  $D$  y dividiendo cada fila de  $U$  por su correspondiente pivote. Se obtiene así:

$$\underbrace{\begin{bmatrix} 2 & -1 & 0 \\ -1 & 2 & -1 \\ 0 & -1 & 2 \end{bmatrix}}_A = \underbrace{\begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ 0 & -\frac{2}{3} & 1 \end{bmatrix}}_L \cdot \underbrace{\begin{bmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{bmatrix}}_U \cdot \underbrace{\begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ & 1 & -\frac{2}{3} \\ & & 1 \end{bmatrix}}_{L^T}$$

En resumen:

$$A = LDL^T$$



factorización llamada Cholesky.

Además,  $D$  se puede factorizar como

$$D = \begin{bmatrix} 2 & & \\ & \frac{3}{2} & \\ & & \frac{4}{3} \end{bmatrix} = \underbrace{\begin{bmatrix} \sqrt{2} & & \\ & \sqrt{\frac{3}{2}} & \\ & & \frac{2}{\sqrt{3}} \end{bmatrix}}_{\sqrt{D}} \cdot \underbrace{\begin{bmatrix} \sqrt{2} & & \\ & \sqrt{\frac{3}{2}} & \\ & & \frac{2}{\sqrt{3}} \end{bmatrix}}_{\sqrt{D}}$$

se llega así:

$$\underbrace{\begin{bmatrix} 1 & & \\ -\frac{1}{2} & 1 & \\ 0 & -\frac{2}{3} & 1 \end{bmatrix}}_L \cdot \underbrace{\begin{bmatrix} \sqrt{2} & & \\ & \sqrt{\frac{3}{2}} & \\ & & \frac{2}{\sqrt{3}} \end{bmatrix}}_{\sqrt{D}} = \underbrace{\begin{bmatrix} \sqrt{2} & & \\ -\frac{\sqrt{2}}{2} & \sqrt{\frac{3}{22}} & \\ 0 & -\frac{2}{3}\sqrt{\frac{3}{2}} & \frac{2}{\sqrt{3}} \end{bmatrix}}_{\tilde{L}}$$

Columnas de  $L \times \sqrt{\text{pivotes}}$

$$\underbrace{\begin{bmatrix} \sqrt{2} & & \\ & \sqrt{\frac{3}{2}} & \\ & & \frac{2}{\sqrt{3}} \end{bmatrix}}_{\sqrt{D}} \cdot \underbrace{\begin{bmatrix} 1 & -\frac{1}{2} & 0 \\ & 1 & -\frac{2}{3} \\ & & 1 \end{bmatrix}}_{L^T} = \underbrace{\begin{bmatrix} \sqrt{2} & -\frac{\sqrt{2}}{2} & 0 \\ & \sqrt{\frac{3}{2}} & -\frac{2}{3}\sqrt{\frac{3}{2}} \\ & & \frac{2}{\sqrt{3}} \end{bmatrix}}_{\tilde{L}^T}$$

$\sqrt{\text{pivotes}} \times \text{filas de } L^T$

Con lo que obtenemos una nueva forma de la factorización Cholesky, a saber,  $A = \tilde{L} \cdot \tilde{L}^T$

- **Nota (Importante)**

Nótese que procedimiento ha influido porque los pivotes han salido positivos. Si hubieran salido negativos no lo hubiéramos podido hacer. Los pivotes son siempre positivos para una clase especial (pero que aparecen con muchísima frecuencia en Ingeniería) de matrices llamadas definidas positivas.

- **Definición**

Una matriz simétrica  $A \in M_n(\mathbb{R})$  se dice **definida positiva** si  $x^T A x > 0 \forall x \neq 0$ . Más adelante veremos una caracterización de este tipo de matrices.

- [Nota \(Coste computacional\)](#)

Para resolver un sistema  $Ax = b$ , con  $A$  simétrica y definida positiva mediante Cholesky hemos de resolver los sistemas triangulares

$$L_y = b, \quad L^T x = y$$

Para dar una respuesta satisfactoria, necesitamos introducir [normas de matrices](#)

Sea  $A \in M_{m \times n}(\mathbb{K})$ . Se define la norma de  $A$ , como  $\|A\| = \max_{x \neq 0} \frac{\|Ax\|}{\|x\|}$

Obviamente se tiene que

$$\frac{\|Ax\|}{\|x\|} \leq \|A\|$$

y así:  $\|Ax\| \leq \|A\| \cdot \|x\|$

La [llave](#) para la estabilidad de sistemas lineales está en el concepto de número de condicionamiento de una matriz.

- [Definición](#)

Dada  $A$  invertible, se define el número de condicionamiento como

$$\kappa(A) = \|A\| \cdot \|A^{-1}\|$$

El coste computacional total es del orden de  $\frac{n^3}{6}$ , justo la mitad que la factorización  $LU$ .

Por eso, para este tipo de matrices simétricas y definidas positivas se ha de usar Cholesky. Nótese que en la factorización  $LU$  hemos de calcular  $L$  y  $U$  mientras que en Choleski sólo  $L$ . Por eso el coste computacional es la mitad.

### [Análisis de errores](#)

Analizamos a continuación la llamada cuestión de la [estabilidad](#) en la resolución de un sistema lineal. La situación es como sigue:

- Para el sistema  $Ax = b$ , si  $b$  se cambia por  $b + \Delta b$ , debido por ejemplo a errores de representación o redondeo, entonces en lugar de  $x$  tendremos  $x + \Delta x$

- [¿Cómo podemos estimar o controlar  \$\Delta x\$  en función de  \$\Delta b\$ ?](#)

Esta es la pregunta de la [estabilidad](#).

- Nota

Como puede intuirse, no es fácil calcular el número de condicionamiento de una matriz. Dejamos para más adelante esta cuestión. La resolveremos vía valores propios y valores singulares al final de este curso.

Ya estamos en condiciones de dar respuesta a la cuestión anterior sobre control de error en la resolución de sistemas lineales.

Sea  $Ax = b$  un sistema compatible determinado. Si se perturba el término independiente de  $b$  por  $\Delta b$ , la nueva solución  $x + \Delta x$

$$A(x + \Delta x) = b + \Delta b$$

de modo que el error  $\Delta x$  es solución de

$$A(\Delta x) = \Delta b,$$

es decir,

$$\Delta x = A^{-1}(\Delta b)$$

y así

$$\|\Delta x\| = \|A^{-1}(\Delta b)\| \leq \|A^{-1}\| \cdot \|\Delta b\|,$$

lo que nos da una cota del [error absoluto](#)  $\Delta x$ .

Más importante que el error absoluto es el [relativo](#), el cual estimamos a continuación:

- Dado que  $Ax = b$ , se tiene que  $\|Ax\| = \|b\|$

Así:

$$\|\Delta x\| \leq \|A^{-1}\| \cdot \|\Delta b\| = \|A^{-1}\| \cdot \|\Delta b\| \cdot \overbrace{\frac{\|Ax\|}{\|b\|}}^1 \leq \|A^{-1}\| \cdot \|A\| \cdot \|x\| \cdot \frac{\|\Delta b\|}{\|b\|}$$

de donde

$$\frac{\|\Delta x\|}{\|x\|} \leq \|A\| \cdot \|A^{-1}\| \cdot \frac{\|\Delta b\|}{\|b\|} = c(A) \cdot \frac{\|\Delta b\|}{\|b\|}$$

Es decir, el [número de condicionamiento](#) controla la propagación del error relativo del término independiente a la solución.

- Nota:

Se puede probar que si el error se produce en  $A$ , (cosa habitual cuando  $A$  se factoriza en forma  $LU$  o Cholesky), entonces también el **número de condicionamiento** controla la propagación del error.

En resumen, si  $c(A)$  está mal condicionada y la solución  $x$  puede ser muy sensible con respecto a pequeños errores en la matriz  $A$  y/o en el término independiente  $b$ .

Por eso hemos de precisar **qué entendemos por un número de condicionamiento grande**. No hay un teorema matemático que de respuesta a esta pregunta, pero sí resultados de experimentos numéricos que sugieren la siguiente regla.

A la hora de resolver el sistema  $Ax = b$  mediante un método directo (Gauss,  $LU$ , Cholesky), el ordenador pierde el orden  $\log_{10}(c(A))$  cifras decimales en errores de redondeo.

Para finalizar, veamos un ejemplo concreto.

• **Ejemplo**

$$A = \begin{bmatrix} 1 & 7 \\ 0 & 1 \end{bmatrix}, \quad b = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$$

Consideramos el sistema  $Ax = b$

$$\begin{bmatrix} 1 & 7 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 7 \\ 1 \end{bmatrix}$$

$$\left. \begin{array}{l} x_1 + 7x_2 = 7 \\ x_2 = 1 \end{array} \right\} \boxed{x_1 = 0, \quad x_2 = 1}$$

Supongamos que  $b$  se perturba por  $\Delta b = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}$ .

El error en  $x$ , denotado  $\Delta x$ , es solución de  $A(\Delta x) = \Delta b$ ;

$$\begin{bmatrix} 1 & 7 \\ 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} \Delta x_1 \\ \Delta x_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0.1 \end{bmatrix}$$

$$\left. \begin{array}{l} \Delta x_1 + 7\Delta x_2 = 0 \\ \Delta x_2 = 0.1 \end{array} \right\} \longrightarrow \Delta x_1 = -7\Delta x_2 = -0.7$$

$$\boxed{\Delta x_1 = -0.7, \quad \Delta x_2 = 0.1}$$

El error relativo en  $x$  es:

$$\frac{\|\Delta x\|}{\|x\|} = \frac{\sqrt{(-0.7)^2 + 0.1^2}}{\sqrt{0^2 + 1^2}} = 0.1\sqrt{50}$$

El error absoluto en  $b$  es:

$$\frac{\|\Delta b\|}{\|b\|} = \frac{0^2 + 0.1^2}{\sqrt{7^2 + 1^2}} = \frac{0.1}{\sqrt{50}}$$

Por tanto,

$$\frac{\|\Delta x\|}{\|x\|} = 0.1\sqrt{50} = \sqrt{50} \cdot \sqrt{50} \frac{0.1}{\sqrt{50}} = 50 \frac{0.1}{\sqrt{50}} = \frac{\|\Delta b\|}{\|b\|}$$

es decir, el error de salida es **50 veces** mayor que en la entrada.

Si calculamos, con Python, el número de condicionamiento de  $A$ , nos devuelve  $c(A) = 51$ . Para la norma de Frobenius observamos pues que el número de condicionamiento de una estimación bastante precisa de la propagación del error

### Métodos iterativos para la resolución de sistemas lineales

- ¿Qué hacer si  $c(A) \gg 1$ ?

En lugar de resolver el sistema  $Ax = b$ , resolveremos el sistema equivalente

$$c(C^{-1} \cdot Ax) = C^{-1} \cdot b$$

de modo que la matriz invertible  $C$  haga que

$$c(C^{-1}A) \ll c(A)$$

- ¿Cómo elegir  $C$ ?

La primera opción sería tomar  $C = A$ . En este caso,  $C^{-1} \cdot A = A^{-1} \cdot A = I$ , cuyo número de condicionamiento es 1.

Sin embargo, calcular  $A^{-1}$  es igual de costoso que resolver  $Ax = b$ .

El intento de tomar  $C = A$  no es útil pero nos sugiere que la matriz  $C$  ha de ser **próxima** a  $A$ .

Otras dos opciones más prácticas son:

- 1) Si  $A$  no tiene ceros en la diagonal, tomamos  $C = \text{diag}(a_{11}, a_{12}, \dots, a_{nn})$ .

### Ejemplo

$$A = \begin{bmatrix} 8 & -2 \\ -2 & 50 \end{bmatrix} \quad \text{Python} \longrightarrow \kappa(A) = 6.3371$$

$$C = \begin{bmatrix} 8 & 0 \\ 0 & 50 \end{bmatrix} \quad C^{-1} = \begin{bmatrix} \frac{1}{8} & 0 \\ 0 & \frac{1}{50} \end{bmatrix}$$

$$\kappa(C^{-1} \cdot A) = 1.3370$$

2) Tomar  $C^{-1} = p(A)$ , un polinomio de  $A$  que resulta de [truncar](#) la serie de  $A^{-1}$ .

Es decir,

$$A^{-1} = I - (I - A)^{-1} = \frac{1}{I - (I - A)} = I + (I - A) + (I - A)^2 + \dots$$

lo cual converge si  $\|I - A\| < 1$

Por tanto,

$$C^{-1} = I + \sum_{k=1}^N (I - A)^k$$

Teniendo en mente la idea de usar una matriz próxima a  $A$  para reducir el número de condicionamiento de la matriz del nuevo sistema se introducen los [métodos iterativos](#) del siguiente modo:

- Dado que  $Ax = b$ , para toda matriz  $P$ , que a partir de ahora llamaremos preconditionador, se tiene que

$$Px = (P - A)x + b$$

Esta descomposición sugiere un algoritmo iterativo en el que a partir de  $x_k$  calculamos  $x_{k+1}$  siguiendo el algoritmo.

$$Px_{k+1} = (P - A)x_k + b$$

El éxito de este algoritmo está garantizado si se cumple:

- 1)  $x_{k+1}$  se calcula fácilmente a partir de  $x_k$ .
- 2) Los errores  $e_k = x - x_k$  se aproximan a cero, es decir, el algoritmo es convergente.

Veamos qué ha de suceder para que el algoritmo sea convergente:

$$\text{Tenemos:} \quad Px_{k+1} = (P - A)x_k + b$$

$$Px = (P - A)x + b$$

Si restamos:  $P(x - x_{k+1}) = (P - A)(x - x_k)$

$$Pe_{k+1} = (P - A) \cdot e_k$$

de donde suponiendo que  $P$  es invertible:

$$e_{k+1} = P^{-1}(P - A)e_k = \underbrace{(I - P^{-1} \cdot A)}_M e_k$$

Repitiendo el proceso y calculando normas:

$$\|e_{k+1}\| = \|M e_k\| = \|M\| \cdot \|e_k\| \leq \|M\| \cdot \|M\| \cdot \|e_{k-1}\| = \|M\|^2 \cdot \|e_{k-1}\| \leq \dots \leq \|M\| \cdot \|e_0\|$$

lo que significa que  $\|e_{k+1}\| \xrightarrow[k \rightarrow +\infty]{} 0$  si  $\|M\| < 1$ .

La pregunta del millón sigue siendo:

- ¿Cómo elegir  $P$ ?

Jacobi propuso tomar  $P = D$ , la parte diagonal de  $A$ . Esto conduce al llamado método de Jacobi. Esta opción es simple pero lenta.

En el lado opuesto, el método de Gauss-Seidel propone tomar como  $P$  una matriz más próxima a  $A$ . En concreto, partimos de la descomposición  $A = L + D + U$ .

$$\underbrace{\begin{bmatrix} a_{11} & \dots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \dots & a_{nn} \end{bmatrix}}_A = \underbrace{\begin{bmatrix} 0 & & 0 \\ a_{21} & & \\ \vdots & & \\ a_{n1} & \dots & a_{n,n-1} & 0 \end{bmatrix}}_L + \underbrace{\begin{bmatrix} a_{11} & & 0 \\ & \ddots & \\ 0 & & a_{nn} \end{bmatrix}}_D + \underbrace{\begin{bmatrix} 0 & a_{12} & \dots & a_{1n} \\ & 0 & & \\ & & \ddots & \\ 0 & & & 0 \end{bmatrix}}_U$$

En el método de Gauss-Seidel se toma  $P = D + L$  de modo que el algoritmo

$$Px_{k+1} = -Ux_k + b, \quad P = D + L$$

Ejemplo: Aplicado al caso de la matriz  $A = k_4$

$$k_4 = \begin{bmatrix} 2 & -1 & 0 & 0 \\ -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 \\ 0 & 0 & -1 & 2 \end{bmatrix}, \quad P = L + D = \begin{bmatrix} 2 & 0 & 0 & 0 \\ -1 & 2 & 0 & 0 \\ 0 & -1 & 2 & 0 \\ 0 & 0 & -1 & 2 \end{bmatrix}, \quad U = \begin{bmatrix} 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

El algoritmo queda:

$$-1 \cdot \begin{bmatrix} 0 \\ x_1 \\ x_2 \\ x_3 \end{bmatrix}_{\text{new}} + 2 \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}_{\text{new}} = \begin{bmatrix} x_2 \\ x_3 \\ x_4 \end{bmatrix}_{\text{old}} + \begin{bmatrix} b_1 \\ b_2 \\ b_3 \\ b_4 \end{bmatrix}$$

- ¿Cuándo tenemos garantizada la convergencia en este método?

Para el caso de las llamadas **matrices diagonal dominantes**.

- **Definición**

Una matriz  $A = (a_{ij})$  se dice estrictamente diagonal dominante si para cada fila  $i$  se cumple

$$|a_{ii}| > \sum_{\substack{j=1 \\ i \neq j}}^n |a_{ij}|$$

Ejemplo

$$A = \begin{bmatrix} 2 & -1 & 0 \\ -1 & 4 & 2 \\ 0 & 2 & 3 \end{bmatrix} \quad \begin{array}{l} |a_{11}| = 2 > 1 \\ |a_{22}| = 4 > 2 + 1 = 3 \\ |a_{33}| = 3 > 2 \end{array}$$

- **Teorema**

Supongamos que  $A$  es estrictamente diagonal dominante. Entonces, los métodos de Jacobi y Gauss-Seidel son convergentes.



## 4) Subespacios Vectoriales

### Subespacios vectoriales

#### • Definición

Un subespacio (vectorial) de  $\mathbb{K}^n$  es un subconjunto de  $W$  de  $\mathbb{K}^n$  que verifica:

- 1) Si  $w, w' \in W \longrightarrow w + w' \in W$ .
- 2) Si  $w \in W$  y  $\alpha \in \mathbb{K} \longrightarrow \alpha w \in W$ .

Nota: Nótese que el vector 0 siempre pertenece a todo subespacio vectorial  $W$ . En efecto: dado  $w \in W$ , se tiene que  $-1 \cdot w = -w \in W$ . Además  $w + (-w) = 0 \in W$ .

#### Ejemplo 1

Sean  $A \in M_{m \times n}(\mathbb{K})$  y  $W$  el conjunto de soluciones del sistema  $Ax = 0$ . Entonces,  $W$  es un subespacio vectorial. En efecto: Si  $w, w' \in W$ , entonces  $Aw = 0$  y  $Aw' = 0$ . Por tanto,  $A(w + w') = Aw + Aw' = 0 + 0 = 0$ , lo que implica que  $w + w' \in W$ . El subespacio  $W$  anterior se llama núcleo de  $A$  y se denota  $\text{Nuc}(A)$ .

#### Ejemplo 2

Sea  $\{u_1, u_2, \dots, u_m\}$  un conjunto de vectores de  $\mathbb{K}^n$ .

Denotemos por  $\langle u_1, u_2, \dots, u_m \rangle$  el conjunto formado por todas las combinaciones lineales de  $\{u_1, u_2, \dots, u_m\}$ , es decir

$$\langle u_1, u_2, \dots, u_m \rangle = \left\{ w = \sum_{i=1}^m \alpha_i u_i, \alpha_i \in \mathbb{K} \right\}$$

Es fácil comprobar que  $\langle u_1, u_2, \dots, u_m \rangle$  es un subespacio vectorial de  $\mathbb{K}^n$ . Este subespacio se llama **subespacio generado** por  $\{u_1, u_2, \dots, u_m\}$ , y al conjunto de vectores,  $\{u_1, u_2, \dots, u_m\}$  se le llama **conjunto generador**.

#### • Definición (Base de un subespacio vectorial)

Sean  $W$  un subconjunto vectorial de  $\mathbb{K}^n$  y  $B = \{u_1, u_2, \dots, u_m\}$  un conjunto de vectores. Se dice que  $B$  es una base de  $W$  si se cumple:

- 1)  $W = \langle u_1, u_2, \dots, u_m \rangle$
- 2)  $\{u_1, u_2, \dots, u_m\}$  son linealmente independientes.

### Ejemplo 1

Sea  $W = \mathbb{K}^n$ . Entonces los vectores

$$B = \{u_1 = (1, 0, \dots, 0), u_2 = (0, 1, 0, \dots, 0), \dots, u_n = (0, \dots, 0, 1)\}$$

Son una base de  $W$ . En efecto:

1)  $B$  es un conjunto generador ya que si  $v = (v_1, \dots, v_n) \in \mathbb{K}^n$ , entonces

$$v = \sum_{i=1}^n v_i u_i$$

2)  $B$  es literalmente independiente:

$$\sum \alpha_i u_i = 0 \longleftrightarrow (\alpha_1, \alpha_2, \dots, \alpha_n) = (0, \dots, 0) \longrightarrow \alpha_i = 0 \quad \forall 1 \leq i \leq n$$

### Ejemplo 2 Sean

$$A = \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

y  $W = \text{nuc}(A)$ , es decir,  $W$  está formado por las soluciones del sistema

$$Ax = 0 \longleftrightarrow \begin{bmatrix} 1 & -1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = [0, 0] \quad \left. \begin{array}{l} x_1 - x_2 + x_3 = 0 \\ x_1 + x_2 + x_3 = 0 \end{array} \right\}$$

Como  $\text{rango}(A) = 2$  se trata de un sistema compatible indeterminado. La solución depende de un parámetro. Tomamos  $x_3 = \lambda$ :

$$\left. \begin{array}{l} x_1 - x_2 = -\lambda \\ x_1 + x_2 = -\lambda \end{array} \right\} \longrightarrow \begin{array}{l} 2x_1 = -2\lambda \\ x_1 = -\lambda \\ x_2 = -\lambda - x_1 = -\lambda - (-\lambda) = 0 \end{array}$$

$$\left. \begin{array}{l} x_1 = -\lambda \\ x_2 = 0 \\ x_3 = \lambda \end{array} \right\} (x_1, x_2, x_3) = \lambda(-1, 0, 1)$$

$B = \{(-1, 0, 1)\}$  es una base de  $W$ .

### • Teorema (de la base)

Sea  $W$  un subespacio vectorial de  $\mathbb{K}^n$ . Entonces todas las bases de  $W$  tienen el mismo número de elementos.

A este número se le llama dimensión de  $W$  y se denota  $\dim(W)$ .

## Subespacios vectoriales y matrices

### • Subespacios asociados a una matriz. Rango

Dada  $A \in M_{m \times n}(\mathbb{K})$ , se definen los siguientes subespacios asociados a  $A$ :

- 1)  $\text{Col}(A)$  = subespacio generados por las columnas de  $A$ .
- 2)  $\text{Fil}(A)$  = subespacio generados por las filas de  $A$ .
- 3)  $\text{Nuc}(A)$  = subespacio generados por las soluciones de  $Ax = 0$ .

—→ Núcleo por la izquierda: subespacio generados por las soluciones de  $x^T A = 0$ .

- Recordemos que hemos definido el rango de una matriz como el número de pivotes que resultan de escalar una matriz, es decir, el número de filas no nulas que resultan de escalar una matriz. Nuestro objetivo es ver que  $r(A) = \dim(\text{Col}(A)) = \dim(\text{Fil}(A))$ .

### • Proposición

Sean  $A$ ,  $P$ ,  $Q$  matrices de tamaños adecuados para que tengan sentido los productos correspondientes. Supongamos que  $P$  y  $Q$  son invertibles. Entonces:

- 1)  $\dim \text{Col}(PA) = \dim \text{Col}(A)$
- 2)  $\dim \text{Fil}(PA) = \dim \text{Fil}(A)$
- 3)  $\text{rango}(A) = \dim \text{Fil}(A) = \dim \text{Col}(A)$

### • Demostración

- 1) Sean  $r = \dim \text{Col}(A)$  y  $\{u_1, u_2, \dots, u_r\}$  una base de  $\text{Col}(A)$ . Entonces, es evidente que al ser  $P$  invertible,  $\{Pu_1, Pu_2, \dots, Pu_r\}$  es una base de  $\text{Col}(A)$ , entonces  $\{P^{-1}Pu_1, P^{-1}Pu_2, \dots, P^{-1}Pu_r\}$  son columnas linealmente independientes de  $A$ . En consecuencia,  $\dim \text{Col}(PA) \leq r$ .

Por tanto,  $\dim \text{Col}(A) = \dim \text{Col}(PA)$

- 2) Análogo

- 3) Veamos que dadas dos matrices  $A$  y  $B$ , de modo que  $B$  es invertible, entonces  $\text{col}(A, B) = \text{col}(A)$ . En

efecto: las columnas de  $AB$  son combinación lineal de las columnas de  $A$ . Veámoslo:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} \cdot \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & & \vdots \\ b_{n1} & \cdots & b_{nn} \end{bmatrix} = b_{11} \cdot \begin{bmatrix} a_{11} \\ \vdots \\ a_{m1} \end{bmatrix} + b_{12} \cdot \begin{bmatrix} a_{12} \\ a_{22} \\ \vdots \\ a_{m2} \end{bmatrix} + \cdots$$

Por tanto,  $\text{Col}(AB) \subset \text{Col}(A)$ .

De igual forma, como  $B$  es invertible,

$$\text{Col}((AB) \cdot B^{-1}) = \text{col}(A) \subset \text{Col}((AB) \cdot B^{-1}B) = \text{col}(AB)$$

Ya estamos en disposición de probar (3).

Consideremos la factorización  $PAQ$  de  $A$ , es decir,

$$\left( \begin{array}{c|c} Ir & 0 \\ \hline 0 & 0 \end{array} \right) = PAQ$$

Entonces, por lo visto anteriormente:

$$\dim \text{Col}(A) = \dim \text{Col}(PA) = \dim \text{Col}(PAQ) = r$$

$$\dim \text{Fil}(A) = \dim \text{Fil}(PA) = \dim \text{Fil}(PAQ) = r$$

### • Teorema (Rank Nulity)

Sea  $A \in M_{m \times n}(\mathbb{K})$ . Entonces:

$$r(A) + \dim \text{Nuc}(A) = n$$

### Demostración

Consideremos el sistema  $Ax = 0$  y sea  $r = r(A)$ . Entonces, por el Teorema de Rouché-Frobenius, el sistema anterior es compatible indeterminado (si  $r(A) < m$ ) y la solución depende de

$$n^0 \text{ de incógnitas} - r(A) \text{ parámetros}$$

Como  $n^0$  de incógnitas es  $n$ , y el número de parámetros de la dimensión de  $\text{Nuc}(A)$  el resultado se tiene ya que

$$n - r(A) = \dim \text{Nuc}(A)$$

## 2) Cálculo del rango de una matriz con determinantes

Sea  $M_{m \times n}(\mathbb{K})$ . Se llaman **menores** de  $A$  a los determinantes de las submatrices cuadradas de  $A$ . Se llama **orden** de un menor al número de sus filas (o columnas).

- Sea  $A \in M_{m \times n}(\mathbb{K})$ . El rango de  $A$  coincide con el mayor de los órdenes de los menores no nulos de  $A$ .

### Ejemplos

$$A = \begin{bmatrix} 2 & 2 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 1 & 1 & 1 & 1 \end{bmatrix} \quad \begin{vmatrix} 2 & 2 \\ 1 & 1 \end{vmatrix} = 2 - 2 = 0, \quad \begin{vmatrix} 2 & 1 \\ 1 & 0 \end{vmatrix} = -1 \neq 0$$
$$\begin{vmatrix} 2 & 2 & 1 \\ 1 & 1 & 0 \\ 1 & 1 & 1 \end{vmatrix} = 0, \quad \begin{vmatrix} 2 & 1 & 1 \\ 1 & 0 & 0 \\ 1 & 1 & 1 \end{vmatrix} = 0$$

$$r(A) = 2$$

## 2) Coordenadas respecto de una base

Sean  $W \subset \mathbb{K}^n$  un subespacio vectorial y  $B = \{u_1, u_2, \dots, u_m\}$  una base de  $W$ . Entonces todo vector de  $W$  se expresa **de modo único** como combinación lineal de los vectores de  $B$ . En efecto: sea  $w \in W$  y supongamos que

$$w = \sum_{i=1}^m \alpha_i u_i = \sum_{i=1}^m \beta_i u_i$$

Entonces

$$\sum_{i=1}^m (\alpha_i - \beta_i) u_i = 0$$

y como los vectores  $\{u_1, u_2, \dots, u_m\}$  son linealmente independientes, entonces  $\alpha_i - \beta_i = 0$ ,  $1 \leq i \leq m$ , lo que implica  $\alpha_i = \beta_i$ ,  $1 \leq i \leq m$ .

### • Definición

Sean  $W$  un subespacio vectorial de  $\mathbb{K}^n$  y  $B = \{u_1, u_2, \dots, u_m\}$  una base de  $W$ . Dado  $w \in W$ , se llaman **coordenadas** de  $w$  en la base  $B$  a los únicos escalares  $(x_1, \dots, x_m) \in \mathbb{K}^n$  tales que  $w = x_1 u_1 + \dots + x_m u_m$ . Se denota  $[w]_B = (x_1, \dots, x_m)$

### Matriz de cambio de base

Sean  $B = \{u_1, u_2, \dots, u_m\}$  y  $B' = \{u'_1, u'_2, \dots, u'_m\}$  dos bases de  $W \subset \mathbb{K}^n$ .

Dado  $w \in W$ , se tiene

$$w = \sum_{i=1}^m x_i v_i = \sum_{i=1}^m x'_i v'_i$$

Vamos a encontrar la relación que existe en  $[w]_B = (x_i)$  y  $(x'_i) = [w]_{B'}$ .

Como  $v_i \in W$  se tiene:

$$\begin{array}{l} v_1 = a_{11}v'_1 + a_{21}v'_2 + \dots + a_{m1}v'_m \\ v_2 = a_{12}v'_1 + a_{22}v'_2 + \dots + a_{m2}v'_m \\ \hline v_m = a_{1m}v'_1 + a_{2m}v'_2 + \dots + a_{mm}v'_m \end{array}$$

Llamamos matriz de cambio de base de  $B$  a  $B'$  a la matriz

$$M_{B \rightarrow B'} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & & a_{mm} \end{bmatrix}$$

Se tiene:

$$\begin{aligned} w = \sum_{i=1}^m x_i v_i &= x_1 v_1 + x_2 v_2 + \dots + x_m v_m \\ &= x_1 (a_{11}v'_1 + a_{21}v'_2 + \dots + a_{m1}v'_m) \\ &\quad + x_2 (a_{12}v'_1 + a_{22}v'_2 + \dots + a_{m2}v'_m) \\ &\quad + \dots \\ &\quad + x_m (a_{1m}v'_1 + a_{2m}v'_2 + \dots + a_{mm}v'_m) \\ &= (a_{11}x_1 + a_{21}x_2 + \dots + a_{m1}x_m)v'_1 \\ &\quad + (a_{12}x_1 + a_{22}x_2 + \dots + a_{m2}x_m)v'_2 \\ &\quad + \dots \\ &\quad + (a_{1m}x_1 + a_{2m}x_2 + \dots + a_{mm}x_m)v'_m \end{aligned}$$

Por tanto,

$$\begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_m \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1m} \\ a_{21} & a_{22} & \dots & a_{2m} \\ \hline a_{m1} & a_{m2} & \dots & a_{mm} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_m \end{bmatrix}$$

es decir,

$$[w]_{B'} = M_{B \rightarrow B'} [w]_B$$

Ejemplo

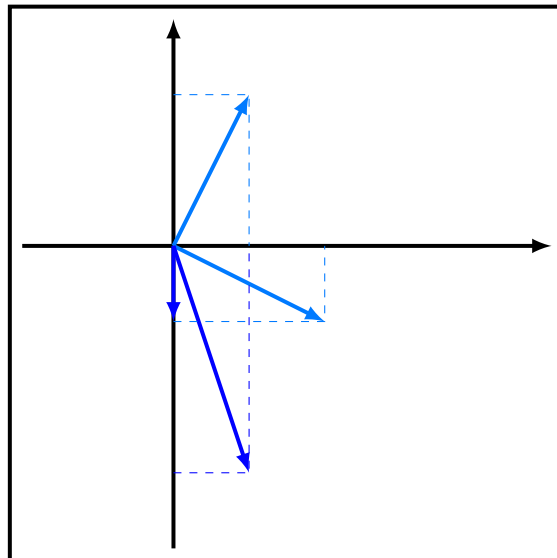
$$W = \mathbb{R}^2$$

$$B = \{v_1 = (1, 2), v_2 = (2, -1)\}$$

$$B' = \{v'_1 = (1, -3), v'_2 = (0, -1)\}$$

Vamos a calcular  $M_{B \rightarrow B'}$ .

$$\begin{aligned} v_1 = (1, 2) &= a_{11}v'_1 + a_{21}v'_2 \\ &= a_{11}(1, -3) + a_{21}(0, -1) \\ &= (a_{11}, -3a_{11} - a_{21}) \end{aligned}$$



$$\left. \begin{aligned} 1 &= a_{11} \\ 2 &= -3a_{11} - a_{21} \end{aligned} \right\} \longrightarrow a_{22} = -3a_{12} + 1 = -6 + 1 = -5$$

Por tanto,

$$M_{B \rightarrow B'} = \begin{bmatrix} 1 & 2 \\ -5 & -5 \end{bmatrix}$$

Sea ahora  $[v]_B = (1, 1)$ . ¿ $[v]_{B'}$ ?

$$[v_{B'}] = M_{B \rightarrow B'} [v]_B = \begin{bmatrix} 1 & 2 \\ -5 & -5 \end{bmatrix} \cdot \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 3 & -10 \end{bmatrix}$$

[Coordenadas en bases ortonormales](#)

Supongamos que  $B = \{u_1, u_2, \dots, u_m\}$  es una base ortonormal de  $W$ , es decir,

$$u_i \cdot u_j = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases} \quad \text{delta de Kronecker}$$

$$\|u_i\|, \quad 1 \leq i \leq m$$

Vamos a calcular la forma que tiene  $[w]_B$ , siendo  $w \in W$ . Se tiene:

$$w = \sum_{n=1}^{+\infty} x_n u_n = x_j$$

Por tanto,

$$[w]_B = (w \cdot u_j), \quad 1 \leq j \leq m$$

$$\begin{aligned} \text{Además, como } w \cdot u_j &= \|w\| \cdot \|u_j\| \cos(w, u_j) \\ &= \|w\| \cdot \cos(w, u_j) \end{aligned}$$

Se tiene que si  $\|w\| = 1$ , entonces

$$[w]_B = (w \cdot u_j) = \underbrace{(\cos(w, u_j))}_{\substack{\text{cosenos directores} \\ \text{del vector } w}}$$

Trabajar en bases ortonormales simplifica enormemente muchos de los cálculos que se hacen en Álgebra Lineal. Por ello, dada una base de un subespacio vectorial, interesa calcular, a partir de ella una nueva base ortonormal. El algoritmo de Gram-Schmidt nos permite conseguir este objetivo.

### Método de ortogonalización de Gram-Schmidt y su versión matricial, factorización $A = QR$

#### • Teorema (Factorización $QR$ )

Sean  $\{a_1, a_2, \dots, a_n\}$  un conjunto de vectores de  $\mathbb{R}^m$  linealmente independientes. Entonces, existe un conjunto ortonormal  $\{q_1, q_2, \dots, q_n\}$ , entonces existe una matriz  $R$  triangular superior con elementos positivos en la diagonal de modo que

$$A = QR$$

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & \cdots & a_{mn} \end{bmatrix}_{m \times n} = \begin{bmatrix} q_{11} & q_{12} & \cdots & q_{1n} \\ \vdots & \vdots & & \vdots \\ q_{m1} & q_{m2} & \cdots & q_{mn} \end{bmatrix}_{m \times n} \cdot \begin{bmatrix} r_{11} & \cdots & r_{1n} \\ & \ddots & \\ & & r_{nn} \end{bmatrix}_{n \times n}$$

$\begin{matrix} \uparrow & \uparrow & & \uparrow \\ a_1 & a_2 & & a_n \end{matrix} \quad \begin{matrix} \uparrow & \uparrow & & \uparrow \\ q_1 & q_2 & & q_n \end{matrix}$

#### • Demostración

El algoritmo, el cual es constructivo, consta de las siguientes etapas o pasos:

Paso 1: Tomamos  $q = 1 = \frac{a_1}{\|a_1\|}$

Con lo que  $a_1 = \|a_1\|q_1$ . Tenemos entonces  $r_{11} = \|a_{11}\|$

Paso 2: Tomamos  $v_2 = a_2 + \alpha q_1$



Imponemos  $q_1 \cdot v_2 = 0$

$$0 = q_1 \cdot v_2 = q_1 \cdot a_2 + \alpha q_1 \cdot q_1 = q_1 \cdot a_2 + \alpha \longrightarrow \alpha = -q_1 \cdot a_2$$

Por tanto:  $v_2 = a_2 - (q_1 \cdot a_2)q_1$

$$\begin{aligned} \text{Calculamos } \|v_2\|^2 &= (a_2 - (q_1 \cdot a_2)q_1) \cdot (a_2 - (q_1 \cdot a_2)q_1) \\ &= a_2 \cdot a_2 - (q_1 \cdot a_2)^2 - (q_1 \cdot a_2)^2 + (q_1 \cdot a_2)^2 \\ &= a_2 \cdot a_2 - (q_1 \cdot a_2)^2 \end{aligned}$$

Con lo que  $\|v_2\| = (\|a_2\|^2 - (q_1 \cdot a_2)^2)^{\frac{1}{2}}$

$$\text{Tomamos entonces } q_2 = \frac{v_2}{\|v_2\|} = \frac{1}{(\|a_2\|^2 - (q_1 \cdot a_2)^2)^{\frac{1}{2}}} \cdot (a_2 - (q_1 \cdot a_2)q_1)$$

$$r_{12} = -q_1 \cdot a_2, \quad r_{22} = (\|a_2\|^2 - (q_1 \cdot a_2)^2)^{\frac{1}{2}} \text{ ya que } a_2 = \underbrace{q_1 \cdot a_2}_{r_{12}} q_1 + \underbrace{(\|a_2\|^2 - (q_1 \cdot a_2)^2)^{\frac{1}{2}}}_{r_{22}} \cdot q_{12}$$

Paso 3: Tomamos  $v_3 = a_3 + \alpha q_1 + \beta q_{12}$

Imponemos:

$$\text{a) } 0 = v_3 \cdot q_1 = a_3 \cdot q_1 + \alpha \overbrace{q_1 \cdot q_1}^1 + \beta \overbrace{q_2 \cdot q_1}^0 \longrightarrow \alpha = -a_3 q_1$$

$$\text{b) } 0 = v_3 \cdot q_2 = a_3 \cdot q_2 + \alpha \underbrace{q_1 \cdot q_2}_0 + \beta \underbrace{q_1 \cdot q_1}_1 \longrightarrow \beta = -a_3 q_2$$

Finalmente, calculamos

$$\begin{aligned} \|v_3\|^2 &= (a_3 + \alpha q_1 + \beta q_2) \cdot (a_3 + \alpha q_1 + \beta q_2) \\ &= \|a_3\|^2 + \overline{\alpha q_1} q_3 + \beta q_2 \cdot a_3 + \overline{\alpha q_1} q_3 + \alpha \overbrace{q_1 \cdot q_1}^1 + \\ &\quad \alpha \beta q_1 q_2 + \beta q_2 a_3 + \alpha \beta q_1 q_2 + \beta \underbrace{q_2 \cdot q_2}_1 \\ &= \|a_3\|^2 + 2\alpha \beta q_1 q_3 + 2\beta q_2 a_3 + \alpha + \beta \end{aligned}$$

$$\text{Tomamos: } q_3 = \frac{v_3}{\|v_3\|} = \frac{1}{\|v_3\|} (a_3 - (a_3 \cdot q_1)q_1 - (a_3 \cdot q_2)q_2)$$

Con lo que:

$$a_3 = \underbrace{(a_3 \cdot q_1)}_{r_{13}} q_1 + \underbrace{(a_3 \cdot q_2)}_{r_{23}} q_2 + \underbrace{\|v_3\|}_{r_{33}} q_3$$

y así sucesivamente.

Ejemplo

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{3}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{3}} \\ 0 & \frac{2}{\sqrt{6}} & \frac{1}{\sqrt{3}} \end{bmatrix} \cdot \begin{pmatrix} \sqrt{2} & \frac{1}{\sqrt{2}} & \frac{1}{2} \\ 0 & \frac{\sqrt{2}}{3} & \frac{1}{\sqrt{6}} \\ 0 & 0 & \frac{2\sqrt{3}}{\sqrt{3}} \end{pmatrix}$$

$\begin{matrix} \uparrow & \uparrow & \uparrow \\ a_1 & a_2 & a_3 \end{matrix}$

$$1) \quad q_1 = \frac{a_1}{\|a_1\|}, \quad \|a_1\| = \sqrt{1^2 + 1^2} = \sqrt{2}$$

$$q_1 = \frac{1}{\sqrt{2}}a_1 \longrightarrow a_1 = \underbrace{\sqrt{2}}_{r_{11}} q_{11}$$

$$2) \quad v_2 = a_2 + \alpha q_1$$

$$\begin{aligned} 0 &= v_2 \cdot q_1 = a_2 \cdot q_1 + \alpha q_1 \cdot q_1 \longrightarrow \alpha = -a_2 \cdot q_1 = -(1, 0, 0) \cdot \frac{1}{\sqrt{2}}(1, 1, 0) \\ &= a_2 - \frac{1}{\sqrt{2}}q_1 = (1, 0, 1) - \left(\frac{1}{2}, \frac{1}{2}, 0\right) = \left(\frac{1}{2}, -\frac{1}{2}, 1\right) \end{aligned}$$

$$\|v_2\| = \sqrt{\frac{1}{4} + \frac{1}{4} + 1} = \sqrt{\frac{3}{2}} = \frac{3}{\sqrt{6}}$$

$$q_2 = \frac{v_2}{\|v_2\|} = \frac{\sqrt{2}}{\sqrt{3}} \left(\frac{1}{2}, \frac{1}{2}, 1\right) = \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right)$$

$$a_2 = \frac{1}{\sqrt{2}}q_1 + v_2 = \underbrace{\frac{1}{\sqrt{2}}}_{r_{12}} q_1 + \underbrace{\|v_2\|}_{r_{22}} q_2$$

$$3) \quad V_3 = a_3 + \alpha q_1 + \beta q_2$$

$$0 = q_1 \cdot v_3 \longrightarrow \alpha = a_3 \cdot q_1 = -(0, 1, 1) \cdot \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right) = -\frac{1}{\sqrt{2}}$$

$$0 = q_2 \cdot v_3 \longrightarrow \beta = -a_3 \cdot q_2 = -(0, 1, 1) \cdot \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right) = \frac{1}{\sqrt{6}} - \frac{2}{\sqrt{6}} = -\frac{1}{\sqrt{6}}$$

$$v_3 = (0, 1, 1) - \frac{1}{\sqrt{2}} \left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}, 0\right) - \frac{1}{\sqrt{6}} \left(\frac{1}{\sqrt{6}}, -\frac{1}{\sqrt{6}}, \frac{2}{\sqrt{6}}\right) =$$

$$(0, 1, 1) - \left(\frac{1}{2}, \frac{1}{2}, 0\right) - \left(\frac{1}{6}, -\frac{1}{6}, \frac{2}{6}\right) = \left(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right)$$

$$\|v_3\| = \sqrt{\frac{4}{9} + \frac{4}{9} + \frac{4}{9}} \sqrt{\frac{12}{9}} = \frac{2}{3}\sqrt{3}$$

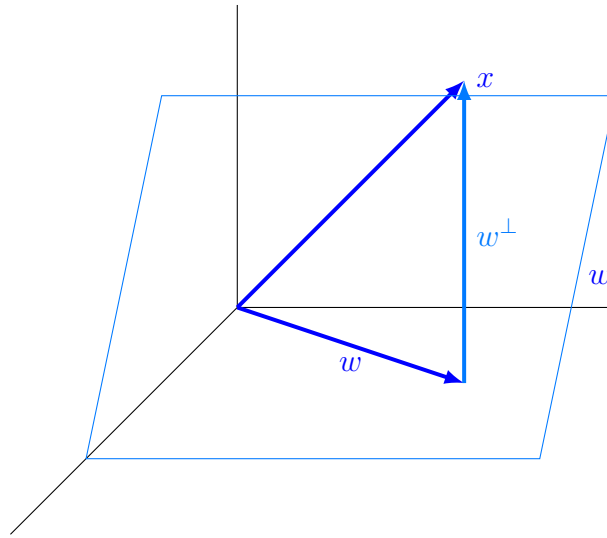
$$q_3 = \frac{v_3}{\|v_3\|} = \frac{2\sqrt{3}}{3} \left(-\frac{2}{3}, \frac{2}{3}, \frac{2}{3}\right) = \left(-\frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}, \frac{1}{\sqrt{3}}\right)$$

$$a_3 = -\alpha q_1 - \beta q_2 + \|v_3\| q_3 = \underbrace{\frac{1}{\sqrt{2}}}_{r_{13}} q_1 + \underbrace{\frac{1}{\sqrt{6}}}_{r_{23}} q_2 + \underbrace{\frac{2\sqrt{3}}{3}}_{r_{33}} q_3$$

Nota: Si  $A$  es una matriz cuadrada, entonces la condición  $\det(A) \neq 0$  garantiza que las columnas de  $A$  son linealmente independientes y por tanto, la existencia de la factorización  $A = QR$  está garantizado.

### Proyección ortogonal

Sean  $W \subset \mathbb{R}^n$  un subespacio y  $x \in \mathbb{R}^n$ .



¿Cuál es el vector de  $W$  que está más cerca de  $x$ ?

Es decir,  $\min_{w \in W} \|w - x\|$

Dado  $x \in \mathbb{R}^n$ , se puede ver que  $x = w + w^\perp$  con  $w \in W$  y  $w^\perp \in W^\perp$ , subespacio al que llamaremos ortogonal de  $W$ . En concreto:

$$w^\perp : \{ \overline{w} \in \mathbb{R}^n : \overline{w} \cdot w = 0 \quad \forall w \in W \}$$

Es decir,  $\mathbb{R}^n = W \oplus W^\perp$ , donde  $\oplus$  significa:

i)  $\mathbb{R}^n = W + W^\perp$

ii)  $W \cap W^\perp = \{0\}$

y se habla de [suma discreta](#).

Al vector  $w$  en la descomposición  $x = w + w^\perp$  se le llama [proyección ortogonal de  \$x\$  en  \$W\$](#) .

Escribimos  $w = \mathbb{P}_w(x)$ .

Veamos cómo se calcula. Sea  $\{u_1, \dots, u_m\}$  una base ortonormal de  $W$ , y  $\{u_{m+1}, \dots, u_n\}$  una base ortonormal de  $W^\perp$ . Resulta entonces que

$$\{u_1, u_2, \dots, u_m, u_{m+1}, \dots, u_n\}$$

es una base ortonormal de  $\mathbb{R}^n$ . Por tanto

$$x = w + w^\perp = \sum_{i=1}^m x_i u_i = \sum_{i=1}^m x_i u_i + \sum_{i=m+1}^n x_i u_i$$

Se tiene entonces que

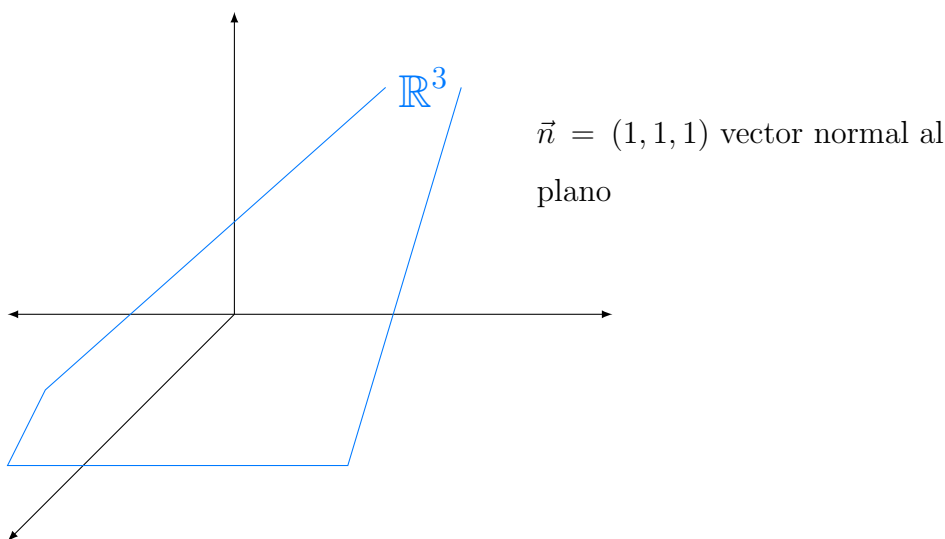
$$x \cdot u_j = x_j$$

Por tanto,  $w = \sum_{i=1}^m x_i u_i$  tiene coordenadas

$$x_i = x \cdot u_i$$

### Ejemplo

Calcula la proyección ortogonal en  $\mathbb{R}^3$  del vector  $v = (0, -1, 0)$  sobre el plano de ecuación  $x + y + z = 0$ .



$$W = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 0\}$$

Tomamos dos parámetros

$$\begin{aligned} z &= \alpha \\ y &= \beta \\ x &= -z - y = -\alpha - \beta \end{aligned} \quad \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \alpha \cdot \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} + \beta \cdot \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$$

$\uparrow \quad \quad \uparrow$   
 $v_1 \quad \quad v_2$

$$\|v_1\| = \sqrt{2}; \quad \|v_2\| = \sqrt{2}$$

Construimos una base ortonormal de  $W$  usando el algoritmo de Gram-Schmidt.

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

$$v_2 = (-1, 0, 1) + \alpha \frac{1}{\sqrt{2}}(-1, 1, 0)$$

$$0 = v_2 \cdot u_1 = (-1, 0, 1) \cdot \frac{1}{\sqrt{2}}(-1, 1, 0) + \alpha$$

$$\alpha = -\frac{1}{\sqrt{2}}(-1, 0, 1) \cdot (-1, 1, 0) = -\frac{1}{\sqrt{2}}$$

$$v_2 = (-1, 0, 1) - \frac{1}{\sqrt{2}} \cdot \frac{1}{\sqrt{2}}(-1, 1, 0) = (-1, 0, 1) + \left(\frac{1}{2}, -\frac{1}{2}, 0\right) = \left(-\frac{1}{2}, -\frac{1}{2}, 1\right)$$

$$\|v_2\| = \sqrt{\frac{1}{4} + \frac{1}{4} + 1} = \sqrt{\frac{3}{2}}$$

$$\text{Finalmente, } u_2 = \frac{\sqrt{3}}{\sqrt{2}} \left(-\frac{1}{2}, -\frac{1}{2}, 1\right)$$

$$\mathbb{P}_w(u) = (v \cdot u_1)u_1 + (v \cdot u_2)u_2$$

### Transformaciones (Aplicaciones) lineales

#### • Definición

Una transformación o aplicación lineal es una aplicación

$$f : \mathbb{K}^n \longrightarrow \mathbb{K}^n$$

que cumple las dos siguientes propiedades:

$$1) \quad f(u + v) = f(u) + f(v) \quad \forall u, v \in \mathbb{K}^n$$

$$2) \quad f(\alpha u) = \alpha f(u)$$

#### Ejemplo

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

$$(x, y) \longmapsto f(x, y) = (x \cos \theta - y \sin \theta, x \sin \theta + y \cos \theta)$$

$\theta$  es un ángulo fijo.

Veamos que  $f$  cumple las dos propiedades anteriores:

$$1) \quad u = (x_1, y_1), v = (x_2, y_2)$$

$$u + v = (x_1 + x_2, y_1 + y_2)$$

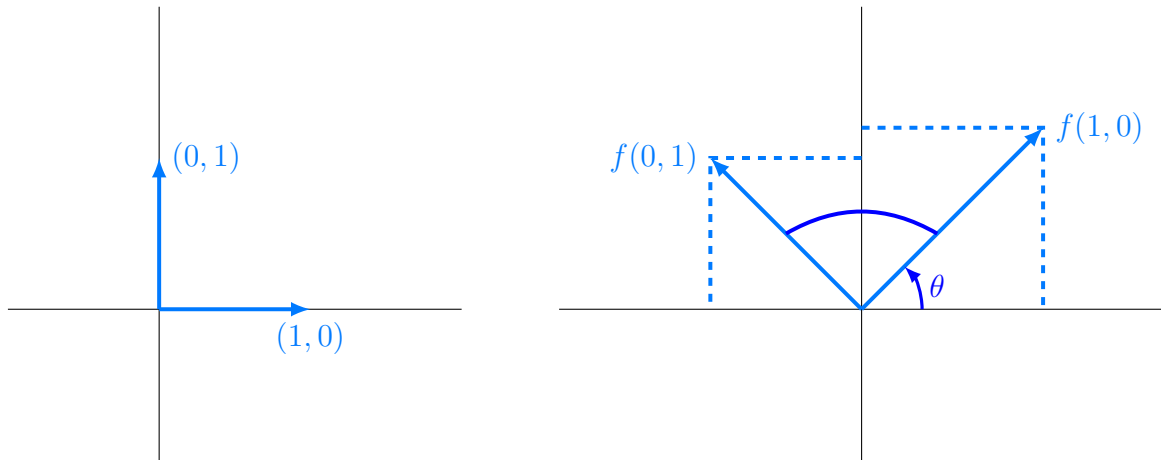
$$f(u + v) = f(x_1 + x_2, y_1 + y_2) = ((x_1 + x_2) \cos \theta - (y_1 + y_2) \sin \theta, (x_1 + x_2) \sin \theta + (y_1 + y_2) \cos \theta) = (x_1 \cos \theta - y_1 \sin \theta, x_1 \sin \theta + y_1 \cos \theta) + (x_2 \cos \theta - y_2 \sin \theta, x_2 \sin \theta + y_2 \cos \theta) = f(u) + f(v)$$

$$2) f(\alpha u) = f(\alpha(x_1, y_1)) = f(\alpha x_1, \alpha y_1) = (\alpha x_1 \cos \theta - \alpha y_1 \sin \theta, \alpha x_1 \sin \theta + \alpha y_1 \cos \theta) = \alpha(x_1 \cos \theta - y_1 \sin \theta, x_1 \sin \theta + y_1 \cos \theta) = \alpha f(u)$$

• Significado geométrico de  $f$

$$f(1, 0) = (\cos \theta, \sin \theta)$$

$$f(0, 1) = (-\sin \theta, \cos \theta)$$



$f$  es un giro de  $\theta$  grados.

Además, por linealidad:

$$\begin{aligned} f(x, y) &= f(x(1, 0) + y(0, 1)) \\ &= xf(1, 0) + yf(0, 1) \\ &= x(\cos \theta, \sin \theta) + y(-\sin \theta, \cos \theta) \\ &= \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos \theta \\ \sin \theta \end{bmatrix} x + \begin{bmatrix} -\sin \theta \\ \cos \theta \end{bmatrix} y \end{aligned}$$

Lo anterior nos sugiere que dada una aplicación lineal  $f : \mathbb{K}^n \longrightarrow \mathbb{K}^n$  y fijados bases  $B$  de  $\mathbb{K}^n$  y  $B'$  de  $\mathbb{K}^m$  existe  $f$  se dice una aplicación lineal ortogonal porque su matriz es ortogonal.

• Matriz asociada a una aplicación lineal

Sean  $B = \{v_1, v_2, \dots, v_n\}$  y  $B' = \{v'_1, v'_2, \dots, v'_m\}$  bases de  $\mathbb{K}^n$  y  $\mathbb{K}^m$ , respectivamente.

Dada  $f : \mathbb{K}^n \longrightarrow \mathbb{K}^m$ , se tiene

$$\begin{aligned} f(v_1) &= a_{11}v'_1 + a_{21}v'_2 + \cdots + a_{m1}v'_m \\ f(v_2) &= a_{12}v'_1 + a_{22}v'_2 + \cdots + a_{m2}v'_m \\ &\vdots \\ f(v_n) &= a_{1n}v'_1 + a_{2n}v'_2 + \cdots + a_{mn}v'_m \end{aligned}$$

A la matriz

$$M_{B \rightarrow B'}(f) = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & & \vdots \\ a_{m1} & a_{m2} & & a_{mn} \end{bmatrix}$$

se le llama matriz asociada a  $f$  en las bases  $B$  y  $B'$ .

Obviamente, se tiene que dado  $x = x_1v_1 + \cdots + x_nv_n$ , entonces

$$f(x) = M_{B \rightarrow B'}(f) \cdot x_B$$

con

$$x_B = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

- Cambio de coordenadas (o de base)

Sea  $f : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ . Consideremos la base canónica de  $\mathbb{R}^n$ , denotada por  $\mathcal{C}$ , y  $B$  es una base distinta de  $\mathbb{R}^n$ . Entonces, obviamente

$$M_{\mathcal{C} \rightarrow \mathcal{C}}(f) \neq M_{BB}(f)$$

$$\begin{array}{ccc} f : \mathbb{R}^n & \longrightarrow & \mathbb{R}^n \\ B & & B \\ \mathcal{C} & & \mathcal{C} \end{array}$$

¿Qué relación existe entre las matrices anteriores?

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{\quad f \quad} & \mathbb{R}^n \\ \mathcal{C} \uparrow & & \downarrow \mathcal{C} \\ \mathbb{R}^n & \xrightarrow{\quad f \quad} & \mathbb{R}^n \\ B & & B \end{array} \quad \begin{array}{l} M_{\mathcal{C} \rightarrow \mathcal{C}}(f) \\ \\ M_{B \rightarrow B}(f) \end{array}$$

- Proposición

Se cumple

$$M_{B \rightarrow B}(f) = M_{C \rightarrow B} \cdot M_{C \rightarrow C}(f) \cdot M_{B \rightarrow C}$$

donde  $M_{B \rightarrow C}$  es la misma matriz de cambio de base de  $B$  a  $C$

$M_{C \rightarrow B} = (M_{B \rightarrow C})^{-1}$  es la misma matriz de cambio de base de  $C$  a  $B$

### • Demostración

Recordemos

$$x_B = M_{C \rightarrow B} X_C$$

y por tanto

$$X_C = (M_{C \rightarrow B}^{-1}) \cdot x_B = M_{B \rightarrow C} \cdot x_B$$

Por tanto:

$$\begin{aligned} (f(x_B))_B &= M_{C \rightarrow B} f(x_B)_C = M_{C \rightarrow B} \cdot M_{C \rightarrow C}(f)(x_B)_C = \\ &= M_{C \rightarrow B} \cdot M_{C \rightarrow C}(f) \cdot (M_{C \rightarrow B}^{-1}) \cdot x_B = M_{C \rightarrow B} \cdot M_{C \rightarrow C}(f) \cdot M_{B \rightarrow C} x_B \end{aligned}$$

Importante: Si dos matrices  $M$  y  $N$  están relacionadas en la forma  $N = P^{-1}MP$ , con  $P$  invertible, entonces ambas matrices están asociadas a la misma aplicación lineal pero en dos bases distintas.

## Valores y vectores propios

### • Definición

Sea  $A \in M_n(\mathbb{K}^n)$ . Se dice que  $\lambda$  es un valor propio o autovalor (eigenvalue) de  $A$  si existe un vector no nulo  $v \in \mathbb{K}^n$ , llamando vector propio o autovector (eigenvector) de modo que  $Au = \lambda v$ .

Al conjunto de todos los valores propios de  $A$  se le llama espectro de  $A$ . Al mayor, en valor absoluto, de los autovalores se le llama radio espectral, denotado  $\rho(A)$ , o valor propio dominante.

—> ¿Cómo calculamos los autovalores y autovectores de una matriz?

$$\begin{aligned} Au = \lambda v &\longleftrightarrow (A - \lambda I)v = 0 \longleftrightarrow v \in \text{Nuc}(A - \lambda I) \neq \{0\} \longleftrightarrow \dim \text{Nuc}(A - \lambda I) = n - r(A - \lambda I) \geq 1 \longleftrightarrow \\ r(A - \lambda I) &< n \longleftrightarrow \det(A - \lambda I) = 0 \end{aligned}$$

### • Definición (Polinomio característico)

Sea  $A \in M_n(\mathbb{K}^n)$ . Se llama polinomio característico de  $A$ , denotado  $P_A(\lambda)$ , al polinomio

$$P_A(\lambda) = \det(A - \lambda I)$$

Valores propios  $\lambda_1 = 9$ ,  $\lambda_2 = 1$



• Vectores propios

$$\lambda_1 = 9 \quad K_{er}(A - 9I) = K_{er} \begin{pmatrix} -4 & 4 \\ 4 & -4 \end{pmatrix}$$

$$\begin{bmatrix} -4 & 4 \\ 4 & -4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left. \begin{array}{l} -4x + 4y = 0 \\ 4x - 4y = 0 \end{array} \right\} \quad y = \alpha \longrightarrow x = \alpha$$

$$K_{er}(A - 9I) = \langle (1, 1) \rangle$$

$$\lambda_2 = 1 \quad K_{er}(A - I) = K_{er} \begin{pmatrix} 4 & 4 \\ 4 & 4 \end{pmatrix}$$

$$\begin{bmatrix} 4 & 4 \\ 4 & 4 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

$$\left. \begin{array}{l} 4x + 4y = 0 \\ 4x + 4y = 0 \end{array} \right\} \quad y = \alpha \longrightarrow x = -\alpha$$

$$K_{er}(A - I) = \langle (-1, 1) \rangle$$

Consideremos la matriz  $P$  formada por los valores propios anteriores pero unitarios, es decir

$$P = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

Nótese que  $P$  es ortogonal. Por tanto  $P^{-1} = P^T$ .

Interpretación en términos de aplicaciones lineales

$$f : \mathbb{R}^2 \longrightarrow \mathbb{R}^2$$

$$(x, y) \longmapsto f(x, y) = (5x + 4y, 4x + 5y)$$

$$C = \{(1, 0), (0, 1)\}$$

$$A = M_{C \rightarrow C}(f) = \begin{bmatrix} 5 & 4 \\ 4 & 5 \end{bmatrix}$$

$$f(1, 0) = (5, 4) = 5 \cdot (1, 0) + 4 \cdot (0, 1)$$

$$f(0, 1) = (4, 5) = 4 \cdot (1, 0) + 5 \cdot (0, 1)$$

$$B = \left\{ \left( \frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right), \left( -\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}} \right) \right\}$$

$$f\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \left(5 \cdot \frac{1}{\sqrt{2}} + 4 \cdot \frac{1}{\sqrt{2}}, 4 \cdot \frac{1}{\sqrt{2}} + 5 \cdot \frac{1}{\sqrt{2}}\right) = 9\left(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

$$f\left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right) = \left(-5 \cdot \frac{1}{\sqrt{2}} + 4 \cdot \frac{1}{\sqrt{2}}, 4 \cdot -\frac{1}{\sqrt{2}} + 5 \cdot \frac{1}{\sqrt{2}}\right) = \left(-\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}}\right)$$

$$A = M_{B \rightarrow B}(f) = \begin{bmatrix} 9 & 0 \\ 0 & 1 \end{bmatrix}$$

- Definición (Factorización en valores propios)

Sea  $A \in M(\mathbb{K}^n)$ . Se dice que  $A$  es diagonalizable si existe una matriz diagonal  $D$ , con los valores propios de  $A$  en su diagonal, y una matriz invertible  $P$  cuyas columnas están formadas por vectores propios de  $A$ , de modo que

$$A = PDP^{-1}$$

A esta factorización se le llama factorización en valores propios.

### Matrices simétricas

- Teorema

Si  $A \in M_n(\mathbb{R})$  es simétrica, entonces

$$A = PDP^{-1}$$

con  $D = \text{diag}(\lambda_1, \dots, \lambda_n)$ ,  $\lambda_j$  valores propios y  $P = [u_1, \dots, u_n]$  base ortonormal de vectores propios.

- Método  $QR$

→ Se utiliza para calcular todos los valores y vectores propios.

→ Es muy estable porque al ser  $Q$  ortogonal, no se propagan errores en los datos.

- Algoritmo

1) Inicialización: tomar  $A_0 = A$

2) Iteración: para  $k \geq 0$   $d_0$ :

2.1)  $A_0 = Q_1 R_1$  factorización  $QR$  de  $A_0$

2.2)  $A_1 = R_1 Q_1$

Nota:  $A_1 = Q_1^T Q_1 R_1 Q_1 = Q_1^T A_0 Q_1$

2.3) Factorización  $QR$  de  $A_1$

$$A_1 = Q_2 R_2$$

$$A_2 = R_2 Q_2$$

Nota:  $A_2 = Q_2^T Q_2 R_2 Q_2 = Q_2^T A_1 Q_2 = Q_2^T Q_1^T A_0 Q_1 Q_2$  y así sucesivamente

### Factorización en valores singulares (SVD)

El principal inconveniente de la factorización en valores propios es que sólo se puede aplicar a matrices cuadradas. De hecho, además las matrices han de ser simétricas para garantizar dicha factorización. Sin embargo, en Ciencia de Datos, aparecen muchas matrices que no son cuadradas ni tampoco simétricas. La factorización *SVD* soluciona ambos problemas y para muchos [es el resultado más importante de la Ciencia de Datos](#).

#### • Definición

Dada  $A \in M_{m \times n}(\mathbb{R})$ , se llama factorización *SVD* (Singular Value Decomposition) de  $A$  a una factorización de la forma

$$A = U \Sigma V^T$$

donde  $\Sigma$  es una matriz diagonal con  $\sigma_1, \dots, \sigma_n$  en su diagonal  $\Sigma \in M_{m \times n}(\mathbb{R})$ .

#### • Demostración

$$\boxed{A} = \boxed{U} \cdot \begin{array}{|c|} \hline \sigma_1 \\ \hline \ddots \\ \hline \sigma_n \\ \hline \end{array} \cdot \boxed{V^T}$$

→ Para que podamos colocar todos los valores singulares  $m \geq n$ , es decir, el número de filas de  $\Sigma$  tiene que ser mayor o igual que el de columnas.

Sea  $m \leq n$ , aplicamos el resultado a  $A^T$  con lo que

$$A^T = U \Sigma V^T$$

de donde tomamos traspuestas

$$A^T = V \Sigma^T U^T$$

Definimos  $U = [u_1, \dots, u_m]$ .

Se tiene

$$U^T A \cdot V = \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} \cdot A[v_1, \dots, v_n] = \begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} \cdot [Av_1, \dots, Av_n] =$$

$$\begin{bmatrix} u_1^T \\ \vdots \\ u_m^T \end{bmatrix} \cdot [\sigma_1 u_1, \dots, \sigma_r u_r, 0, \dots, 0] = [\sigma_1 e_1, \dots, \sigma_r e_r, 0, \dots, 0] = \Sigma$$

Multiplicando a la derecha por  $V^{-1} = V^T$  y a la izquierda por  $U$  se tiene

$$A = U \Sigma V^T$$

### • Vectores propios

$$\lambda_1 = 8 \quad K_{er}(A^T A - 8I)$$

$$\left\{ \begin{array}{l} \begin{bmatrix} -3 & 3 \\ 3 & -3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ -3x + 3y = 0 \\ 3x - 3y = 0 \end{array} \right\} \longrightarrow x = y \longrightarrow v_1 = \frac{1}{\sqrt{2}}(1, 1)$$

$$\lambda_2 = 2 \quad K_{er}(A^T A - 2I)$$

$$\left\{ \begin{array}{l} \begin{bmatrix} 3 & 3 \\ 3 & 3 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \\ 3x + 3y = 0 \\ 3x + 3y = 0 \end{array} \right\} \longrightarrow x = -y \longrightarrow v_1 = \frac{1}{\sqrt{2}}(-1, 1)$$

$$\text{Matriz } V = \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$\text{Valores singulares: } \sigma = \sqrt{8}, \sigma_2 = \sqrt{2} \longrightarrow \Sigma = \begin{bmatrix} \sqrt{8} & 0 \\ 0 & \sqrt{2} \end{bmatrix}$$

### • Cálculo de la matriz $U$

$$u_1 = \frac{1}{\sigma_1} A v_1 = \frac{1}{\sqrt{8}} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

$$u_2 = \frac{1}{\sigma_2} A v_2 = \frac{1}{\sqrt{2}} \begin{bmatrix} 2 & 2 \\ -1 & 1 \end{bmatrix} \cdot \begin{bmatrix} -\frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} \end{bmatrix} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

### Aplicaciones de la SVD

#### 1) Relación de la SVD con los cuatro subespacios fundamentales de una matriz

Recordemos que dada  $A \in M_{m \times n}(\mathbb{R})$ , los cuatro subespacios fundamentales de  $A$  son:

- $\text{Nuc}(A) = \{x \in \mathbb{R}^n : Ax = 0\}$
- $\text{Nuc}(A^T) = \{x \in \mathbb{R}^n : A^T x = 0\}$
- $\text{Fil}(A) \equiv$  subespacio generado por las filas de  $A$
- $\text{Col}(A) \equiv$  subespacio generado por las columnas de  $A$

Por otra parte, la factorización  $SVD$  de  $A$  es

$$A = U\Sigma V^T$$

$$\Sigma \equiv \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_r, 0, \dots, 0)$$

$U \equiv$  matriz ortogonal cuyas columnas son vectores propios de  $AA^T$  con los valores propios asociados  $(\sigma_1^2, \dots, \sigma_r^2, 0, \dots, 0)$ .

$V \equiv$  matriz ortogonal cuyas columnas son vectores propios de  $A^T A$  con los mismos valores propios.

• Observación

$$u_j = \frac{1}{\sigma_j} A v_j, \quad i \leq j \leq r$$

Por tanto,  $u_1, \dots, u_r \in \text{Col}(A)$  y como

$$\boxed{r(A) = r(U\Sigma V^T)}$$

se tiene que  $\text{Col}(A) = \langle u_1, \dots, u_r \rangle$

• Observación

$$A^T = (U\Sigma V^T)^T = V\Sigma U^T$$

$$\text{Fil}(A) = \text{Col}(A^T) = \langle v_1, \dots, v_r \rangle$$

$$\text{Nuc}(A) = \langle u_{r+1}, \dots, u_m \rangle$$

• Interpretación en términos de aplicaciones lineales

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{A} & \mathbb{R}^m \\ V \downarrow & & \uparrow U \\ \mathbb{R}^n & \xrightarrow{\Sigma} & \mathbb{R}^m \end{array}$$

$$\begin{aligned} B &= B_A^n = \{e_1, \dots, e_n\} \text{ canónica} \\ B' &= B_A^m = \{e'_1, \dots, e'_m\} \text{ canónica} \\ B_v &= \{v_1, \dots, v_n\} \text{ ortonormal} \\ B_u &= \{u_1, \dots, u_n\} \text{ ortonormal} \end{aligned}$$

$$A = U\Sigma V^T$$

$V$  = matriz de cambio de base de  $B_A^n$  a  $B_v$

$U$  = matriz de cambio de base de  $B_u$  a  $B_A^m$

## 2) Low-rank approximations

Sea  $A \in M_{m \times n}(\mathbb{R})$  y consideremos su descomposición  $SVD$ , es decir,

$$A = U\Sigma V^T = [u_1, \dots, u_m]_{m \times m} \cdot [\sigma_1 e_1, \dots, \sigma_r e_r, 0, \dots, 0]_{m \times n} \cdot \begin{bmatrix} v_1^T \\ \vdots \\ v_m^T \end{bmatrix}_{n \times n} = [u_1, \dots, u_m]_{m \times m} \cdot \begin{bmatrix} \sigma_1 v_1^T \\ \vdots \\ \sigma_r v_r^T \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{m \times n} = \sigma_1 u_1 v_1^T + \dots + \sigma_r u_r v_r^T$$

Cada sumando  $\sigma_j u_j v_j^T$  es una matriz de tamaño  $m \times n$  pero de rango 1.

Si aproximamos  $A$  por  $A_K = \sigma_1 u_1 v_1^T + \dots + \sigma_K u_K v_K^T$  la matriz  $Ax$  tiene a lo sumo rango  $K$ . Se tiene el siguiente resultado fundamental.

### • Teorema (Eckart-Young-Mirsky)

Con la notación anterior, el mínimo

$$\min_{\text{rango}(x) \leq K} \|A - x\|$$

se alcanza para  $Ax = \sum_{j=1}^K \sigma_j u_j v_j^T$ , es decir, la descomposición  $SVD$  truncada anterior es la mejor aproximación para la norma  $\|\cdot\|_2$ .