

# **ALNAE**

**Mis notas de estudio**

Miguel Trias

2026-06-16

Éste libro compila el material de estudio del curso Álgebra Lineal Para Aprendizaje Estadístico del año 2026 en la Facultad de Ciencias Económicas y Administración de la Universidad de la República (Uruguay). Está basado en el contenido de las clases de los profesores Diego Armentano y Manuel Hernández y en las cátedras del Profesor Gilbert Strang.

# Tabla de contenidos

|   |           |
|---|-----------|
| <b>Prefacio</b>   | <b>8</b>  |
| Sobre este sitio . . . . .                                    | 8         |
| Motivacion . . . . .  | 8         |
| Atribución Intelectual . . . . .                              | 8         |
| ¡Colabora y Reporta! . . . . .                                | 9         |
| Licencia y Uso . . . . .                                      | 9         |
| <b>1 Preliminares</b>   | <b>11</b> |
| 1.1 Numeros Complejos . . . . .                               | 11        |
| 1.1.1 Propiedades . . . . .                                   | 11        |
| 1.1.2 Interpretación geométrica . . . . .                     | 11        |
| 1.1.3 Descomposición Polar de un Número Complejo . . . . .    | 11        |
| 1.2 Determinante de una Matriz Cuadrada . . . . .             | 13        |
| 1.2.1 Propiedades Relevantes . . . . .                        | 13        |
| 1.3 Independencia Lineal, Rango, Menores . . . . .            | 14        |
| <b>2 Matrices de Rango 1</b>                                  | <b>15</b> |
| 2.1 Estructura Espectral . . . . .                            | 16        |
| 2.2 Autovectores . . . . .                                    | 16        |
| 2.3 Resumen Espectral . . . . .                               | 17        |
| 2.4 El Polinomio Característico . . . . .                     | 17        |
| 2.4.1 Observaciones . . . . .                                 | 17        |
| 2.4.2 El Caso $\lambda = 0$ repetido . . . . .                | 18        |
| <b>3 Productos Matriciales, Factorización <math>CR</math></b> | <b>19</b> |
| 3.1 Multiplicación de Matrices por vector $Ax$ . . . . .      | 19        |
| 3.1.1 Perspectiva de Filas . . . . .                          | 19        |
| 3.1.2 Perspectiva de Columnas . . . . .                       | 19        |
| 3.2 Multiplicación de Matrices $AB$ . . . . .                 | 20        |
| 3.2.1 Perspectiva de Columnas . . . . .                       | 20        |
| 3.2.2 Perspectiva de Suma de Productos Externos . . . . .     | 21        |
| 3.3 Teorema de Factorización $A = CR$ . . . . .               | 22        |
| 3.4 Igualdad del Rango por Filas y Columnas . . . . .         | 22        |
| <b>4 Subespacios Fundamentales</b>                            | <b>24</b> |
| 4.1 El Espacio Columna $C(A)$ . . . . .                       | 24        |

|          |   |           |
|----------|---|-----------|
| 4.2      | El Espacio Fila $C(A^T)$ . . . . .                                  | 25        |
| 4.3      | Núcleo o Espacio Nulo $N(A)$ . . . . .                              | 25        |
| 4.4      | Núcleo Izquierdo $N(A^T)$ . . . . .                                 | 25        |
| 4.5      | Teorema de las Dimensiones . . . . .                                | 25        |
| 4.6      | Descomposición Ortogonal . . . . .                                  | 26        |
| 4.6.1    | Corolario: Igualdad del rango por filas y columnas . . . . .        | 27        |
| 4.6.2    | Resumen . . . . .   | 27        |
| 4.7      | El “Gran Mapa” del Álgebra Lineal . . . . .                         | 27        |
| 4.8      | Restricciones y Codimensión . . . . .                               | 28        |
| 4.8.1    | La Fila como Hiperplano . . . . .                                   | 28        |
| 4.8.2    | Intersección y Reducción de Dimensión . . . . .                     | 28        |
| 4.8.3    | Conclusión sobre la Dimensión del Núcleo . . . . .                  | 29        |
| 4.8.4    | Ejemplo de Codimensión en $\mathbb{R}^3$ . . . . .                  | 29        |
| <b>5</b> | <b>Ortogonalidad, Matrices de Stiefel, Proyectores</b> . . . . .    | <b>30</b> |
| 5.1      | Fundamentos de Ortogonalidad . . . . .                              | 30        |
| 5.1.1    | Definición: Vectores Ortogonales . . . . .                          | 30        |
| 5.1.2    | Definición: Conjuntos y Bases Ortonormales . . . . .                | 30        |
| 5.1.3    | Coordenadas en Bases no Ortonormales . . . . .                      | 31        |
| 5.1.4    | Coordenadas en Bases Ortonormales . . . . .                         | 31        |
| 5.2      | Matrices de Stiefel . . . . .                                       | 32        |
| 5.2.1    | Teorema: Preservación de la Norma (Propiedad Isométrica) . . . . .  | 32        |
| 5.3      | El Operador de Proyección Ortogonal $QQ^T$ . . . . .                | 33        |
| 5.3.1    | Prueba Geométrica y Algebraica . . . . .                            | 33        |
| 5.3.2    | Caracterización Geométrica . . . . .                                | 34        |
| 5.3.3    | Aplicación: La Proyección como Mejor Aproximación . . . . .         | 35        |
| <b>6</b> | <b>Matrices de Householder</b> . . . . .                            | <b>36</b> |
| 6.1      | Proyector $vv^T$ . . . . .  | 36        |
| 6.2      | Proyección sobre $v^\perp$ . . . . .                                | 36        |
| 6.3      | Reflexión . . . . .   | 36        |
| 6.3.1    | Definición . . . . .  | 37        |
| 6.3.2    | Propiedades Estructurales . . . . .                                 | 37        |
| 6.3.3    | Espectro y Autovectores . . . . .                                   | 38        |
| 6.3.4    | Utilidad en Factorizaciones . . . . .                               | 38        |
| <b>7</b> | <b>Autovalores, Autovectores, Factorización Espectral</b> . . . . . | <b>40</b> |
| 7.1      | La Ecuación Fundamental . . . . .                                   | 40        |
| 7.2      | Motivación: Convergencia de Algoritmos Numéricos . . . . .          | 41        |
| 7.3      | El Polinomio Característico . . . . .                               | 42        |
| 7.4      | Espectro de la Traspuesta $A^T$ . . . . .                           | 42        |
| 7.5      | Espectro de Matrices Semejantes . . . . .                           | 42        |
| 7.6      | Espectro del Producto $AB$ . . . . .                                | 43        |

|           |   |           |
|-----------|---|-----------|
| 7.7       | Espectro de $A + kI$ . . . . .                                | 43        |
| 7.8       | Vínculo entre Espectro, Determinante y Traza . . . . .        | 43        |
| 7.8.1     | Determinante . . . . .  | 43        |
| 7.8.2     | Traza . . . . .   | 44        |
| 7.8.3     | Verificación en Matrices Diagonalizables . . . . .            | 44        |
| 7.9       | Factorización $X\Lambda X^{-1}$ . . . . .                     | 45        |
| 7.9.1     | Aplicaciones . . . . .  | 45        |
| <b>8</b>  | <b>Matrices Ortogonales</b> . . . . .                         | <b>46</b> |
| 8.0.1     | Definición . . . . .  | 46        |
| 8.0.2     | La Identidad Fundamental . . . . .                            | 46        |
| 8.0.3     | Preservación de la Geometría (Isometría) . . . . .            | 47        |
| 8.0.4     | El Determinante y la Orientación . . . . .                    | 47        |
| 8.1       | Espectro . . . . .  | 48        |
| 8.2       | Autovectores . . . . .  | 49        |
| 8.3       | Ejemplos . . . . .  | 49        |
| <b>9</b>  | <b>Matrices Simétricas</b> . . . . .                          | <b>52</b> |
| 9.1       | Definición . . . . .  | 52        |
| 9.2       | Autovalores Reales . . . . .                                  | 52        |
| 9.3       | Ortogonalidad de Autovectores . . . . .                       | 53        |
| 9.4       | Teorema Espectral para Matrices Simétricas . . . . .          | 54        |
| 9.5       | Factorización Espectral . . . . .                             | 54        |
| 9.6       | Componentes de la Factorización . . . . .                     | 54        |
| 9.7       | Interpretación Geométrica . . . . .                           | 55        |
| 9.8       | Interpretación como Suma de Rango 1 . . . . .                 | 55        |
| <b>10</b> | <b>Matrices Definidas y Semidefinidas Positivas</b> . . . . . | <b>56</b> |
| 10.1      | Definición . . . . .  | 56        |
| 10.2      | Definición . . . . .  | 56        |
| 10.3      | Criterios de verificación . . . . .                           | 57        |
| 10.3.1    | 1: Autovalores Positivos . . . . .                            | 57        |
| 10.3.2    | 2: Energía Positiva . . . . .                                 | 57        |
| 10.3.3    | 3: Factorización $S = A^T A$ . . . . .                        | 57        |
| 10.3.4    | 4: Determinante y Menores Principales Positivos . . . . .     | 58        |
| 10.3.5    | 5: Pivotes de eliminación Gaussiana Positivos . . . . .       | 58        |
| 10.4      | Ejemplos y Aplicaciones Fundamentales . . . . .               | 59        |
| 10.4.1    | Relación con la Matriz Identidad . . . . .                    | 59        |
| 10.4.2    | Desplazamientos (Shifts) . . . . .                            | 59        |
| 10.4.3    | La Matriz de Unos ( $J$ ) . . . . .                           | 59        |
| 10.4.4    | Matrices de Proyección Ortogonal . . . . .                    | 60        |
| 10.4.5    | Matrices de Markov Simétricas . . . . .                       | 60        |

|  |           |
|--|-----------|
| <b>11 SVD(Descomposición en Valores Singulares)</b>                      | <b>61</b> |
| 11.1 Motivación: Limitaciones de los Autovectores . . . . .              | 61        |
| 11.2 La Ecuación Fundamental . . . . .                                   | 61        |
| 11.3 Intuición Geométrica: Rotación y Estiramiento . . . . .             | 62        |
| 11.4 Dedución de la SVD . . . . .  | 62        |
| 11.5 Relación con la Estructura Espectral . . . . .                      | 64        |
| 11.6 Construcción y Prueba de Ortogonalidad . . . . .                    | 65        |
| 11.7 Observaciones y Propiedades . . . . .                               | 65        |
| 11.8 La Pseudo-inversa ( $A^+$ ) . . . . .                               | 66        |
| 11.8.1 Dedución . . . . .  | 66        |
| 11.8.2 Propiedades . . . . .   | 67        |
| <b>12 Normas</b>   | <b>68</b> |
| 12.1 Definición . . . . .  | 68        |
| 12.1.1 La Identidad de Polarización y la Ley del Paralelogramo . . . . . | 68        |
| 12.1.2 Normas Inducidas por Producto Interno . . . . .                   | 69        |
| 12.2 Normas Vectoriales . . . . .  | 69        |
| 12.2.1 Geometría de las Normas . . . . .                                 | 70        |
| 12.2.2 El Caso Especial de $\ell_0$ . . . . .                            | 71        |
| 12.2.3 Normas $S$ . . . . .  | 71        |
| 12.3 Normas en Espacios de Funciones . . . . .                           | 72        |
| 12.3.1 Norma $L^p$ Funcional . . . . .                                   | 72        |
| 12.4 Cauchy - Schwarz . . . . .  | 73        |
| 12.4.1 Dedución Algebraica y Geométrica . . . . .                        | 73        |
| 12.5 Desigualdad Triangular . . . . .                                    | 74        |
| <b>13 Normas Matriciales</b>   | <b>76</b> |
| 13.1 Producto Hilbert-Schmidt . . . . .                                  | 76        |
| 13.2 Norma Nuclear $\ A\ _N$ . . . . .                                   | 77        |
| 13.2.1 Caracterización Variacional . . . . .                             | 78        |
| 13.3 Norma Operador $\ A\ $ . . . . .                                    | 78        |
| 13.4 Norma de Frobenius $\ A\ _F$ . . . . .                              | 80        |
| 13.5 Propiedades de las normas matriciales . . . . .                     | 81        |
| 13.5.1 Regla de Submultiplicatividad para Normas de Matrices . . . . .   | 81        |
| 13.5.2 Invarianza Ortogonal . . . . .                                    | 83        |
| 13.5.3 Vinculación con la SVD . . . . .                                  | 84        |
| <b>14 Teorema de Eckart-Young - Aproximación de Bajo Rango</b>           | <b>87</b> |
| 14.0.1 Enunciado . . . . .   | 87        |
| 14.0.2 Medida del Error de Aproximación . . . . .                        | 87        |
| 14.0.3 Argumento de la Demostración (para la norma $L_2$ ) . . . . .     | 88        |
| 14.0.4 Demostración para Norma $\ell_2$ . . . . .                        | 88        |
| 14.0.5 Extensión a la Norma de Frobenius . . . . .                       | 89        |

|  |            |
|--|------------|
| <b>15 PCA (Análisis de Componentes Principales)</b>                                | <b>91</b>  |
| 15.0.1 Ejemplo . . . . .   | 92         |
| <b>16 Minimos Cuadrados</b>  | <b>93</b>  |
| 16.1 Las Ecuaciones Normales . . . . .   | 94         |
| 16.2 Proyecciones y Ortogonalidad . . . . .  | 94         |
| 16.3 Solución vía SVD La Pseudoinversa de Moore-Penrose $A^+$ . . . . .            | 95         |
| 16.4 Ortogonalización: El método de Gram-Schmidt . . . . .                         | 96         |
| <b>Referencias</b>   | <b>97</b>  |
| <b>Apéndices</b>   | <b>98</b>  |
| <b>Espectro de Matrices Aleatorias</b>   | <b>98</b>  |
| Existencia de Autovectores en el Plano Complejo . . . . .                          | 98         |
| La Probabilidad de Autovalores Reales . . . . .                                    | 98         |
| Diagonalizabilidad casi segura . . . . .   | 99         |
| El Determinante y la Invertibilidad . . . . .                                      | 99         |
| Resumen de Probabilidades Espectrales . . . . .                                    | 99         |
| <b>Factorización <math>LU</math></b>   | <b>101</b> |
| Factorización $A = LU$ . . . . .   | 101        |
| Factorización $PA = LU$ . . . . .  | 102        |
| Complejidad Computacional . . . . .  | 102        |
| <b>Diagonalización</b>   | <b>103</b> |
| Definiciones Previas . . . . .   | 103        |
| Criterio de Diagonalización . . . . .  | 103        |
| Explicación Intuitiva: La Falta de “Direcciones” . . . . .                         | 104        |
| ¿ Puede una matriz 3x3 no nula tener sus 3 valores propios iguales a 0 ? . . . . . | 104        |
| <b>Sintesis</b>  | <b>105</b> |
| <b>Problemas de Optimización</b>   | <b>108</b> |
| Basis Pursuit . . . . .  | 108        |
| Lasso . . . . .  | 109        |
| Ridge Regression . . . . .   | 109        |

# Prefacio

## Sobre este sitio

Este libro digital fue creado utilizando el ecosistema de **Código Abierto**:

- **Quarto**: El sistema de publicación científica y técnica de próxima generación.
- **RStudio IDE**: El entorno de desarrollo integrado para computación estadística.
- **Markdown**: El lenguaje de marcado ligero para la redacción de contenido.
- **GitHub Pages**: Para el alojamiento y despliegue automático del libro.

Todo el código fuente está disponible en el repositorio de Github.

## Motivacion

Este libro digital surge como un esfuerzo personal para organizar y profundizar en los conceptos del curso **Álgebra Lineal para Aprendizaje Estadístico** de la **Facultad de Ciencias Económicas y Administración (UdelaR)**. El objetivo es recopilar toda la información disponible en los materiales del curso, enriquecidos con explicaciones y ejemplos asistidos por inteligencia artificial (**Gemini 1.5**).

## Atribución Intelectual

Es fundamental aclarar que la autoría intelectual de los conceptos, teoremas y la estructura lógica presentados en este libro pertenece a las [fuentes originales citadas](#). El núcleo teórico y la secuencia pedagógica provienen directamente de la obra del **Prof. Gilbert Strang (MIT)** (Strang 2019), sus lecciones magistrales en *OpenCourseWare* (Strang 2018), las notas de clase del **Prof. Diego Armentano** (Armentano 2026) y los repartidos de ejercicios prácticos del curso (Hernández 2026).

Mi rol en este proyecto se limita exclusivamente a la **curaduría, organización y edición de contenidos** para facilitar su estudio. Mi contribución personal se centra en la digitalización de los conocimientos, la creación de un formato interactivo mediante **Quarto**, y la generación de ejemplos aclaratorios asistidos por IA. Las contribuciones originales son ocasionales y tienen

como único fin conectar las ideas de los autores mencionados o adaptar la notación para mayor claridad del lector.

#### Descargo de responsabilidad (Disclaimer)

Este material debe ser utilizado como una **guía de estudio complementaria** y no como una fuente oficial de verdad absoluta. Debido a que gran parte de la redacción y los ejemplos han sido procesados mediante IA:

- **No se ofrecen garantías sobre la exactitud del contenido.**
- Pueden existir errores técnicos, matemáticos o de transcripción que aún no han sido detectados.
- Se recomienda siempre contrastar los conceptos con la bibliografía oficial y las fuentes primarias citadas.

## ¡Colabora y Reporta!

Este es un proyecto abierto. Si encuentras una errata en una fórmula, una explicación confusa o quieres proponer una mejora en el código, no dudes en contribuir.

Puedes contribuir de las siguientes maneras:

1. **Reportando un error:** Abre un *Issue* en el repositorio de GitHub.
2. **Corrigiendo directamente:** Haz un *Fork* del proyecto y envía un *Pull Request* con tus correcciones.

› [Acceder al Repositorio en GitHub](#)

## Licencia y Uso

Este libro se distribuye bajo una licencia [Creative Commons Atribución-NoComercial-CompartirIgual 4.0 Internacional \(CC BY-NC-SA 4.0\)](#).

Esto significa que eres libre de:

- **Compartir:** Copiar y redistribuir el material en cualquier medio o formato.
- **Adaptar:** Remezclar, transformar y construir a partir del material.

Bajo los siguientes términos:

1. **Atribución:** Debes dar crédito de manera adecuada (citando este libro y las fuentes originales).

2. **No Comercial:** No puedes utilizar este material con fines comerciales.
3. **Compartir Igual:** Si remezclas o transformas este material, debes distribuir tus contribuciones bajo la misma licencia que el original.

# 1 Preliminares

Esta sección contiene una breve referencia sobre conocimientos previos, necesarios para entender el contenido del curso

## 1.1. Numeros Complejos

### 1.1.1. Propiedades

- **Suma/Resta:** sumamos las partes reales y las imaginarias
- **Conjugado:** Cambiamos el signo a la parte imaginaria
- **Producto:** Expresamos los numeros como  $z = a + ib$  y aplicamos distributiva
- **Cociente:**

$$\frac{z_1}{z_2} = \frac{z_1 \overline{z_2}}{|z_2|^2}, \quad \text{si } z_2 \neq 0$$

### 1.1.2. Interpretación geométrica

- **Suma:** suma de vectores en el plano
- **Multiplicación:** escala el módulo y **rota** el ángulo
- **Conjugado:** reflejo respecto al eje real
- **División:** divide módulos y **resta** ángulos

**Observacion:** El producto interno de complejos, (vector x vector = vector) no es generalizable a otros espacios. En  $\mathbb{R}^3$  tenemos el producto vectorial usual, pero en  $\mathbb{R}^2$  no hay un producto vectorial que sirva para definir un cuerpo.

### 1.1.3. Descomposición Polar de un Número Complejo

Cualquier número complejo  $z$  en forma binómica se escribe como  $z = a + bi$ , donde  $a, b \in \mathbb{R}$  e  $i$  es la unidad imaginaria ( $i^2 = -1$ ). La **descomposición polar** expresa este mismo punto utilizando coordenadas polares  $(r, \theta)$ .

**El Módulo**  $r$  Representa la magnitud o “longitud” del vector que une el origen con el punto  $z$ . Se obtiene mediante el teorema de Pitágoras aplicado a las componentes real e imaginaria:  
 $r = |z| = \sqrt{a^2 + b^2}$

**El Argumento**  $\theta$  Es el ángulo formado por el vector con el eje real positivo. Se define mediante la función arcotangente, teniendo en cuenta el cuadrante donde se ubica  $z$ :  $\theta = \arg(z) = \arctan\left(\frac{b}{a}\right)$

## Representación Matemática

Existen tres formas estrechamente vinculadas para expresar la descomposición polar:

### 1. Forma Trigonométrica

: Utilizando proyecciones sobre los ejes:

$$\begin{cases} a = r \cos \theta \\ b = r \sin \theta \end{cases} \implies z = r(\cos \theta + i \sin \theta)$$

### 2. Forma Exponencial (Fórmula de Euler)

: Basada en la identidad de Euler:  $e^{i\theta} = \cos \theta + i \sin \theta$ <sup>1</sup>, esta es la forma más compacta y potente para el análisis:  $z = r e^{i\theta}$

### 3. Forma Polar (Notación de Fasores)

: Común en ingeniería eléctrica y física:  $z = r \angle \theta$

## Propiedades y Ventajas Operacionales

La descomposición polar simplifica drásticamente ciertas operaciones algebraicas en comparación con la forma binómica:

**Multiplicación** Se multiplican los módulos y se suman los ángulos.

$$z_1 z_2 = (r_1 r_2) e^{i(\theta_1 + \theta_2)}$$

**Potenciación (Fórmula de De Moivre)** Se eleva el módulo y se multiplica el argumento

$$z^n = r^n (\cos(n\theta) + i \sin(n\theta))$$

$$(r e^{i\theta})^n = r^n e^{in\theta}$$

---

<sup>1</sup>De la fórmula de Euler  $e^{i\theta} = \cos \theta + i \sin \theta$ , evaluando en  $\theta = \pi$  surge

$$e^{i\pi} + 1 = 0$$

En consecuencia,  $e^{i\pi}$  parametriza todos los puntos de  $S^1 = z \in \mathbb{C} : |z| = 1$

**Raíces** Permite hallar las  $n$  raíces de un número complejo distribuyéndolas uniformemente en un círculo de radio  $\sqrt[n]{r}$ .

## 1.2. Determinante de una Matriz Cuadrada

El determinante de una matriz cuadrada  $A \in \mathbb{R}^{n \times n}$  es un escalar que representa el **factor de escala de volumen** de la transformación lineal asociada. Geométricamente, el volumen de un hiperparalelepípedo definido por las columnas de  $A$  es igual a  $|\det(A)|$ .

### 1.2.1. Propiedades Relevantes

#### Determinante del Producto de Matrices

$$\det(AB) = \det(A) \det(B).$$

**Idea de la prueba:** Se basa en la unicidad del determinante como función multilineal alternada. Al definir  $d(B) = \det(AB)$ , se demuestra que  $d(I) = \det(A)$ , lo que fuerza la relación proporcional  $d(B) = \det(A) \det(B)$ .

#### Determinante como Producto de Autovalores

Si  $A$  tiene autovalores  $\lambda_1, \dots, \lambda_n$ :  $\det(A) = \prod_{i=1}^n \lambda_i$ .

**Idea de la prueba:** Evaluando el polinomio característico  $p(\lambda) = \det(A - \lambda I)$  en  $\lambda = 0$ . La forma factorizada  $(\lambda_1 - 0) \dots (\lambda_n - 0)$  coincide con el término constante del polinomio, que es  $\det(A)$ .

#### Determinante de Matrices Semejantes

Si  $B = P^{-1}AP$ , entonces  $\det(B) = \det(A)$ .

**Idea de la prueba:** Por la propiedad del producto,  $\det(P^{-1}AP) = \det(P^{-1}) \det(A) \det(P)$ . Dado que  $\det(P^{-1}) = 1/\det(P)$ , los términos externos se anulan.

## El Determinante en Matrices Ortogonales

Una propiedad fundamental de las matrices ortogonales ( $Q^T Q = I$ ) es que su determinante está restringido a dos valores posibles: 1 o  $-1$ .

**Prueba:** Partiendo de la definición de matriz ortogonal:  $Q^T Q = I$ , aplicamos el determinante a ambos lados  $\det(Q^T Q) = \det(I)$ . Como  $\det(I) = 1$  y el determinante del producto es el producto de los determinantes:  $\det(Q^T) \det(Q) = 1$ . Dado que el determinante de una transpuesta es igual al de la matriz original ( $\det(Q^T) = \det(Q)$ ):

$$(\det(Q))^2 = 1 \implies \det(Q) = \pm 1$$

### 1.3. Independencia Lineal, Rango, Menores

El **rango** de una matriz se define como el número de columnas (o filas) linealmente independientes. Se dice que una matriz  $A \in \mathbb{R}^{m \times n}$  de rango  $r$ , tiene **rango completo** si  $r = \min(m, n)$ .

Una herramienta analítica para determinar el rango de una matriz es el estudio de sus **menores**. Para una matriz de rango  $r$ , existe al menos un menor de tamaño  $r \times r$  con determinante no nulo.

Si una matriz  $A \in \mathbb{R}^{3 \times 2}$  tiene rango 2, al menos uno de sus menores de  $2 \times 2$  debe tener un determinante distinto de cero.

En una matriz genérica (con valores aleatorios), la probabilidad de que un menor sea singular es cero. Formalmente, en el espacio de matrices  $\mathbb{R}^{3 \times 2}$ , el conjunto de matrices con rango menor a 2 (donde todos los menores de  $2 \times 2$  son singulares) forma una **variedad algebraica**. Dado que las raíces de un polinomio no trivial definen un conjunto de medida cero, la probabilidad de que una matriz elegida aleatoriamente sea singular es nula.

## 2 Matrices de Rango 1

§

§

Las matrices de rango 1, representadas por el producto externo entre dos vectores constituyen los bloques de construcción fundamentales del álgebra lineal y la ciencia de datos. Su estructura es notablemente simple: toda su información se comprime en una única dirección. A pesar de esta sencillez, estas matrices son esenciales por varios motivos.

**Definición** Si una matriz  $A$  tiene rango 1, entonces todas sus columnas pueden obtenerse como combinaciones lineales de la primera (llamémosle  $u$ ). Por lo tanto, esta matriz puede expresarse como el **producto exterior** de dos vectores  $u$  y  $v^\top$ , si elegimos las componentes adecuadas de  $v$ :

$$A = uv^\top$$

**Espacio Columna**  $C(A)$  Es el subespacio en el codominio  $\mathbb{R}^m$  generado por las combinaciones lineales de las columnas de  $A$ . Para una matriz de rango 1, este espacio es simplemente la **recta** generada por el vector  $u$ .

**Espacio Fila**  $C(A^\top)$  Es el subespacio en el dominio  $\mathbb{R}^n$  generado por las filas de  $A$ . En el caso de rango 1, es la **recta** generada por el vector  $v$ .

### Ejemplo

Consideremos la matriz  $A$  de tamaño  $2 \times 2$ :

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

Aquí, el vector  $u = [1, 3]^\top$  define la dirección del espacio columna en  $\mathbb{R}^2$ , mientras que el vector  $v = [1, 2]^\top$  define la dirección del espacio fila en  $\mathbb{R}^2$ .

#### **i** Nota

De forma general, cualquier matriz de rango  $r$  puede descomponerse en la suma de  $r$

matrices de rango 1:

$$A = \sum_{i=1}^r c_i r_i^\top$$

Esta perspectiva es esencial para entender factorizaciones avanzadas como SVD o la diagonalización, donde la información de la matriz se “fragmenta” en componentes de importancia decreciente.

## 2.1. Estructura Espectral

Las matrices de rango 1 son los bloques de construcción fundamentales del álgebra lineal numérica. Una matriz  $A \in \mathbb{R}^{n \times n}$  definida como el producto exterior de dos vectores no nulos,  $A = uv^\top$ , posee una geometría espectral muy simple pero poderosa. Este se divide en dos partes: el autovalor asociado a la imagen, o autovalor “vivo” y los autovalores asociados al núcleo o autovalores “nulos” (Strang 2019, I.1; Armentano 2026 clase 4).

### El Autovalor “Vivo” $\lambda_1$

Si multiplicamos la matriz por el vector  $u$ , observamos:

$$Au = (uv^\top)u = u(v^\top u) = u\langle v, u \rangle = \langle v, u \rangle u$$

Esto revela que  $\lambda_1 = v^\top u = \langle v, u \rangle$  (el producto punto entre  $v$  y  $u$ ) es el único autovalor que puede ser distinto de cero (Hernández 2026, 3, ejercicio 12).

### Los Autovalores Nulos $\lambda_2, \dots, \lambda_n$

Dado que el rango de la matriz es  $r = 1$ , la dimensión del núcleo es  $n - r = n - 1$  (Armentano 2026 clase 4). Por lo tanto, existen  $n - 1$  autovalores adicionales que deben ser cero:  $\lambda_2 = 0, \dots, \lambda_n = 0$

## 2.2. Autovectores

La ubicación de los autovectores depende de los subespacios fundamentales de  $A$  (Strang 2019, I.6; Armentano 2026 clase 4):

### Autovector para $\lambda_1 = v^\top u$

Como se demostró arriba, el vector  $x_1 = u$  es el autovector asociado al valor propio no nulo. Este vector define la dirección del espacio columna  $C(A)$ .

### Autovectores para $\lambda = 0$

Los autovectores asociados al cero son todos los vectores que pertenecen al núcleo  $N(A)$ . Por definición, estos son los vectores perpendiculares al espacio fila, generado por  $v$ :

$$Ax = (uv^\top)x = u(v^\top x) = 0 \iff v^\top x = 0$$

Para completar la base de autovectores en  $\mathbb{R}^3$ , simplemente elegimos  $n - 1$  vectores  $x_2, x_3, \dots$  que sean linealmente independientes y perpendiculares a  $v$  (Hernández 2026, Pr3 Ej1)

## 2.3. Resumen Espectral

| Tipo      | Valor Propio                                  | Vector Propio                     | Ubicación |
|-----------|---|-----------------------------------|-----------|
| Principal | $\lambda_1 = v^\top u = \langle u, v \rangle$ | $x_1 = u$                         | $C(A)$    |
| Nulos     | $\lambda_2 \dots \lambda_n = 0$               | $x_2 \perp v, \dots, x_n \perp v$ | $N(A)$    |

### ! Importante

La matriz es diagonalizable solo si  $\lambda_1 \neq 0$  (es decir,  $u \not\perp v$ ).

Si  $v^\top u = 0$ , la matriz tiene todos sus autovalores en cero y no posee una base completa de autovectores, convirtiéndose en una matriz defectiva (Strang 2019, I.6).

## 2.4. El Polinomio Característico

La estructura espectral de una matriz de rango 1,  $A = uv^\top$ , se refleja de manera simplificada en su polinomio característico  $\chi_A(\lambda) = \det(A - \lambda I)$  (Armentano 2026 clase 7; Strang 2019, I.6).

Para una matriz  $n \times n$  de rango 1, el polinomio característico siempre toma la forma:

$$\chi_A(\lambda) = (-\lambda)^{n-1}(v^\top u - \lambda)$$

En el caso específico de una matriz  $3 \times 3$ , esto se reduce a:

$$\chi_A(\lambda) = -\lambda^2(\lambda - v^\top u) = -\lambda^3 + (v^\top u)\lambda^2$$

### 2.4.1. Observaciones

#### Multiplicidad Algebraica del Cero

El factor  $\lambda^{n-1}$  confirma que el valor propio  $\lambda = 0$  tiene una multiplicidad algebraica de al menos  $n - 1$ . Esto es una consecuencia directa de que la matriz tiene un núcleo de dimensión  $n - 1$  (Armentano 2026 clase 4).

#### Relación con la Traza

El coeficiente del término  $\lambda^{n-1}$  (en este caso  $v^\top u$ ) corresponde a la **traza** de la matriz (Strang

2018 Lec. 4 '144). En una matriz de rango 1, la traza es exactamente el producto interno de los vectores que la generan:  $tr(uv^\top) = v^\top u$ .

### **Determinante Nulo**

Como el término constante del polinomio es el determinante y aquí es cero (para  $n > 1$ ), se confirma que toda matriz de rango 1 es singular (Armentano 2026 clase 4; Strang 2019, I.6).

### **2.4.2. El Caso $\lambda = 0$ repetido**

Si los vectores son ortogonales ( $v^\top u = 0$ ), el polinomio se convierte en  $\chi_A(\lambda) = (-\lambda)^n$ . En este escenario, todos los autovalores son cero, pero la matriz solo posee  $n - 1$  autovectores independientes, lo que la hace no diagonalizable.

## 3 Productos Matriciales, Factorización $CR$

§

§

Este capítulo explora la dualidad en la interpretación de la multiplicación matriz-vector y la estructura fundamental de una matriz a través de sus columnas.

### 3.1. Multiplicación de Matrices por vector $Ax$

Sea  $A \in \mathcal{M}_{m \times n}$  una matriz que podemos denotar por sus filas como  $A_{i*}$ , por sus columnas como  $A_j$  o por sus entradas  $a_{ij}$ . Sea  $x = (x_1, x_2, \dots, x_n)^\top \in \mathbb{R}^n$  un vector columna. Entonces, el producto  $Ax$  puede interpretarse de varias maneras (Strang 2018 Lec. 1).

#### 3.1.1. Perspectiva de Filas

En esta interpretación, el resultado es un vector en  $\mathbb{R}^m$  donde cada componente es el **producto interno** (también llamado producto punto) entre una fila de la matriz  $A$  y el vector  $x$ .

$$\begin{bmatrix} - & A_{1*} & - \\ - & A_{2*} & - \\ & \vdots & \\ - & A_{m*} & - \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = \begin{bmatrix} \sum_{j=1}^n a_{1j}x_j \\ \sum_{j=1}^n a_{2j}x_j \\ \vdots \\ \sum_{j=1}^n a_{mj}x_j \end{bmatrix} = \begin{bmatrix} \langle A_{1*}, x \rangle \\ \langle A_{2*}, x \rangle \\ \vdots \\ \langle A_{m*}, x \rangle \end{bmatrix}$$

#### 3.1.2. Perspectiva de Columnas

Alternativamente, el producto  $Ax$  puede verse como una **combinación lineal** de las columnas de la matriz  $A$ , donde los coeficientes son las componentes del vector  $x$ :

$$\begin{bmatrix} | & | & \dots & | \\ A_{*1} & A_{*2} & \dots & A_{*n} \\ | & | & & | \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 A_{*1} + x_2 A_{*2} + \dots + x_n A_{*n} = \sum_{j=1}^n x_j A_{*j}$$

**i** Nota

El vector resultante  $b = Ax$  reside necesariamente en el subespacio generado por las columnas de la matriz  $A$ . De otra manera, el vector solución  $x$ , muestra como expresar el termino independiente  $b$  como combinación lineal de las columnas de  $A$ . Para algunos  $b$ , esto es imposible porque no estan en el espacio generado por las columnas de  $A$ .

La interpretación como combinación lineal de las columnas de  $A$  es un pilar central del curso, ya que permite visualizar el producto como un movimiento dentro del subespacio generado por las columnas de la matriz.

Otra manera de visualizar el producto  $Ax$  en la perspectiva de columnas, es que, tomando un vector aleatorio  $x$  de  $\mathbb{R}^n$  obtenemos un vector del espacio  $C(A)$ . A su vez, si pensamos a  $A$  como la matriz que representa a una cierta transformación lineal  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  entonces  $C(A)$  corresponde al subespacio  $\text{Im}(T)$

Es fácil ver como consecuencia de lo anterior, que la dimensión del espacio generado por  $A$  coincide con su rango.  $\dim C(A) = \dim \text{Im}(t) = \text{rango}(A)$

### Ejemplo

En este ejemplo, las primeras igualdades corresponden a la interpretación usual en el cálculo, donde cada entrada de la matriz resultante es el producto punto entre las filas de la matriz  $A$  y el vector  $x$ . La última igualdad, corresponde a la interpretación de  $Ax$  como combinación lineal de las columnas de  $A$ . (Strang 2019, 1.1)

$$\begin{bmatrix} 2 & 3 \\ 2 & 4 \\ 3 & 7 \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} \langle A_{1*}, x \rangle \\ \langle A_{2*}, x \rangle \\ \langle A_{3*}, x \rangle \end{bmatrix} = \begin{bmatrix} 2x_1 + 3x_2 \\ 2x_1 + 4x_2 \\ 3x_1 + 7x_2 \end{bmatrix} = x_1 \begin{bmatrix} 2 \\ 2 \\ 3 \end{bmatrix} + x_2 \begin{bmatrix} 3 \\ 4 \\ 7 \end{bmatrix}$$

## 3.2. Multiplicación de Matrices $AB$

La interpretación del producto  $Ax$  como una combinación lineal de columnas se extiende de manera natural al producto de dos matrices  $A \in \mathbb{R}^{m \times n}$  y  $B \in \mathbb{R}^{n \times p}$ . Existen dos perspectivas fundamentales para descomponer esta operación.

### 3.2.1. Perspectiva de Columnas

Si denotamos las columnas de  $B$  como  $B_{*1}, B_{*2}, \dots, B_{*p}$ , el producto  $AB$  puede visualizarse como una colección de productos matriz-vector:

$$AB = A \left[ \begin{array}{c|c|c|c} | & | & \dots & | \\ B_{*1} & B_{*2} & & B_{*p} \\ | & | & & | \end{array} \right] = \left[ \begin{array}{c|c|c|c} | & | & \dots & | \\ AB_{*1} & AB_{*2} & & AB_{*p} \\ | & | & & | \end{array} \right]$$

Bajo esta óptica, **cada columna de  $AB$  es una combinación lineal de las columnas de  $A$** , donde los coeficientes de la  $j$ -ésima combinación provienen de la  $j$ -ésima columna de  $B$ .

Observación: Esto muestra que el espacio columna del producto está contenido en el espacio columna de la matriz de la izquierda:  $C(AB) \subseteq C(A)$ .

### 3.2.2. Perspectiva de Suma de Productos Externos

Al multiplicar  $A \in \mathbb{R}^{m \times n}$  por  $B \in \mathbb{R}^{n \times p}$ , cada columna  $k$  de  $A$  se encuentra exclusivamente con la fila  $k$  de  $B$ .

El producto de la columna  $k$  de  $A$  ( $m \times 1$ ) por la fila  $k$  de  $B$  ( $1 \times p$ ) genera una matriz completa de tamaño  $m \times p$ .

Este producto, denominado **producto exterior**, produce una matriz donde todas las columnas son múltiplos de la columna de  $A$  y todas las filas son múltiplos de la fila de  $B$ , lo que define una matriz de rango 1.

La matriz final  $AB$  es simplemente la suma de estos  $n$  bloques de construcción de rango 1 (Strang 2018 Lec. 1, '7):

$$AB = \sum_{k=1}^n A_{*k} B_k = A_{*1} B_1 + A_{*2} B_2 + \cdots + A_{*n} B_n$$

donde:

- $A_{*k}$  es la  $k$ -ésima **columna** de  $A$ .
- $B_k$  es la  $k$ -ésima **fila** de  $B$ .

Cada término  $A_{*k} B_k$  genera una matriz donde todas las columnas son múltiplos de  $A_{*k}$  y todas las filas son múltiplos de  $B_k$ .

Veamos por que el elemento  $(AB)_{ij}$  de la matriz resultante es idéntico en ambos métodos.

La entrada  $(i, j)$  de la  $k$ -ésima matriz de la suma es exactamente  $A_{ik} B_{kj}$ . Al sumarlas todas, recuperamos la fórmula estándar:  $\sum_{k=1}^n A_{ik} B_{kj}$ .

En el método de productos internos (filas por columnas), tenemos:

$$(AB)_{ij} = \sum_{k=1}^n A_{ik} B_{kj}$$

En la suma de productos exteriores, el elemento  $(i, j)$  de la  $k$ -ésima matriz  $A_{*k} B_k$  es precisamente  $a_{ik} b_{kj}$ . Al sumar las  $n$  matrices, obtenemos exactamente la misma sumatoria para cada entrada de la matriz final. <sup>1</sup>

---

<sup>1</sup>Esta visión de sumar piezas de rango 1 es un pilar de las grandes factorizaciones del curso. La factorización

### **i** Nota

Desde el punto de vista computacional, multiplicar una matriz  $m \times n$  por una  $n \times p$  requiere  $mnp$  multiplicaciones, independientemente del orden o la perspectiva utilizada.

## 3.3. Teorema de Factorización $A = CR$

(Strang 2019, 1.1) Sea  $A \in \mathbb{R}^{m \times n}$  una matriz de rango  $r$ . Existen una matriz  $C \in \mathbb{R}^{m \times r}$ , y una matriz  $R \in \mathbb{R}^{r \times n}$  tales que:<sup>2</sup>

$$A = CR$$

- La matriz  $C \in \mathbb{R}^{m \times r}$  se construye seleccionando las primeras  $r$  columnas linealmente independientes de  $A$ .
- La matriz  $R \in \mathbb{R}^{r \times n}$  contiene los coeficientes necesarios para reconstruir todas las columnas de  $A$  a partir de la base  $C$ . Notablemente,  $R$  se encuentra a menudo en forma escalonada reducida (o contiene un bloque identidad  $I_r$ ).

### Ejemplo:

(Strang 2018 Lec. 1 '20)

$$A = \begin{pmatrix} 2 & 1 & 3 \\ 3 & 1 & 4 \\ 5 & 7 & 12 \end{pmatrix} = \begin{pmatrix} 2 & 1 \\ 3 & 1 \\ 5 & 7 \end{pmatrix} \cdot \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix}$$

### Observaciones:

- Las columnas de  $C$  forman una base de  $C(A)$ .
- Las filas de  $R$  forman una base de  $C(A^T)$ .

## 3.4. Igualdad del Rango por Filas y Columnas

Un pilar del álgebra lineal es que  $\dim C(A) = \dim C(A^T) = r$ . La factorización  $A = CR$  ofrece una prueba constructiva:

---

*LU* resta sucesivamente matrices de rango 1 para eliminar entradas, y la factorización SVD y Descomposición Espectral, reconstruyen la matriz original sumando sus componentes más importantes de mayor a menor valor singular/autovalor.

<sup>2</sup>Esta factorización permite la **reducción de dimensionalidad**. Si  $r \ll n$  y  $r \ll m$ , la matriz puede almacenarse utilizando  $r(m+n)$  coeficientes en lugar de  $mn$ . Esto es la base conceptual de técnicas de compresión y de aproximación de matrices de bajo rango en ciencia de datos.

- Las columnas de  $C$  son una base de  $C(A)$  por construcción ( $\dim C(A) = r$ ).
- De la igualdad  $A = CR$ , se observa que cada fila de  $A$  es una combinación lineal de las filas de  $R$  (los coeficientes de dicha combinación se encuentran en las filas de  $C$ ). Como  $R$  tiene  $r$  filas independientes, el espacio fila tiene dimensión  $r$ , igual a la dimensión del espacio de columnas.

**Observar:** (Strang 2018 Lec. 1 '29)

- Rango de columnas de  $A = \#$  columnas de  $C =$  rango de filas de  $A = \#$  filas de  $R = \text{rango}(A)$
- Obtenemos las columnas de  $A$  como combinaciones lineales de las columnas de  $C$  utilizando como coeficientes las columnas de  $R$
- Obtenemos las filas de  $A$  como combinaciones lineales de las filas de  $R$  utilizando como coeficientes las filas de  $C$

## 4 Subespacios Fundamentales

§

§

Toda matriz  $A \in \mathbb{R}^{m \times n}$  actúa como una transformación lineal  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  que vincula subespacios vectoriales específicos en el dominio y en el codominio mediante la transformación  $T(x) = Ax$ . Esta acción se comprende a través de **cuatro subespacios fundamentales**, organizados en pares de complementos ortogonales. La comprensión de estos subespacios es fundamental para el análisis de solubilidad y estabilidad numérica. El objetivo de esta sección es comprender a una matriz a través de estas cuatro piezas que la componen.

### 4.1. El Espacio Columna $C(A)$

El **espacio columna** de una matriz  $A \in \mathbb{R}^{m \times n}$ , denotado como  $C(A)$ , es el subespacio de  $\mathbb{R}^m$  formado por todas las combinaciones lineales posibles de sus columnas.

$$C(A) = \{b \in \mathbb{R}^m : b = Ax \quad \forall x \in \mathbb{R}^n\}$$

El subespacio  $C(A) \subseteq \mathbb{R}^m$  representa la **imagen** de la transformación lineal  $T$ , es decir, todos los vectores  $b$  para los cuales el sistema  $Ax = b$  tiene solución.

Geoméricamente, si las columnas de una matriz  $3 \times 2$  son linealmente independientes, el espacio columna representa un plano en  $\mathbb{R}^3$  que pasa por el origen.

El sistema de ecuaciones  $Ax = b$  tiene solución si y solo si el vector  $b$  pertenece al espacio columna  $C(A)$ . Si en cambio,  $b \notin C(A)$ , el sistema se denomina **incompatible** o inconsistente. En tales casos, el álgebra lineal numérica busca la “mejor aproximación” mediante métodos como mínimos cuadrados, proyectando  $b$  sobre el subespacio  $C(A)$ .

## 4.2. El Espacio Fila $C(A^\top)$

Similarmente podemos definir el espacio Fila de  $A$  como el espacio generado por las filas de la matriz. Para no introducir otro simbolo, observamos que las filas de  $A$  son las columnas de su transpuesta, por tanto utilizamos la notacion  $C(A^\top)$ .

El subespacio  $C(A^\top) \subseteq \mathbb{R}^n$  contiene las direcciones de los datos de entrada que la matriz realmente “ve” y transforma hacia la imagen.

## 4.3. Núcleo o Espacio Nulo $N(A)$

El conjunto de todos los vectores  $x \in \mathbb{R}^n$  tales que  $Ax = 0$ . Reside en el dominio  $\mathbb{R}^n$ . Estos vectores representan la información que se “pierde” o se colapsa al origen durante la transformación. A veces tambien se lo denota  $\text{Ker}(A)$ .

## 4.4. Núcleo Izquierdo $N(A^\top)$

El conjunto de todos los vectores  $y \in \mathbb{R}^m$  tales que  $A^\top y = 0$ . Representa las combinaciones lineales de las filas de  $A$  que resultan en el vector nulo.

## 4.5. Teorema de las Dimensiones

(tambien llamado Teorema del Rango o Teorema fundamental del algebra)

Para cualquier matriz  $A \in \mathbb{R}^{m \times n}$  de rango  $r$  que actúa como una transformación lineal  $A : \mathbb{R}^n \rightarrow \mathbb{R}^m$ , el Teorema de las Dimensiones establece que la dimensión del dominio se compone por la suma de las dimensiones del núcleo y de la imagen:

$$n = \dim N(A) + \dim C(A) \tag{4.1}$$

En términos prácticos, esto significa que el rango ( $r$ ) más la nulidad ( $n - r$ ) siempre es igual al número de columnas de la matriz.  $n = \dim N(A) + r$

## 4.6. Descomposición Ortogonal

Al resolver el sistema homogéneo  $Ax = 0$ , buscamos el espacio nulo  $N(A)$ , compuesto por todos los vectores  $x$  que la transformación lineal  $A$  mapea al vector nulo.

Desde una perspectiva algebraica, esto equivale a encontrar los coeficientes de una combinación lineal de las columnas de  $A$  que resulte en cero; si existen vectores  $x \neq 0$  que cumplen esto, las columnas son linealmente dependientes.

Desde una perspectiva geométrica, y utilizando la perspectiva de filas del producto  $Ax$  Sección 3.1.1, las componentes del vector resultante son los productos internos de cada fila  $A_{*i}$  con  $x$ , y estos tienen que dar cero:  $\langle A_{*i}, x \rangle = 0$ .

Es decir, que para todo  $v \in N(A)$  y todo  $w \in C(A^\top)$ , se cumple  $\langle v, w \rangle = 0$ . Esto implica que cualquier vector en el núcleo es perpendicular a todas las filas de la matriz. (Strang 2018 Lec. 2 '35)

### ! Importante

El espacio nulo  $N(A)$  es el conjunto de vectores ortogonales al espacio de filas  $C(A^\top)$ , esto es, el **complemento ortogonal**

$$N(A) = C(A^\top)^\perp$$

Esto significa que el espacio de partida se puede descomponer como una suma directa de estos subespacios perpendiculares. Es decir, el dominio  $\mathbb{R}^n$  admite una **descomposición ortogonal** dada por:

$$\mathbb{R}^n = N(A) \oplus C(A^\top) \quad (4.2)$$

De manera análoga, para el codominio se verifica:

$$N(A^\top) = C(A)^\perp$$

Por lo que:

$$\mathbb{R}^m = C(A) \oplus N(A^\top)$$

Utilizando estas propiedades de los subespacios fundamentales, podemos ver que al aplicar la transformación  $A$  a un vector  $x$ ; si éste tiene una componente en el núcleo, esa parte será anulada por  $A$ , mientras que la componente en el espacio fila será mapeada biunívocamente hacia el espacio columna.

Por lo tanto, para conocer totalmente la transformación  $A$ , solo necesitamos conocer como transforma los vectores en  $C(A^\top)$  porque los vectores de  $N(A)$  siempre van a parar al 0.

### 4.6.1. Corolario: Igualdad del rango por filas y columnas

Al tomar dimensiones en ambos lados de la Ecuación 4.2:

$$n = \dim N(A) + \dim C(A^\top)$$

Lo que a su vez, junto con la Ecuación 4.1 implica:

$$\dim C(A^\top) = \dim C(A) = r$$

Esta igualdad revela que toda la información linealmente independiente de una matriz se refleja de manera simétrica tanto en sus filas como en sus columnas. Mientras que las filas describen las restricciones sobre las soluciones del sistema ( $Ax = 0$ ), las columnas describen la imagen o el alcance de la transformación. El rango  $r$  cuantifica estas “direcciones independientes” de manera unívoca para ambos espacios.

El rango por filas es idénticamente igual al rango por columnas. Este hecho garantiza que la matriz  $A$  y su traspuesta  $A^\top$  comparten la misma capacidad de generación de dimensiones independientes en sus respectivos subespacios.

### 4.6.2. Resumen

$A \in \mathcal{M}_{m \times n}$  de rango  $r$ , representa una transformación lineal  $T : \mathbb{R}^n \rightarrow \mathbb{R}^m$  que puede caracterizarse a través de los subespacios:

| Subespacio              | Símbolo     | Dimensión | Espacio        |
|-------------------------|-------------|-----------|----------------|
| <b>Espacio Columna</b>  | $C(A)$      | $r$       | $\mathbb{R}^m$ |
| <b>Núcleo Izquierdo</b> | $N(A^\top)$ | $m - r$   | $\mathbb{R}^m$ |
| <b>Espacio Fila</b>     | $C(A^\top)$ | $r$       | $\mathbb{R}^n$ |
| <b>Núcleo</b>           | $N(A)$      | $n - r$   | $\mathbb{R}^n$ |

## 4.7. El “Gran Mapa” del Álgebra Lineal

La acción de  $A$  puede visualizarse como una correspondencia biunívoca entre el espacio fila y el espacio columna (Armentano 2026 Clase 4). Los vectores en  $N(A)$  son mapeados al vector nulo en el codominio. Si restringimos el dominio de  $A$  únicamente al subespacio  $C(A^\top)$ , el operador resultante:

$$A : C(A^\top) \rightarrow C(A)$$

es un isomorfismo (es decir, es invertible sobre estos subespacios específicos). Esto implica que para cada vector en la imagen  $C(A)$ , existe un único vector en el espacio fila  $C(A^T)$  que lo produce.<sup>1</sup>

## 4.8. Restricciones y Codimensión

Para comprender la dimensión del núcleo  $N(A)$ , resulta útil visualizar las filas de la matriz  $A \in \mathbb{R}^{m \times n}$  como una serie de restricciones geométricas aplicadas sobre el espacio  $\mathbb{R}^n$ .

Un vector  $x$  pertenece a  $N(A)$  si y solo si es ortogonal a cada vector fila  $A_{i*}$  de la matriz  $A$ . Si denotamos el rango de la matriz como  $r$ , la existencia de  $r$  filas independientes impone  $r$  restricciones lineales sobre el dominio, lo que reduce la dimensión del conjunto solución a  $n - r$ .

Por esta razón es útil hablar de **espacios de codimensión**  $r$ , es decir, todo lo que le falta al espacio para abarcar la dimensión máxima. Esta manera de expresarlo es útil porque no depende de la dimensión del espacio de partida  $n$ .

### 4.8.1. La Fila como Hiperplano

Consideremos una única ecuación lineal  $A_{1*}x = 0$ , donde  $A_{1*}$  es la primera fila de la matriz. Geométricamente, esta ecuación define un subespacio de **codimensión 1**.

- Si estamos en  $\mathbb{R}^n$ , el conjunto de soluciones es un hiperplano de dimensión  $n - 1$ .
- Este hiperplano es, por definición, el subespacio ortogonal al vector fila  $A_{1*}$ .

### 4.8.2. Intersección y Reducción de Dimensión

Cuando añadimos una segunda fila  $A_{2*}$ , estamos buscando la intersección de dos subespacios.

- **Si las filas son linealmente independientes** la intersección de estos dos hiperplanos es “transversal”, lo que reduce la dimensión del espacio solución en exactamente uno más. Así, el espacio de soluciones pasa a tener dimensión  $n - 2$ .
- **Si las filas son dependientes**, por ejemplo, si  $A_{2*} = 2A_{1*}$ , la segunda ecuación no aporta información nueva. La intersección de los subespacios es el mismo hiperplano original, y la dimensión no decrece.

---

<sup>1</sup>Esta correspondencia es la base para entender la pseudo-inversa y la resolución de sistemas mediante SVD.

### 4.8.3. Conclusión sobre la Dimensión del Núcleo

Este proceso se generaliza para las  $m$  filas de la matriz. Si la matriz tiene rango  $r$  (es decir,  $r$  filas linealmente independientes), la dimensión del núcleo se calcula como:

$$\dim N(A) = n - r$$

#### Tip

Cada “dirección independiente” capturada por el espacio fila  $C(A^\top)$  consume una dimensión del dominio, dejando el resto para el núcleo. Esto refuerza la idea de que el núcleo y el espacio fila son complementarios: lo que una fila independiente “restringe”, se pierde para el núcleo pero se gana para la descripción del sistema.

### 4.8.4. Ejemplo de Codimensión en $\mathbb{R}^3$

Si tenemos una matriz con  $n = 3$  y una sola fila independiente ( $r = 1$ ), el núcleo es un plano (dimensión  $3 - 1 = 2$ ). Al añadir una segunda fila independiente ( $r = 2$ ), el núcleo colapsa a una recta (dimensión  $3 - 2 = 1$ ). Si añadimos una tercera fila independiente, el único vector perpendicular a las tres direcciones es el vector nulo (dimensión 0).

# 5 Ortogonalidad, Matrices de Stiefel, Proyectores

§

§

Las matrices ortogonales, representan las transformaciones más estables en el cálculo numérico. Estas preservan la geometría del espacio, lo que garantiza que los errores de redondeo no se amplifiquen durante las factorizaciones.

## 5.1. Fundamentos de Ortogonalidad

### 5.1.1. Definición: Vectores Ortogonales

Dos vectores  $x, y \in \mathbb{R}^n$  se definen como ortogonales si su producto escalar es nulo:

$$x^\top y = \sum_{i=1}^n x_i y_i = 0$$

Para el caso complejo ( $\mathbb{C}^n$ ), la definición requiere el uso del conjugado traspuesto (producto hermitiano) para asegurar que la norma sea un número real no negativo:

$$x^* y = \bar{x}^\top y = \sum_{i=1}^n \bar{x}_i y_i = 0$$

### 5.1.2. Definición: Conjuntos y Bases Ortonormales

Un conjunto de vectores  $\{q_1, q_2, \dots, q_k\}$  es **ortonormal** si cumple la condición de la delta de Kronecker:

$$q_i^\top q_j = \delta_{ij} = \begin{cases} 1 & \text{si } i = j \\ 0 & \text{si } i \neq j \end{cases}$$

Decimos que un conjunto de vectores es una **base ortonormal** si es un conjunto ortogonal de vectores unitarios ( $\|q_i\| = 1 \quad i = 1, 2, \dots, k$ ) que genera todo el espacio.

### 5.1.3. Coordenadas en Bases no Ortonormales

Sea  $\mathcal{B} = \{b_1, b_2, \dots, b_n\}$  una base **no ortogonal** de  $V$ . Para expresar un vector  $v \in V$  como una combinación lineal de los elementos de dicha base, debemos encontrar un conjunto de escalares  $\{c_i\}$  tales que:

$$v = \sum_{i=1}^n c_i b_i$$

Este planteamiento requiere resolver el sistema lineal  $Bc = v$ , donde  $B$  es la matriz cuyas columnas son los vectores de la base. Esta operación conlleva varias dificultades desde una perspectiva computacional.<sup>1</sup>

### 5.1.4. Coordenadas en Bases Ortonormales

Cualquier vector  $v \in \mathbb{R}^n$  puede expresarse de manera única como una combinación lineal de las columnas  $q_1, \dots, q_n$  de una matriz ortogonal  $Q$ :

$$v = c_1 q_1 + c_2 q_2 + \dots + c_n q_n$$

Multiplicando la expresión de  $v$  por  $q_1^\top$  a la izquierda:

$$q_1^\top v = c_1 (q_1^\top q_1) + c_2 (q_1^\top q_2) + \dots + c_n (q_1^\top q_n)$$

Por la propiedad de ortonormalidad,  $q_1^\top q_1 = 1$  y  $q_1^\top q_j = 0$  para todo  $j \neq 1$ , resultando en  $c_1 = q_1^\top v$ .

#### Conclusión

Sea  $Q = \{q_1, q_2, \dots, q_n\}$  una base ortonormal de  $\mathbb{R}^n$ .

$\forall v \in \mathbb{R}^n : v = c_1 q_1 + c_2 q_2 + \dots + c_n q_n$

Los coeficientes  $c_i$  se obtienen mediante productos internos individuales:  $c_i = q_i^\top v$

---

1

- **Complejidad Algorítmica:** La resolución mediante métodos directos (como la eliminación gaussiana o la descomposición LU) implica un costo de  $O(n^3)$  operaciones de punto flotante.
- **Interdependencia de Coeficientes:** El cálculo de cualquier coeficiente  $c_i$  está acoplado al resto del sistema, lo que impide una resolución parcial o asíncrona de las coordenadas. A diferencia de las bases arbitrarias, donde el cálculo de las coordenadas exige resolver un sistema lineal mediante eliminación ( $O(n^3)$ ), en una base ortonormal los coeficientes se obtienen mediante productos internos individuales ( $O(n)$ ):

En notación matricial:

$$c = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix} = \begin{bmatrix} q_1^\top v \\ q_2^\top v \\ \vdots \\ q_n^\top v \end{bmatrix} = \begin{bmatrix} \langle v, q_1 \rangle \\ \langle v, q_2 \rangle \\ \vdots \\ \langle v, q_n \rangle \end{bmatrix} = \begin{bmatrix} - & q_1^\top & - \\ \vdots & \vdots & \vdots \\ - & q_n^\top & - \end{bmatrix} v = Q^\top v$$

Por lo tanto, el vector de coeficientes completo se obtiene aplicando la matriz  $Q^\top$  al vector  $v$ :

$$c = Q^\top v$$

## 5.2. Matrices de Stiefel

Una matriz  $Q$  de tamaño  $n \times r$  se denomina **Matriz de Stiefel** si sus columnas son vectores ortonormales en  $\mathbb{R}^n$ .

Una consecuencia directa de esta definición es la identidad fundamental:

$$Q^\top Q = \mathbb{I}_r$$

donde  $\mathbb{I}_r$  es la matriz identidad de tamaño  $r \times r$ .

$$Q^\top Q = \begin{bmatrix} - & q_1^\top & - \\ \vdots & \vdots & \vdots \\ - & q_r^\top & - \end{bmatrix} \begin{bmatrix} | & \dots & | \\ q_1 & \dots & q_r \\ | & \dots & | \end{bmatrix} = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix} = \mathbb{I}_r$$

### ! Importante

Si  $n > r$ , la matriz  $QQ^\top$  no es la identidad, sino un operador de proyección sobre el subespacio generado por las columnas de  $Q$ .

### 5.2.1. Teorema: Preservación de la Norma (Propiedad Isométrica)

La multiplicación por una matriz de Stiefel preserva la norma euclídea de cualquier vector  $x \in \mathbb{R}^r$ . (Armentano 2026 clase 4; Strang 2018 Lec. 3 '6)

#### Demostración:

Sea  $x \in \mathbb{R}^r$ . La norma al cuadrado del vector transformado  $Qx$  es:

$$\|Qx\|^2 = (Qx)^\top (Qx) = x^\top (Q^\top Q)x = x^\top I_r x = x^\top x = \|x\|^2$$

Por lo tanto,  $\|Qx\| = \|x\|$ . Geométricamente, esto significa que  $Q$  actúa como una rotación o reflexión, manteniendo intactas las distancias y los ángulos originales del dominio.

### 5.3. El Operador de Proyección Ortogonal $QQ^\top$

Es imperativo distinguir entre los dos posibles productos de una matriz de Stiefel  $Q \in \mathbb{R}^{n \times r}$  y su transpuesta:

**Matriz Identidad**  $Q^\top Q = \mathbb{I}_r$ . Puede visualizarse como el producto interno de filas por columnas, donde cada entrada de la matriz resultante es el producto punto de una fila de  $Q$  por una columna. Este producto interno produce la identidad de tamaño  $r$ , confirmando su ortonormalidad.

**Proyección Ortogonal**  $QQ^\top = P$ . En este caso, ayuda ver a  $Q$  como un conjunto de columnas  $q_1, q_2, \dots, q_r$ , y a  $QQ^\top$  como el producto de columnas por filas lo que convierte cada término de  $q_i q_i^\top$  en un producto exterior. A diferencia del producto punto (que da un número), el producto exterior de un vector por sí mismo genera una matriz de rango 1 de tamaño  $n \times n$ .

$$P = q_1 q_1^\top + q_2 q_2^\top + \dots + q_r q_r^\top$$

Si  $r < n$ , esta matriz es singular, de rango  $r$  y actúa como un **proyector ortogonal** sobre el espacio columna  $C(Q)$ .

#### 💡 Tip

---

|                           |  |
|---------------------------|--|
| $Q^\top Q = \mathbb{I}_r$ | Identidad de tamaño $r$                                  |
| $QQ^\top = P_Q$           | Proyección Ortogonal sobre $C(Q)$ $Pv = QQ^\top v = v_Q$ |

#### 5.3.1. Prueba Geométrica y Algebraica

Para que un operador lineal sea una proyección ortogonal, debe satisfacer algunas condiciones estructurales que garantizan la estabilidad y la geometría del mapeo.

##### Idempotencia

Aplicar la proyección sucesivamente no debe alterar el vector tras la primera aplicación.  $P^2 = (QQ^\top)(QQ^\top) = Q(Q^\top Q)Q^\top = Q(I_r)Q^\top = QQ^\top = P$  Entonces,  $P^2 = P$ .

## Invarianza en el subespacio de proyeccion

Si  $v \in C(Q)$ , entonces existe un  $x$  tal que  $v = Qx$ . Al aplicar el proyector:  $Pv = P(Qx) = (QQ^\top)Qx = Q(Q^\top Q)x = Qx = v$ . Por lo tanto,  $P$  deja invariantes a los vectores que ya pertenecen a su imagen.

## Simetría

Si estamos proyectando un vector  $x$  sobre  $Q$ , su proyección es  $Px$  y el error es  $x - Px$ . Para que esta proyección sea ortogonal, el error de la proyección debe ser perpendicular al espacio donde se proyecta, es decir, que el error debe ser perpendicular a cualquier vector  $P_y$  en la imagen de  $P$ . Por lo tanto debe cumplirse

$$(Py)^\top(x - Px) = 0, \forall y$$

Desarrollando esta formula:

$$(Py)^\top(x - Px) = y^\top P^\top(x - Px) = y^\top(P^\top x - P^\top Px) = y^\top(P^\top - P^\top P)x$$

Para que esto sea cero para todo  $x$  e  $y$ , debe cumplirse  $P^\top = P^\top P$ .

Si además la matriz  $P$  fuese simétrica ( $P = P^\top$ ), entonces tendríamos  $P^\top = P^\top P = PP = P$ . Por lo tanto, la simetría de  $P$  garantiza que la proyección sea ortogonal y no oblicua. Verifiquemos entonces que  $P$  es simétrica:  $P^\top = (QQ^\top)^\top = (Q^\top)^\top Q^\top = QQ^\top = P$ .

### 5.3.2. Caracterización Geométrica

El operador  $P$  descompone cualquier vector  $w \in \mathbb{R}^n$  en una componente dentro del subespacio y un error ortogonal:  $w = Pw + (\mathbb{I} - P)w$ .

Para cualquier vector  $w$ , el “vector de error” definido como  $e = w - Pw$  es perpendicular a toda combinación lineal de las columnas de  $Q$ , es decir  $w - Pw \perp C(Q)$ .

Un vector es perpendicular a  $C(Q)$  si su producto interno con cada columna de  $Q$  es nulo, lo cual equivale a decir que  $Q^\top e = 0$ .

$$Q^\top e = Q^\top(w - Pw) = Q^\top(w - QQ^\top w) = Q^\top w - (Q^\top Q)Q^\top w = 0$$

### 5.3.3. Aplicación: La Proyección como Mejor Aproximación

La proyección ortogonal es la piedra angular del método de **Mínimos Cuadrados**. Cuando un sistema  $Ax = b$  es incompatible, es decir,  $b \notin C(A)$ , la solución numérica óptima no busca resolver la igualdad, sino minimizar el tamaño del error:  $\|b - Ax\|$ . En otras palabras, buscamos el vector  $\hat{x}$  en  $C(A)$  más cercano a  $b$  en términos de distancia euclídea.

Si las columnas de  $A$  se transforman en una base ortonormal  $Q$ , la “mejor aproximación” de  $b$  en el subespacio es simplemente su proyección ortogonal  $Pb = QQ^\top b$ .

Pero estamos interesados en encontrar una solución para cualquier  $A$ . Como vimos antes, para que el error sea mínimo, el vector de error  $(b - A\hat{x})$  debe ser perpendicular a todas las columnas de  $A$ . En lenguaje matricial, esto significa que el producto de  $A^\top$  por el error debe ser cero:

$$A^\top(b - A\hat{x}) = 0$$

Si distribuimos el producto, obtenemos las famosas **Ecuaciones Normales**:

$$A^\top b - A^\top A\hat{x} = 0 \tag{5.1}$$

$$A^\top A\hat{x} = A^\top b \tag{5.2}$$

Ahora, para encontrar  $\hat{x}$ , multiplicamos por la inversa de  $(A^\top A)$  (asumiendo que las columnas de  $A$  son linealmente independientes):

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

Sabemos que la proyección de  $b$  sobre el espacio de  $A$  es, por definición,  $Pb$ . Pero también sabemos que esa proyección es el resultado de aplicar la matriz  $A$  a nuestra mejor aproximación  $\hat{x}$ :

$$Pb = A\hat{x}$$

Sustituimos el valor de  $\hat{x}$  que acabamos de despejar:

$$Pb = A(A^\top A)^{-1} A^\top b$$

Si quitamos la  $b$  de ambos lados, nos queda la estructura de la matriz de proyección:

$$P = A(A^\top A)^{-1} A^\top$$

## 6 Matrices de Householder

§

§

### 6.1. Proyector $vv^\top$

Sea  $v \in \mathbb{R}^n$ ,  $\|v\| = 1$ , la expresión  $vv^\top$  produce una matriz de rango 1. Nos preguntamos, que hace esta matriz, cuando se la aplico a otro vector  $u$ . Es decir, quien es  $vv^\top u$  ?

Aplicando asociatividad, podemos ver que  $vv^\top u = v(v^\top u) = v\langle v, u \rangle$ , es decir que el resultado, es un vector en la dirección de  $v$ , y de módulo  $\|u\|\cos\theta$ . Es decir, es la proyección de  $u$  en la dirección de  $v$ :

$$u_v = vv^\top u$$

Por comodidad, normalmente tomamos  $v$  unitario. Si  $v$  no fuese unitario, el proyector correcto sería  $\frac{vv^\top}{v^\top v}$  (hay que normalizarlo). El concepto de “quitar una dimensión” para colapsar sobre la dirección de  $v$  sigue siendo el mismo.

**La expresión  $vv^\top$  es un proyector sobre la dirección de  $v$**

### 6.2. Proyección sobre $v^\perp$

Si restamos a  $u$  la componente de  $u$  en la dirección de  $v$ , esto “colapsa” el vector  $u$  sobre el plano ortogonal a  $v$ .

$$u_{v^\perp} = (\mathbb{I} - vv^\top)u$$

### 6.3. Reflexión

Si en cambio restamos la componente de  $u$  en la dirección de  $v$  **dos veces**, obligamos al vector a cruzar el eje y quedar exactamente a la misma distancia en el lado opuesto. A esta transformación se le llama **Reflexión de Householder**.

### 6.3.1. Definición

Dado un vector  $v \in \mathbb{R}^n$  unitario ( $\|v\| = 1$ ), la matriz de Householder se define como:

$$H = \mathbb{I} - 2vv^\top$$

La matriz  $H$  actúa como un operador de reflexión:

- Si un vector  $x$  es ortogonal a  $v$ , entonces  $Hx = x$  (el vector reside en el hiperplano de reflexión).
- Si un vector es paralelo a  $v$ , entonces  $Hv = -v$  (el vector se refleja completamente).

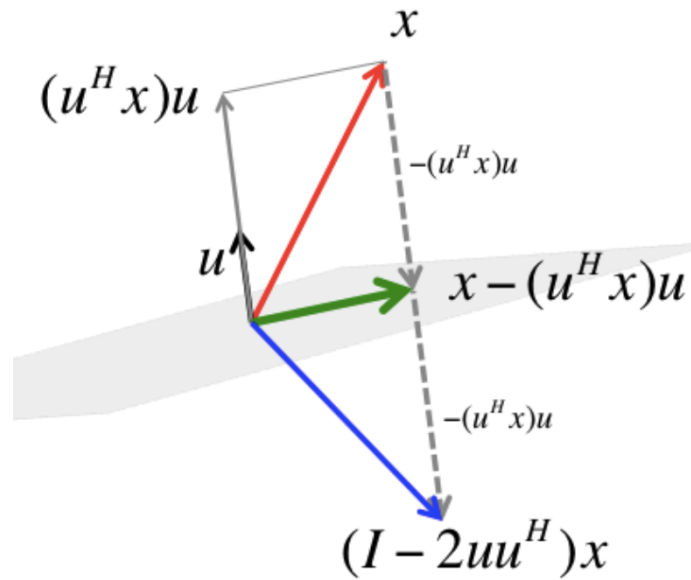


Figura 6.1: Reflexión  $Hx$  de  $x$  a través de  $u^\perp$

Créditos: Robert van de Geijn y Margaret Myers. Fuente: [LAFF-ALAF](#)

### 6.3.2. Propiedades Estructurales

La potencia de Householder en el cálculo numérico reside en sus propiedades estructurales:

**Simetría** La matriz es idéntica a su traspuesta.  $H^\top = (I - 2vv^\top)^\top = \mathbb{I}^\top - 2(v^\top)^\top v^\top = \mathbb{I} - 2vv^\top = H$

**Involución** Aplicar la reflexión dos veces regresa el espacio a su estado original:  $H^2 = (\mathbb{I} - 2vv^\top)(\mathbb{I} - 2vv^\top) = \mathbb{I} - 4vv^\top + 4v(v^\top v)v^\top$  Como  $v^\top v = 1$ , la expresión se simplifica a  $\mathbb{I} - 4vv^\top + 4vv^\top = \mathbb{I}$ . Por lo tanto:  $H^2 = \mathbb{I}$ .

**Ortogonal** Sabemos que es simétrica, por lo que  $H^\top = H$  y sabemos que es involutiva, por lo que  $H^2 = I$  entonces  $HH = H^\top H = I$  por lo tanto es ortogonal, y  $H^{-1} = H^\top$ .

**Isometría** Como es ortogonal, preserva la norma del vector original y nunca amplifica el error numérico.  $\|Hu\| = \|u\|$ .

### 6.3.3. Espectro y Autovectores

La matriz de Householder tiene un espectro extremadamente simple, compuesto únicamente por los valores propios  $\lambda_1 = 1$  y  $\lambda_2 = -1$ .

**Autovalor**  $\lambda_2 = -1$  El vector unitario  $v$  que define la reflexión es el autovector asociado a  $-1$ . Esto se debe a que la matriz actúa invirtiendo su dirección:

$$Hv = (I - 2vv^\top)v = v - 2v = -v$$

**Autovalor**  $\lambda_1 = 1$  El hiperplano ortogonal a  $v$ , denotado como  $v^\perp$  es el espacio propio asociado al autovalor  $\lambda_1 = 1$ . Cualquier vector  $w$  contenido en este plano satisface  $Hw = w$ , lo que significa que el valor propio 1 tiene una multiplicidad de  $n - 1$ .

Desde una perspectiva geométrica, esto confirma que la matriz es una reflexión: los vectores sobre el “espejo”  $v^\perp$  no cambian, mientras que el vector normal al espejo ( $v$ ) se refleja exactamente al lado opuesto.

Como consecuencia de este espectro, el determinante de cualquier matriz de Householder es siempre  $-1$ .

### 6.3.4. Utilidad en Factorizaciones

En la práctica, utilizamos Householder para “limpiar” columnas de una matriz. Si queremos que un vector  $x$  se convierta en un vector con ceros en todas las posiciones excepto la primera (proyectarlo sobre el eje  $e_1$ ), diseñamos un espejo que esté justo a mitad de camino entre  $x$  y el eje deseado. La “limpieza” de una columna mediante una reflexión de Householder consiste en diseñar un hiperplano (espejo) que sea el bisector perpendicular del camino entre el vector original  $x$  y el objetivo deseado  $z$  (Armentano 2026 clase 6; Strang 2018 Lec. 3, '70).

Generalmente se busca que toda la norma del vector  $x$  se concentre en la primera componente para crear ceros debajo. Por tanto, el objetivo es  $z = \|x\|e_1$ . El vector perpendicular al espejo debe ser la línea que conecta ambos puntos:  $v = x - z$ . Este vector  $v$  apunta directamente desde el origen del vector hacia su imagen reflejada. Para asegurar que la matriz sea ortogonal, se utiliza el vector unitario  $u = \frac{v}{\|v\|}$ . La matriz resultante  $H = I - 2uu^\top$  funciona restando exactamente el doble de la proyección de  $x$  sobre el vector normal  $u$ . Matemáticamente:

$$Hx = (I - 2uu^\top)x = x - 2u(u^\top x)$$

Al aplicar  $Hx$ , la matriz resta exactamente el doble de la proyección de  $x$  sobre el vector normal  $u$ . Esto cancela toda la parte de  $x$  que “sobra” fuera del eje  $e$  y lo empuja a aterrizar

exactamente en el objetivo  $z$ , dejando la columna “limpia” con ceros debajo de la primera posición

# 7 Autovalores, Autovectores, Factorización Espectral

El estudio de los autovalores y autovectores permite entender la acción de una matriz  $A$  no a través de sus entradas individuales, sino a través de sus direcciones invariantes. Mientras que la mayoría de los vectores cambian de dirección al ser multiplicados por  $A$ , los autovectores permanecen en la misma línea, experimentando únicamente un escalamiento.

## 7.1. La Ecuación Fundamental

Sea  $A \in \mathbb{R}^{n \times n}$  una matriz cuadrada. Un vector no nulo  $x$  es un **autovector** (o eigenvector) de  $A$  si existe un escalar  $\lambda$ , denominado **autovalor** (o eigenvalor), tal que:

$$Ax = \lambda x \tag{7.1}$$

Geoméricamente, la aplicación de  $A$  sobre  $x$  no altera su dirección, solo su magnitud por un factor  $\lambda$ .

Esta definición es exclusiva para matrices cuadradas porque el transformado de  $x$ , es decir,  $Ax$  está en el mismo subespacio vectorial que  $x$ .

### Nota

La definición de autovector requiere que  $x \neq \vec{0}$ . Sin embargo, el autovalor 0 si es permitido. De hecho, cualquier vector en el núcleo es un autovector con  $\lambda = 0$ .

### Tip

Si la matriz es invertible y  $\lambda \neq 0$ , se verifica que  $x$  también es autovector de  $A^{-1}$  con autovalor  $1/\lambda$ . En este caso,  $\lambda$  necesariamente tiene que ser distinto de 0 para que  $1/\lambda$  exista, pero el caso  $\lambda = 0$  tiene sentido igualmente.  $\lambda = 0$  significa que  $Ax = 0$  es decir, que  $x \in N(A)$ , lo cual a su vez implica que el rango de  $A$  es menor que  $n$  y que  $A$  no es invertible, por lo que la expresión  $A^{-1}x = \frac{1}{\lambda}x$  no es válida tampoco del lado izquierdo.

## 7.2. Motivación: Convergencia de Algoritmos Numéricos

Una de las mayores utilidades de los autovectores es el cálculo de potencias de matrices. Si aplicamos  $A$  sucesivamente  $k$  veces sobre un autovector  $x$ :

$$A^k x = \lambda^k x$$

Esto implica que el comportamiento asintótico de  $A^k$  está gobernado por la magnitud de sus autovalores:

- Si  $|\lambda| > 1$ , la componente en esa dirección crece exponencialmente.
- Si  $|\lambda| < 1$ , la componente decae o se atenúa hacia cero.

Para la mayoría de los vectores, la multiplicación por una matriz  $A$  cambia tanto su longitud como su dirección. Sin embargo, los **autovectores** son direcciones especiales que permanecen invariantes: el producto  $Ax$  se mantiene en la misma línea que  $x$ , escalado únicamente por el **autovalor**  $\lambda$ .

Esta propiedad se vuelve importante cuando calculamos potencias de una matriz. Si aplicamos  $A$  sucesivamente  $k$  veces sobre un autovector, el resultado es simplemente  $A^k x = \lambda^k x$ .

La verdadera utilidad aparece cuando una matriz  $n \times n$  posee  $n$  autovectores linealmente independientes  $\{x_1, x_2, \dots, x_n\}$ . En este escenario, estos vectores forman una base completa del espacio  $\mathbb{R}^n$ , por lo que cualquier vector arbitrario  $v$  puede expresarse como una combinación lineal de estos autovectores (Strang 2018 Lec. 4 '10):

$$v = c_1 x_1 + c_2 x_2 + \dots + c_n x_n$$

donde los coeficientes  $c_i$  representan las coordenadas de  $v$  en esta “base natural” de la matriz.

Calcular  $A^k v$  mediante multiplicaciones matriciales directas es computacionalmente costoso y numéricamente inestable. Sin embargo, al utilizar la base de autovectores, el cálculo se vuelve inmediato. Gracias a la linealidad, la acción de  $A^k$  se distribuye sobre cada componente:

$$A^k v = c_1 (\lambda_1^k x_1) + c_2 (\lambda_2^k x_2) + \dots + c_n (\lambda_n^k x_n)$$

Bajo esta óptica:

1. **Descomponemos  $v$ :** Encontramos cuánta “magnitud” tiene  $v$  en cada dirección invariante.
2. **Evolucionamos el sistema:** Cada dirección evoluciona de forma independiente según la magnitud de su autovalor  $\lambda_i^k$ .
3. **Resultado Asintótico:** Si  $|\lambda_i| < 1$ , esa componente desaparecerá con el tiempo; si  $|\lambda_i| > 1$ , esa dirección dominará el crecimiento del sistema.

### 7.3. El Polinomio Característico

Para hallar los autovalores, reescribimos la Ecuación 7.1 como un sistema homogéneo singular  $(A - \lambda \mathbb{I})x = 0$ . Para que exista una solución no trivial, la matriz  $(A - \lambda \mathbb{I})$  debe ser singular, lo que conduce al **polinomio característico**:

$$\chi_A(\lambda) = \det(A - \lambda \mathbb{I}) = 0$$

Las raíces de este polinomio de grado  $n$  constituyen el **espectro** de la matriz.

### 7.4. Espectro de la Traspuesta $A^\top$

Toda matriz cuadrada  $A$  comparte exactamente los mismos autovalores con su traspuesta  $A^\top$ , porque ambas generan el mismo polinomio característico (Strang 2019, I.6; Strang 2018 Lec. 4 '131; Armentano 2026 clase 7).

Sin embargo, aunque los autovalores son los mismos, los **autovectores** suelen ser diferentes. Los autovectores de  $A^\top$  se conocen como **autovectores izquierdos** de  $A$ , ya que satisfacen  $y^\top A = \lambda y^\top$ . En el “Gran Mapa”, esto significa que mientras los autovectores de  $A$  definen direcciones invariantes en el dominio, los de  $A^\top$  definen direcciones invariantes respecto a las filas (Armentano 2026 clase 4).

### 7.5. Espectro de Matrices Semejantes

Dos matrices  $A$  y  $B$  son **semejantes** si existe una matriz invertible  $X$  tal que  $A = XBX^{-1}$ .

Matrices semejantes comparten el mismo espectro de autovalores (pero no comparten autovectores).  $Av = \lambda v \Rightarrow X^{-1}BXv = \lambda v \Rightarrow BXv = X\lambda v \Rightarrow B(Xv) = \lambda(Xv)$ . Esta última igualdad implica que  $Xv$  es vector propio de  $B$  con valor propio  $\lambda$ .

**Prueba** Sean dos matrices  $A$  y  $B$  semejantes ( $\exists M/B = M^{-1}AM$ ). La semejanza preserva el polinomio característico, incluso si los autovalores  $\lambda$  residen en el plano complejo  $\mathbb{C}$  (Strang 2018 Lec. 4 '125):

$$\det(B - \lambda \mathbb{I}) = \det(M^{-1}AM - \lambda M^{-1}M) = \det(M^{-1}(A - \lambda \mathbb{I})M)$$

Utilizando la propiedad distributiva del determinante ( $\det(ABC) = \det(A)\det(B)\det(C)$ ):

$$\det(M^{-1})\det(A - \lambda \mathbb{I})\det(M) = \frac{1}{\det(M)}\det(A - \lambda \mathbb{I})\det(M) = \det(A - \lambda \mathbb{I})$$

## 7.6. Espectro del Producto $AB$

El producto de matrices  $AB$  tiene los mismos valores propios (no nulos) que  $BA$ . Para probarlo usando la observación anterior, deberíamos probar que  $AB$  y  $BA$  son semejantes.

Esto equivale a buscar una matriz  $X$  tal que  $X(AB)X^{-1} = BA$ . Esta igualdad se satisface simplemente tomando  $X = B$ , ya que  $B(AB)B^{-1} = BA(BB^{-1}) = BA\mathbb{1} = BA$

## 7.7. Espectro de $A + k\mathbb{1}$

Si  $v$  es autovector de  $A$  con autovalor asociado  $\lambda$ , entonces  $Av = \lambda v$ . Veamos que pasa si aplicamos la matriz  $A + k\mathbb{1}$  a  $v$ :

$$(A + k\mathbb{1})v = Av + kv = \lambda v + kv = (\lambda + k)v$$

Esto significa que  $v$  es autovector de  $A + k\mathbb{1}$  con autovalor  $\lambda + k$ . (Hernández 2026, Pr3 Ej15)

### Tip

Sumar  $k\mathbb{1}$  a una matriz, desplaza todos sus autovalores por  $k$  pero mantiene intacta la dirección de sus autovectores

## 7.8. Vínculo entre Espectro, Determinante y Traza

### 7.8.1. Determinante

El producto de los autovalores es igual al determinante de la matriz:

$$\det(A) = \prod \lambda_i$$

La forma más sencilla de demostrar que el determinante es el producto de los autovalores es evaluando el polinomio característico en un punto específico (Hernández 2026 Practico 3 Ej 10). Por el Teorema Fundamental del Álgebra, el polinomio se puede expresar en términos de sus raíces (autovalores):

$$\chi_A(\lambda) = (\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$$

Si evaluamos la expresión anterior en  $\lambda = 0$  obtenemos:

$$\chi_A(0) = (\lambda_1 - 0)(\lambda_2 - 0) \dots (\lambda_n - 0) = \lambda_1 \lambda_2 \dots \lambda_n$$

Por otro lado, por la definición original en  $\lambda = 0$ :

$$\chi_A(0) = \det(A - 0\mathbb{1}) = \det(A)$$

## 💡 Espectro de Matrices Singulares

Esta propiedad implica que si una matriz es singular ( $\det(A) = 0$ ), al menos uno de sus autovalores es cero (Strang 2019, I.6). Recíprocamente, si una matriz tiene un autovalor 0, es singular.

### 7.8.2. Traza

La suma de las entradas de la diagonal principal es igual a la suma de los autovalores:

$$\text{tr}(A) = \sum \lambda_i$$

La relación con la traza surge al comparar los coeficientes del polinomio característico en sus dos formas de expresión (Strang 2018 - Lec. 4, '144).

- **En la forma factorizada:** Al expandir el producto  $(\lambda_1 - \lambda)(\lambda_2 - \lambda) \dots (\lambda_n - \lambda)$ , el coeficiente del término  $(-\lambda)^{n-1}$  es exactamente la suma de las raíces:  $\sum \lambda_i$ .
- **En el determinante:** Al expandir  $\det(A - \lambda I)$  mediante la fórmula de Leibniz o cofactores, el único término que contiene  $(-\lambda)^{n-1}$  proviene del producto de los elementos de la diagonal principal:

$$(a_{11} - \lambda)(a_{22} - \lambda) \dots (a_{nn} - \lambda)$$

En este producto, el coeficiente de  $(-\lambda)^{n-1}$  es la suma de los elementos de la diagonal:  $a_{11} + a_{22} + \dots + a_{nn}$  (Strang 2019, I.6).

Dado que ambos coeficientes deben ser iguales por pertenecer al mismo polinomio, se cumple que  $\text{tr}(A) = \sum_{i=1}^n \lambda_i$ .

### 7.8.3. Verificación en Matrices Diagonalizables

Si  $A = X\Lambda X^{-1}$ , podemos usar las propiedades de invarianza por semejanza (Strang 2019, I.6):

- $\det(A) = \det(X) \det(\Lambda) \det(X)^{-1} = \det(\Lambda) = \prod \lambda_i$ .
- $\text{tr}(A) = \text{tr}(X\Lambda X^{-1}) = \text{tr}(X^{-1}X\Lambda) = \text{tr}(\Lambda) = \sum \lambda_i$ .

## 7.9. Factorización $X\Lambda X^{-1}$

Si una matriz  $A$  posee  $n$  **autovectores linealmente independientes**, estos pueden agruparse como columnas de una matriz  $X$ . La relación  $AX = X\Lambda$  permite la **factorización por diagonalización** (Strang 2018 Lec. 4 '36):

$$A = X\Lambda X^{-1}$$

donde  $\Lambda$  es una matriz diagonal con los autovalores en su diagonal principal (Strang 2018 Lec 4 '13).

Bajo esta perspectiva, multiplicar por  $A$  equivale a:

1. Cambiar a la base de autovectores ( $X^{-1}$ ).
2. Escalar cada componente por su respectivo  $\lambda_i$  ( $\Lambda$ ).
3. Regresar a la base original ( $X$ ).

### Tip

Toda matriz diagonalizable, se puede pensar como una matriz diagonal, solo basta recordar que la base en que son diagonales es la base de vectores propios.

### Tip

La invertibilidad de una matriz depende exclusivamente de su espectro ( $\lambda_i \neq 0$ ). La diagonalizabilidad requiere de la existencia de  $n$  vectores propios linealmente independientes. (Hernández 2026, Pr3 Ej 16 y 18)

### 7.9.1. Aplicaciones

La factorización  $A = X\Lambda X^{-1}$  simplifica el cálculo de funciones de matrices:

**Potencias**  $A^k = X\Lambda^k X^{-1}$ . Si  $|\lambda_i| < 1$ , entonces  $A^k \rightarrow 0$  cuando  $k \rightarrow \infty$ .

**Inversa**  $A^{-1} = X\Lambda^{-1}X^{-1}$  (si  $\lambda_i \neq 0$ ).

**Método de la Potencia** Desde una perspectiva numérica, si deseamos encontrar la dirección del autovector dominante (el asociado al  $|\lambda|$  más grande), podemos iterar un vector aleatorio  $v_0$ :

$$v_{k+1} = \frac{Av_k}{\|Av_k\|}$$

A medida que  $k \rightarrow \infty$ , el vector  $v_k$  se alinea con el autovector dominante.

# 8 Matrices Ortogonales

§

§

## 8.0.1. Definición

Una matriz cuadrada  $Q \in \mathbb{R}^{n \times n}$  se define como ortogonal (o más precisamente, ortonormal) si sus columnas son ortonormales (vectores unitarios perpendiculares entre sí), lo que implica que: (Armentano 2026 clase 8; Strang 2019, I.5)

$$Q^T Q = \mathbb{I}$$

Cuando una matriz de Stiefel  $Q \in \mathbb{R}^{n \times n}$  es cuadrada, sus columnas no solo son ortonormales, sino que forman una **base ortonormal** completa para  $\mathbb{R}^n$ . En este caso, la matriz se denomina simplemente **matriz ortogonal**.

## 8.0.2. La Identidad Fundamental

La propiedad más potente de una matriz ortogonal es que su transpuesta es idéntica a su inversa.<sup>1</sup>

$$Q^{-1} = Q^T$$

**Prueba** Por definición de matriz de Stiefel,  $Q^T Q = \mathbb{I}$ . En el caso cuadrado, una matriz que posee una inversa a izquierda también la posee a derecha y es única. Por tanto,  $Q^T = Q^{-1}$ , lo que implica que también se verifica  $Q Q^T = \mathbb{I}$  (Diego Armentano 2025 Parcial 1 Ej1a).

---

<sup>1</sup>Esta característica simplifica drásticamente la resolución de sistemas lineales  $Qx = b$ . En lugar de recurrir a la eliminación gaussiana ( $O(n^3)$ ), la solución se obtiene mediante una simple transposición  $O(n^2)$ :  $x = Q^T b$ .

### 8.0.3. Preservación de la Geometría (Isometría)

La propiedad definitoria de las matrices ortogonales es que no alteran el producto escalar entre dos vectores cualesquiera  $u$  y  $v$ . Utilizando la propiedad de la transpuesta de un producto,  $(Qu)^\top = u^\top Q^\top$ , tenemos:

$$(Qu) \cdot (Qv) = (Qu)^\top (Qv) = u^\top (Q^\top Q)v = u^\top Iv = u \cdot v$$

De esta invariancia se derivan dos consecuencias geométricas críticas:

1. **Preservación de la Norma (Longitud):** Dado que  $\|v\| = \sqrt{v \cdot v}$ , se cumple que  $\|Qv\| = \|v\|$ . El vector transformado mantiene su magnitud original.
2. **Preservación de Ángulos:** El ángulo  $\theta$  entre dos vectores se define por  $\cos(\theta) = \frac{u \cdot v}{\|u\| \|v\|}$ . Al ser el numerador y el denominador invariantes bajo  $Q$ , el ángulo permanece constante.

En definitiva, las matrices ortogonales actúan como operadores que preservan la estructura métrica del espacio. No solo mantienen las longitudes (normas), sino también los ángulos entre vectores.<sup>2</sup>

Por lo tanto, las matrices ortogonales  $Q \in \mathbb{R}^{n \times n}$  representan una rotación (o una combinación de rotación y reflexión) porque actúan como una **isometría lineal** en el espacio euclídeo.

### 8.0.4. El Determinante y la Orientación

Aunque todas las matrices ortogonales preservan la estructura métrica, el **determinante** de la matriz  $Q$  determina si la transformación preserva o invierte la orientación del espacio:

El determinante de  $Q$ , que es el producto de sus valores propios, siempre es 1 (rotaciones puras) o  $-1$  (reflexión).

**Prueba** A partir de la identidad  $Q^\top Q = I$ , aplicamos la propiedad del determinante de un producto:  $\det(Q^\top) \det(Q) = \det(I)$ . Sabiendo que  $\det(Q^\top) = \det(Q)$  y que  $\det(I) = 1$ , obtenemos la relación  $\det(Q)^2 = 1$  por lo que  $\det(Q) = \pm 1$

| Transformación                       | Determinante   |   |
|--------------------------------------|----------------|---|
| <b>Rotación Propia</b>               | $\det(Q) = 1$  |   |
| <b>Rotación Impropia (Reflexión)</b> | $\det(Q) = -1$ | Invierte la orientación (quiralidad) del espacio. |

<sup>2</sup>Las transformaciones basadas en matrices ortogonales (como las rotaciones de Householder o Givens) son preferidas en aplicaciones donde la estabilidad numérica es crítica porque **nunca amplifican el error ni producen overflow**. Dado que  $\|Qx\| = \|x\|$ , el tamaño de los datos se mantiene constante a lo largo de múltiples factorizaciones.

## 8.1. Espectro

Los valores propios ( $\lambda$ ) de una matriz ortogonal siguen reglas geométricas estrictas derivadas de su capacidad para preservar longitudes.

### Módulo Unitario

Debido a que las matrices ortogonales preservan la norma de los vectores ( $\|Qx\| = \|x\|$ ). Si aplicamos esto a un autovector:

$$\|Qx\| = \|\lambda x\| = |\lambda| \|x\| \implies |\lambda| = 1$$

Por lo tanto, todos los autovalores de una matriz ortogonal deben tener magnitud 1; es decir, residen en el círculo unidad del plano complejo. (Strang 2018 Lec. 3 '62; Strang 2019, I.5)

### Valores Reales Restringidos

Como  $|\lambda| = 1$ , Los únicos valores propios reales que puede tener una matriz ortogonal son 1 o  $-1$ . Cualquier otro valor real violaría la condición de módulo uno.

### Simetría en el Plano Complejo

Como  $Q$  tiene entradas reales, sus autovalores complejos siempre aparecen en pares  $\lambda$  y  $\bar{\lambda}$  (Teorema Fundamental del Álgebra). Esto asegura que el polinomio característico mantenga coeficientes reales. Esto puede verse algebraicamente: Si  $\lambda$  es un valor propio complejo de  $Q$  con vector propio  $v$ , entonces:  $Q\bar{v} = \overline{Qv} = \overline{\lambda v} = \bar{\lambda} \bar{v}$ . Por lo tanto,  $\bar{\lambda}$  también es valor propio con vector propio  $\bar{v}$ .

### Espectro de $Q^\top = Q^{-1}$

Como se demostró en Sección 7.4,  $Q$  y  $Q^\top$  siempre comparten los mismos autovalores. Sin embargo, en el caso de las matrices ortogonales, existe una relación adicional:  $Q^\top = Q^{-1}$ . Esto implica que: **los autovalores de  $Q^\top$  son también los autovalores del inverso de  $Q$ .**

Si  $x$  es un autovector de  $Q$  con autovalor  $\lambda$ , entonces (Strang 2018 Lec. 4 '115):

$$Qx = \lambda x \implies x = Q^{-1}(\lambda x) \implies Q^{-1}x = \frac{1}{\lambda}x$$

Por lo tanto: **si  $\lambda$  es un autovalor de  $Q$ , su recíproco  $1/\lambda$  es un autovalor de  $Q^{-1}$  (y por ende de  $Q^\top$ ).**

### **i** Nota

Para cualquier número complejo  $\lambda$  con  $|\lambda| = 1$ , se cumple que su recíproco es igual a su conjugado:

$$\frac{1}{\lambda} = \bar{\lambda}$$

Los autovalores de  $Q^\top$  son los recíprocos de los de  $Q$ , pero como tienen magnitud 1, los autovalores de la traspuesta son simplemente los **complejos conjugados** de los autovalores originales (Strang 2019, I.5).

## 8.2. Autovectores

Al igual que las matrices simétricas, las matrices ortogonales (y en general las matrices normales) poseen vectores propios que son ortogonales entre sí.

**Prueba** Sean  $v$  y  $w$  autovectores de  $Q$  con autovalores asociados  $\lambda_1$  y  $\lambda_2$  respectivamente.

La definición de autovector nos dice que  $Qv = \lambda_1 v$  y  $Qw = \lambda_2 w$ . Trasponiendo la primera y multiplicando por la segunda obtenemos:

$$v^\top Q^\top Qw = \lambda_1 v^\top \lambda_2 w$$

Lo que implica  $\langle v, w \rangle = \lambda_1 \lambda_2 \langle v, w \rangle$  y de aquí se deduce:

$$(1 - \lambda_1 \lambda_2) \langle v, w \rangle = 0$$

Los autovalores de las matrices ortogonales tienen módulo 1, pero no necesariamente satisfacen  $1 = \lambda_1 \lambda_2$ . Como esta igualdad debe cumplirse para matrices con cualquier autovalor, entonces necesariamente  $\langle v, w \rangle = 0$  por lo que:

$$v \perp w$$

## 8.3. Ejemplos

**Matrices de Permutación ( $P$ )** Reordenan las entradas de un vector. Sus columnas son las de la identidad en distinto orden. Siempre verifican  $P^\top = P^{-1}$ .

**Matrices de Reflexión** En  $\mathbb{R}^2$ , la matriz  $Q = \begin{pmatrix} \cos \theta & \sin \theta \\ \sin \theta & -\cos \theta \end{pmatrix}$  refleja los vectores respecto de la recta que pasa por el origen y tiene un ángulo  $\theta$  con el eje horizontal.

**Matriz de Fourier ( $F$ )** Utilizada en el procesamiento de señales, sus columnas son autovectores ortonormales que permiten transitar entre el dominio del tiempo y la frecuencia.

**Matrices Simétricas y Ortogonales** Si una matriz  $S$  es simétrica ( $S = S^\top$ ) y a la vez ortogonal ( $S^{-1} = S^\top$ ), entonces  $S^2 = SS = SS^\top = SS^{-1} = I$ .

En este caso, si  $v$  es un autovector con autovalor asociado  $\lambda$ , al aplicar  $S^2$ :

$$S^2v = S(Sv) = S\lambda v = \lambda Sv = \lambda^2v$$

Pero habíamos visto que  $S^2 = \mathbb{1}$ , por lo que  $\lambda^2v = v$ . Los únicos autovalores que satisfacen esta ecuación son 0, 1 y  $-1$ . Pero además,  $S$  es invertible, por lo que 0 no es un autovalor posible. En definitiva,  $\lambda = \pm 1$ .

Esto nos da una expresión concisa para las posibles matrices de autovalores: (Hernández 2026, Pr4 Ej5)

$$\Lambda = \begin{bmatrix} \mathbb{1} & 0 \\ 0 & -\mathbb{1} \end{bmatrix}$$

**Matrices de Rotación ( $R_\theta$ )** La matriz de rotación estándar para un ángulo  $\theta$  en sentido antihorario es el ejemplo más claro de una matriz ortogonal (Hernández 2026, P3, ejercicio 7):

$$R_\theta = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$$

Podemos verificar su ortogonalidad mediante la suma de los cuadrados de sus componentes en las columnas ( $\sin^2 \theta + \cos^2 \theta = 1$ ) y comprobando que su determinante es siempre 1, lo que garantiza que es una rotación pura sin reflexión.

Una matriz de rotación pura (con  $\theta \neq 0, \pi$ ) en  $\mathbb{R}^2$  típicamente no tiene autovectores reales porque gira todos los vectores un ángulo  $\theta$ . Sin embargo, posee autovalores complejos de la forma  $\lambda = e^{\pm i\theta} = \cos \theta \pm i \sin \theta$ , que cumplen con tener magnitud 1.  $|e^{\pm i\theta}| = 1$ . Solo en los casos donde  $\theta = 0$  ( $\lambda = 1$ ) o  $\theta = \pi$  ( $\lambda = -1$ ) los vectores propios vuelven al espacio real.

Los autovectores asociados son vectores complejos en  $\mathbb{C}^2$ :

- Para  $\lambda_1 = e^{i\theta}$  el autovector es  $x_1 = (1, -i)$ .
- Para  $\lambda_2 = e^{-i\theta}$  el autovector es  $x_2 = (1, i)$ .

Podemos validar estos resultados mediante la traza y el determinante:

- **Traza:**  $\lambda_1 + \lambda_2 = (\cos \theta + i \sin \theta) + (\cos \theta - i \sin \theta) = 2 \cos \theta$ , que coincide exactamente con la suma de la diagonal de  $Q$ .
- **Determinante:**  $\lambda_1 \cdot \lambda_2 = e^{i\theta} \cdot e^{-i\theta} = e^0 = 1$ , consistente con una rotación que preserva la orientación.

**i** Ortogonalidad Compleja

Aunque estos vectores viven en  $\mathbb{C}^2$ , siguen siendo ortogonales entre sí bajo el producto interno complejo  $x_1^* x_2 = 0$ .

# 9 Matrices Simétricas

§

§

Las matrices simétricas  $S$  se definen por la propiedad fundamental de ser iguales a su propia traspuesta. Gilbert Strang las describe como los “reyes del álgebra lineal” debido a que poseen propiedades matemáticas excepcionales que las hacen ideales para una vasta gama de aplicaciones en ciencia e ingeniería. Desde una perspectiva práctica, estas matrices son la base para el estudio de las formas cuadráticas y la optimización. En campos como el aprendizaje profundo y la ciencia de datos, se utilizan para representar funciones de pérdida que deben ser minimizadas, donde la naturaleza de la matriz (si es definida positiva) determina si el sistema tiene un mínimo global estable. Además, son esenciales en estadística para construir matrices de covarianza, las cuales describen la relación y varianza entre múltiples conjuntos de datos.

## 9.1. Definición

Decimos que  $S \in \mathcal{M}_{n \times n}$  es **Simétrica** si  $S = S^T$

## 9.2. Autovalores Reales

Sea  $S \in \mathbb{R}^{n \times n}$  una matriz simétrica. Sea  $\lambda$  un valor propio de  $S$  y  $x$  su vector propio asociado no nulo. Entonces:

**!** Importante

Los autovalores de  $S$  son todos reales:

$$\lambda_i \in \mathbb{R} \quad \forall i = 1 \dots n$$

**Prueba** Aunque  $S$  es real, inicialmente permitimos que  $\lambda \in \mathbb{C}$  y  $x \in \mathbb{C}^n$ .

Partimos de la ecuación fundamental:  $Sx = \lambda x$ . Si tomamos el conjugado complejo, considerando que  $S$  es una matriz real ( $\overline{S} = S$ ):

$$\overline{Sx} = \overline{\lambda x} \implies S \overline{x} = \overline{\lambda} \overline{x}$$

Ahora, aplicamos la transpuesta a esta ecuación conjugada

$$(S \overline{x})^\top = (\overline{\lambda} \overline{x})^\top \implies \overline{x}^\top S^\top = \overline{\lambda} \overline{x}^\top$$

Dado que por hipótesis la matriz es simétrica ( $S^\top = S$ ), la expresión queda:  $\overline{x}^\top S = \overline{\lambda} \overline{x}^\top$

Multiplicamos la expresión resultante por el vector propio original  $x$  a la derecha

$$(\overline{x}^\top S) x = (\overline{\lambda} \overline{x}^\top) x \implies \overline{x}^\top (Sx) = \overline{\lambda} (\overline{x}^\top x)$$

Sustituimos la definición original  $Sx = \lambda x$  en el lado izquierdo

$$\overline{x}^\top (\lambda x) = \overline{\lambda} (\overline{x}^\top x) \implies \lambda (\overline{x}^\top x) = \overline{\lambda} (\overline{x}^\top x)$$

La expresión  $\overline{x}^\top x$  representa el producto interno de  $x$  consigo mismo en  $\mathbb{C}^n$ , lo cual equivale a la norma al cuadrado del vector:  $\overline{x}^\top x = \|x\|^2$ . Como un vector propio  $x$  es, por definición, distinto del vector nulo ( $x \neq 0$ ), entonces  $\|x\|^2 > 0$  y entonces podemos dividir ambos lados por esta cantidad. Finalmente obtenemos:

$$\lambda = \overline{\lambda}$$

En el cuerpo complejo, un número que es igual a su conjugado es, por definición, un número real.

### 9.3. Ortogonalidad de Autovectores

Sea  $S \in \mathbb{R}^{n \times n}$  una matriz simétrica. Sea  $\lambda$  un valor propio de  $S$  y  $x$  su vector propio asociado no nulo. Entonces:

#### ! Importante

Los autovectores correspondientes a autovalores distintos son ortogonales entre si:

$$\left. \begin{array}{l} Sq_1 = \lambda_1 q_1 \\ Sq_2 = \lambda_2 q_2 \\ \lambda_1 \neq \lambda_2 \end{array} \right\} \implies q_1 \perp q_2$$

**Prueba** Sean  $q_1, q_2$  autovectores de  $S$  con  $\lambda_1 \neq \lambda_2$ .

$$Sq_1 = \lambda_1 q_1 \implies (Sq_1)^\top = \lambda_1 q_1^\top \implies q_1^\top S = \lambda_1 q_1^\top$$

$$\text{Multiplicando por } q_2 \text{ a la derecha: } q_1^\top S q_2 = \lambda_1 q_1^\top q_2$$

$$\text{Sustituyendo } S q_2 = \lambda_2 q_2: \lambda_2 q_1^\top q_2 = \lambda_1 q_1^\top q_2 \implies (\lambda_2 - \lambda_1) q_1^\top q_2 = 0$$

Dado que  $\lambda_1 \neq \lambda_2$ , necesariamente  $q_1^\top q_2 = 0$ . Por lo tanto, los autovectores son ortogonales.

## 9.4. Teorema Espectral para Matrices Simétricas

Sea  $S \in \mathcal{M}_{n \times n} : S = S^\top$ , existe una base ortonormal de autovectores, lo que permite una diagonalización mediante una **matriz ortogonal**  $Q$ :

$$S = Q \Lambda Q^\top = \sum_{i=1}^n \lambda_i q_i q_i^\top$$

## 9.5. Factorización Espectral

(Strang 2018 Lec 2 '4)

La descomposición  $Q \Lambda Q^\top$  es una variante específica de la diagonalización aplicada exclusivamente a **matrices simétricas**. Para una matriz cuadrada  $S \in \mathbb{R}^{n \times n}$  que es simétrica (es decir,  $S = S^\top$ ), el Teorema Espectral (Sección 9.4) establece que  $S$  es ortogonalmente diagonalizable. La expresión se define formalmente como:

$$S = Q \Lambda Q^\top$$

## 9.6. Componentes de la Factorización

**La Matriz de Autovalores**  $\Lambda$  es una **matriz diagonal** (todos los elementos fuera de la diagonal principal son nulos) que contiene los autovalores  $\lambda_1, \lambda_2, \dots, \lambda_n$  correspondientes a cada autovector en  $Q$ .

$$\Lambda = \begin{pmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{pmatrix}$$

**La Matriz Ortogonal**  $Q$  es una **matriz ortogonal** cuyas columnas son los autovectores de  $S$ . Al ser ortogonal, cumple la propiedad  $Q^\top = Q^{-1}$ , lo que implica que  $Q Q^\top = I$ .

Las columnas de  $Q$  forman una **base ortonormal** del espacio euclídeo. Esto implica que cada columna tiene norma unitaria ( $\|q_i\| = 1$ ) y es ortogonal a las demás ( $q_i \cdot q_j = 0$  para  $i \neq j$ ). Geométricamente,  $Q$  opera como una **isometría** (rotación o reflexión).

## 9.7. Interpretación Geométrica

Si visualizamos la matriz  $A$  como una transformación lineal, la descomposición  $Q\Lambda Q^T$  desglosa la operación en tres etapas cinemáticas:

---

|           |  |
|-----------|--|
| $Q^T$     | Realiza un cambio de base del vector de entrada al sistema de coordenadas definido por los autovectores (rota el espacio).   |
| $\Lambda$ | Aplica un escalamiento (dilatación o contracción) en cada una de estas nuevas direcciones según los valores $\lambda_i$ . Representa el factor de escala de la transformación lineal a lo largo de sus ejes principales. |
| $Q$       | Devuelve el vector al sistema de coordenadas original mediante la rotación inversa.  |

---

## 9.8. Interpretación como Suma de Rango 1

La identidad  $S = Q\Lambda Q^T$  puede expandirse utilizando la multiplicación columna-fila como una suma de matrices de **rango 1**:

$$S = \lambda_1 q_1 q_1^T + \lambda_2 q_2 q_2^T + \dots + \lambda_n q_n q_n^T$$

Cada término  $\lambda_i q_i q_i^T$  representa la proyección del espacio sobre la dirección del autovector  $q_i$ , escalada por  $\lambda_i$ .<sup>1</sup>

---

<sup>1</sup>En aplicaciones de ciencia de datos, esta expansión permite identificar las componentes que capturan la mayor varianza del sistema, descartando aquellas con autovalores despreciables.

# 10 Matrices Definidas y Semidefinidas Positivas

Las matrices definidas positivas representan una clase especial de matrices simétricas que desempeñan un papel central en la optimización, la estadística y el análisis de estabilidad de sistemas. Su estudio permite pasar de una visión puramente lineal a una comprensión de superficies cuadráticas y “energía” del sistema. (Armentano 2026 Clase 9)

## 10.1. Definición

Se le denomina **Energía** o **Forma Cuadrática** de la matriz  $S$  al número resultante de la operación  $x^\top Sx$ .

## 10.2. Definición

Sea  $S \in \mathbb{R}^{n \times n}$  una matriz **simétrica** ( $S = S^\top$ ).

Decimos que  $S$  es una **matriz definida positiva** si para todo vector no nulo  $x \in \mathbb{R}^n$ , su forma cuadrática es estrictamente mayor que cero:

$$x^\top Sx > 0 \quad \forall x \neq 0$$

Dependiendo del comportamiento de la operación  $x^\top Sx$ , clasificamos la matriz  $S$  en:

|                       |               |  |
|-----------------------|---------------|--|
| Definida Positiva     | $S \succ 0$   | Si $x^\top Sx > 0$ para todo $x \neq 0$ <sup>1</sup>   |
| Semidefinida Positiva | $S \succeq 0$ | Si $x^\top Sx \geq 0$ para todo $x$ Esto permite que existan vectores no nulos (normalmente en el núcleo de la matriz) para los cuales la forma cuadrática sea cero. |
| Indefinida            |               | $\exists x, y / x^\top Sx > 0, y^\top Sy < 0$  |

<sup>1</sup>Geoméricamente, la expresión  $f(x) = x^\top Sx$  define una superficie en  $\mathbb{R}^n$ . Si  $S \succ 0$ , la gráfica de  $f(x)$  se asemeja a un paraboloide que abre hacia arriba (bowl), con un mínimo global único en el origen  $x = 0$ .

### 10.3. Criterios de verificación

Para verificar si una matriz simétrica  $S$  es definida positiva, existen diversos criterios equivalentes que vinculan la forma cuadrática con el espectro de la matriz y sus factorizaciones. (Strang 2018 Lec 5)

#### 10.3.1. 1: Autovalores Positivos

Una matriz simétrica  $S$  es definida positiva si y solo si todos sus autovalores son estrictamente positivos:

$$\lambda_i > 0 \quad \text{para todo } i = 1, \dots, n$$

El teorema espectral nos dice que existe una base de vectores propios que diagonaliza a  $S$  como  $S = Q\Lambda Q^\top$ . La forma cuadrática se puede escribir entonces como  $x^\top(Q\Lambda Q^\top)x$ . Definiendo  $y = Q^\top x$  la expresión se simplifica a  $S = y^\top \Lambda y = \sum \lambda_i y_i^2$ , donde es fácil ver que si todos los  $\lambda_i > 0$ , la suma es necesariamente positiva para cualquier  $y \neq 0$ . Además, como  $Q$  es ortogonal, sus columnas son l.i., por lo que su núcleo es vacío y por lo tanto  $y \neq 0$

En el caso de una matriz  $S$  **Semidefinida Positiva** se permite que algunos autovalores sean cero ( $\lambda_i \geq 0$ ), lo que implica que la matriz puede ser singular.

#### 10.3.2. 2: Energía Positiva

Strang usa el concepto de Energía positiva, que no es más que la definición que usamos aquí.

#### 10.3.3. 3: Factorización $S = A^\top A$

( $\implies$ ) Si  $S \succ 0$ , existe una matriz invertible  $A$  tal que  $S = A^\top A$ <sup>2</sup>.

**Prueba** Como la matriz  $S$  es simétrica, el teorema espectral nos dice que podemos escribirla como  $S = Q\Lambda Q^\top$ . Tomamos  $A = \sqrt{\Lambda}Q^\top$ , entonces  $S = Q\sqrt{\Lambda}\sqrt{\Lambda}Q^\top = AA^\top$ . (Hernández 2026, Pr4 Ej1)

( $\impliedby$ ) Para cualquier matriz  $A \in \mathbb{R}^{m \times n}$ , el producto  $S = A^\top A$  es  $\succeq 0$ .

**Prueba** La forma cuadrática se reduce a la norma al cuadrado del vector transformado, y ésta es siempre no negativa.

$$x^\top(A^\top A)x = (Ax)^\top(Ax) = \|Ax\|^2 \geq 0$$

$A^\top A$  es estrictamente  $\succ 0$  si y solo si las columnas de  $A$  son linealmente independientes (rango  $n$ ). En ese caso,  $Ax = 0$  implica necesariamente  $x = 0$ .

---

<sup>2</sup>En computación numérica, esto se traduce en la **Factorización de Cholesky**:  $S = LL^\top$ , donde  $L$  es una matriz triangular inferior con diagonal positiva.

### 10.3.4. 4: Determinante y Menores Principales Positivos

Una matriz  $S$  es definida positiva si todos sus **menores principales líderes** son positivos. Esto incluye:

- El determinante de la matriz completa:  $\det(S) = \prod \lambda_i > 0$ .
- Todos los determinantes de las submatrices superiores izquierdas de tamaño  $k \times k$ .

### 10.3.5. 5: Pivotes de eliminacion Gaussiana Positivos

Toda matriz simétrica definida positiva permite una eliminación gaussiana sin necesidad de intercambio de filas (pivoteo), lo que conduce a la **descomposición simétrica**  $A = LDL^T$ . En esta estructura:

---

|       |   |
|-------|---|
| $L$   | Es una matriz triangular inferior unitaria (con 1s en la diagonal principal). Representa las operaciones elementales de fila realizadas durante la eliminación. |
| $D$   | Es una matriz diagonal. Contiene los pivotes del proceso de eliminación.  |
| $L^T$ | Es la transpuesta de $L$ , una matriz triangular superior.  |

---

Prueba

Para demostrar por qué los pivotes deben ser positivos, analizamos la “energía” o **forma cuadrática** asociada a la matriz. Comenzamos por multiplicar por un vector no nulo  $x \in \mathbb{R}^n$  en ambos lados:

$$x^T Ax = x^T (LDL^T)x$$

Por la propiedad asociativa del producto matricial, podemos agrupar los términos de la siguiente manera:

$$x^T Ax = (L^T x)^T D(L^T x)$$

Si definimos un nuevo vector columna  $y$  como  $y = L^T x$ , la expresión se simplifica a una suma ponderada de cuadrados:

$$x^T Ax = y^T Dy = \sum_{i=1}^n d_i y_i^2 = d_1 y_1^2 + d_2 y_2^2 + \dots + d_n y_n^2$$

Dado que  $L$  (y por ende  $L^T$ ) es una matriz triangular con 1s en la diagonal, su determinante es 1 y es, por tanto, invertible. Esto garantiza que para cualquier  $y \neq 0$ , siempre existe un único

$x \neq 0$  tal que  $L^\top x = y$ . Para que la suma  $\sum_{i=1}^n d_i y_i^2$  sea estrictamente positiva para cualquier valor de  $y_i$  (donde al menos un  $y_i \neq 0$ ), **todos los coeficientes  $d_i$  deben ser estrictamente mayores a cero.**

Si existiera un pivote  $d_k \leq 0$ , podríamos elegir un vector  $y$  con componentes nulas excepto en la posición  $k$ , lo que resultaría en una forma cuadrática  $x^\top Ax \leq 0$ , invalidando la propiedad de ser definida positiva.

## 10.4. Ejemplos y Aplicaciones Fundamentales

Para identificar la positividad en la práctica, analizamos estructuras matriciales recurrentes en el álgebra lineal numérica y la ciencia de datos, vinculándolas con la estructura de los subespacios y la teoría espectral.

### 10.4.1. Relación con la Matriz Identidad

Cualquier matriz definida positiva  $S$  puede verse como una deformación de la matriz identidad. Mientras que  $x^\top \mathbb{1}x = \|x\|^2$  define una esfera perfecta (todas las direcciones pesan lo mismo),  $x^\top Sx$  define un elipsoide donde los ejes están alineados con los autovectores de  $S$  y sus longitudes dependen de  $1/\sqrt{\lambda_i}$ .

Si  $S \succ 0$ , entonces existe un escalar  $\alpha > 0$  tal que:  $S - \alpha \mathbb{1} \succeq 0$ . Esto indica que  $S$  “crece” al menos tan rápido como una versión escalada de la identidad.

### 10.4.2. Desplazamientos (Shifts)

Una matriz simétrica arbitraria  $S$  puede convertirse en definida positiva mediante un “shift” o desplazamiento espectral utilizando la matriz identidad  $\mathbb{1}$  (Armentano (2026) clase 8). Si  $\lambda_{min}$  es el autovalor más pequeño de  $S$ , entonces para cualquier  $\alpha > |\lambda_{min}|$  se cumple que  $S + \alpha \mathbb{1}$  es estrictamente Definida Positiva.

### 10.4.3. La Matriz de Unos ( $J$ )

La matriz  $J$ , donde cada entrada  $J_{ij} = 1$ , es un caso crítico de matriz de **rango 1**. Puede expresarse como el producto exterior  $J = \mathbf{1}\mathbf{1}^\top$ , donde  $\mathbf{1}$  es el vector columna de unos. Sus autovalores son  $\lambda = \{n, 0, 0, \dots, 0\}$ . Al tener autovalores no negativos, es **semidefinida positiva**. Es singular para cualquier  $n > 1$  porque su núcleo tiene dimensión  $n - 1$ .

#### 10.4.4. Matrices de Proyección Ortogonal

Toda matriz definida positiva debe ser, por definición, simétrica. La simetría garantiza que todos los autovalores sean reales, lo que permite ordenarlos y validar si el autovalor mínimo  $\lambda_{min} > 0$ . Un ejemplo común son las matrices de proyección ortogonal  $P = QQ^T$ , las cuales son  $\succeq 0$  pero no  $\succ 0$  (tienen autovalores 1 y 0).

#### 10.4.5. Matrices de Markov Simétricas

Una matriz de Markov  $M$  cumple que sus columnas suman 1. Aunque no todas las matrices de Markov son  $\succ 0$ , si además  $M$  es simétrica, sus autovalores son reales y el autovalor dominante es  $\lambda_1 = 1$ .

# 11 SVD(Descomposición en Valores Singulares)

§

§

A diferencia de la diagonalización  $A = X\Lambda X^{-1}$ , que requiere que la matriz sea cuadrada y posea suficientes autovectores, y la descomposición espectral  $S = Q\Lambda Q^T$  que se limita a matrices simétricas, la SVD es una generalización universal: existe para cualquier matriz  $A \in \mathbb{R}^{m \times n}$ , ya sea cuadrada o rectangular. (Strang 2018 Lec. 6; Strang 2019, I.7)

## 11.1. Motivación: Limitaciones de los Autovectores

La noción de vectores propios ( $Ax = \lambda x$ ) presenta algunos inconvenientes:

Si  $A$  es  $m \times n$ , el producto  $Ax$  reside en  $\mathbb{R}^m$ , mientras que  $x$  está en  $\mathbb{R}^n$ , haciendo imposible la igualdad  $Ax = \lambda x$ . Pero, incluso si la matriz es cuadrada, los autovalores pueden ser complejos o los autovectores pueden no ser ortogonales.

La SVD supera estos inconvenientes utilizando dos conjuntos distintos de vectores ortonormales: los **vectores singulares a izquierda** ( $U$ ) y a **derecha** ( $V$ ).

## 11.2. La Ecuación Fundamental

Toda matriz  $A$  puede factorizarse en el producto de tres matrices:

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

Donde:

---

|                                      |  |
|--------------------------------------|--|
| $U \in \mathbb{R}^{m \times m}$      | Es una matriz ortogonal cuyas columnas son los <b>vectores singulares a izquierda</b> y representan una base ortonormal para el codominio.   |
| $\Sigma \in \mathbb{R}^{m \times n}$ | Es una matriz diagonal (aunque rectangular) que contiene los <b>valores singulares</b> $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$ |
| $V \in \mathbb{R}^{n \times n}$      | Es una matriz ortogonal cuyas columnas son los <b>vectores singulares a derecha</b> y representan una base ortonormal para el dominio.       |

---

### 11.3. Intuición Geométrica: Rotación y Estiramiento

La acción de multiplicar por una matriz  $A$  puede descomponerse en tres pasos geométricos elementales (Strang 2018 Lec. 3 '6):

---

|   |   |
|---|---|
| <b>Rotación en el dominio (<math>V^\top</math>)</b> | Cambia el vector de entrada a la base ortonormal de vectores singulares sin cambiar su longitud].   |
| <b>Escalamiento (<math>\Sigma</math>)</b>           | Estira o contrae el vector a lo largo de los ejes principales. Aquí, la esfera unitaria se transforma en un elipsoide cuyos semiejes tienen longitudes $\sigma_i$ . |
| <b>Rotación en el codominio (<math>U</math>)</b>    | Orienta el elipsoide resultante en el espacio final.  |

---

### 11.4. Deducción de la SVD

Para construir la SVD, buscamos una relación análoga a la de los autovectores ( $Ax = \lambda x$ ), pero adaptada a matrices rectangulares donde el dominio y el codominio tienen dimensiones distintas.

El punto de partida es encontrar un conjunto de vectores ortonormales en el dominio  $\{v_1, v_2, \dots, v_r\}$  que, al ser transformados por  $A$ , resulten en vectores ortonormales en el codominio  $\{u_1, u_2, \dots, u_r\}$ , escalados por un factor  $\sigma_i$ :

$$Av_i = \sigma_i u_i \quad i = 1, \dots, r$$

Donde  $r$  es el rango de la matriz. Estas ecuaciones representan la acción de la matriz sobre sus direcciones principales (Strang 2018 - Lecture 6 '20).

Podemos agrupar estas  $r$  ecuaciones individuales colocando los vectores  $v_i$  como columnas de una matriz  $V$  y los vectores  $u_i$  como columnas de una matriz  $U$  (Strang 2018 - Lecture 6 '30):

$$A \underbrace{\begin{bmatrix} | & | & \dots & | \\ v_1 & v_2 & \dots & v_r \\ | & | & \dots & | \end{bmatrix}}_V = \underbrace{\begin{bmatrix} | & | & \dots & | \\ u_1 & u_2 & \dots & u_r \\ | & | & \dots & | \end{bmatrix}}_U \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & & \sigma_r \end{bmatrix}}_\Sigma$$

Esta disposición compacta las relaciones vectoriales en la ecuación:

$$AV = U\Sigma$$

Dado que los vectores  $v_i$  son elegidos para ser ortonormales, la matriz  $V$  cumple la propiedad de las matrices de Stiefel:  $V^T V = \mathbb{I}$  (Strang 2018 Lec. 3 '02; Armentano 2026 clase 5). Al multiplicar por  $V^T$  a la derecha en ambos lados de la ecuación anterior, obtenemos la representación final de la SVD  $A = U\Sigma V^T$

$$A = \underbrace{\begin{bmatrix} | & | & & | \\ u_1 & u_2 & \dots & u_r \\ | & | & & | \end{bmatrix}}_{U \in \mathcal{M}_{m \times r}} \underbrace{\begin{bmatrix} \sigma_1 & & & \\ & \ddots & & \\ & & \sigma_r & \\ & & & \end{bmatrix}}_{\Sigma \in \mathcal{M}_{r \times r}} \underbrace{\begin{bmatrix} - & v_1^T & - \\ - & v_2^T & - \\ & \vdots & \\ - & v_r^T & - \end{bmatrix}}_{V^T \in \mathcal{M}_{r \times n}}$$

En esta estructura, las columnas de  $U$  representan la base ortonormal de la imagen, mientras que las filas de  $V^T$  (que son los vectores columna  $v_i$  transpuestos) representan la base ortonormal del espacio fila.

Entonces, la conexión con la estructura espectral de las matrices simétricas permite identificar la procedencia de estos vectores:

- **Las columnas de  $V$**  son los vectores propios de la diagonalización de  $A^T A$ .
- **Las columnas de  $U$**  son los vectores propios de la diagonalización de  $AA^T$ .

**i** Nota

**Observación:** En la SVD completa las matrices  $V^T$  y  $U$  son  $n \times n$  y  $m \times m$  en lugar de  $r \times n$  y  $m \times r$ . Para ello, se completan con vectores ortogonales del espacio nulo de  $A$  hasta formar una base ortonormal de sus espacios correspondientes  $\mathbb{R}^n$  y  $\mathbb{R}^m$ .

A su vez, la matriz  $\Sigma$  en vez de ser cuadrada  $r \times r$  es rectangular  $m \times n$  de la misma forma que  $A$ . Tiene los  $r$  valores singulares en la diagonal y ceros en el resto.

En el producto  $U\Sigma V^T$ , los vectores extra de  $U$  y  $V^T$  simplemente se anulan al multiplicar por los ceros de  $\Sigma$ . Solo se retienen los elementos asociados a los  $r$  valores singulares no nulos.

| SVD Completa          | SVD Reducida            | Descripción en la SVD Reducida                          |
|-----------------------|-------------------------|---|
| $U (m \times m)$      | $U_r (m \times r)$      | Base ortonormal para el espacio columna (Imagen) de $A$ |
| $\Sigma (m \times n)$ | $\Sigma_r (r \times r)$ | Matriz diagonal de valores singulares no nulos          |
| $V^T (n \times n)$    | $V_r^T (r \times n)$    | Base ortonormal para el espacio fila de $A$             |

| SVD Completa     | SVD Reducida     | Descripción en la SVD Reducida                                     |
|------------------|------------------|--|
| $A (m \times n)$ | $A (m \times n)$ | La matriz original se reconstruye exactamente igual en ambos casos |

## 11.5. Relación con la Estructura Espectral

La construcción de la SVD se apoya en las matrices simétricas  $A^T A$  y  $AA^T$ , las cuales siempre son semidefinidas positivas.

$A^T A$  es una matriz cuadrada ( $n \times n$ ), simétrica y **definida positiva**. Sus autovectores forman las columnas de  $V$  y sus autovalores son  $\sigma_i^2$ .

$AA^T$  es una matriz cuadrada ( $m \times m$ ) que posee los mismos autovalores no nulos que  $A^T A$ . Sus autovectores forman las columnas de  $U$ .

|                                     |  |
|-------------------------------------|--|
| $V$ Vectores singulares a derecha   | Son los autovectores de la matriz simétrica $A^T A$ . Se cumple que $(A^T A)v_i = \sigma_i^2 v_i$  |
| $U$ Vectores singulares a izquierda | Son los autovectores de la matriz simétrica $AA^T$ . Se cumple que $(AA^T)u_i = \sigma_i^2 u_i$  |
| $\sigma_i$ Valores singulares       | Son las raíces cuadradas de los autovalores no nulos de $A^T A$ (o $AA^T$ ). Representan el factor de estiramiento en cada dirección invariante. |

Algebraicamente, esto se verifica expandiendo el producto:

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

Dado que  $U^T U = \mathbb{I}$ , se observa que  $V$  diagonaliza a  $A^T A$ , y en esta descomposición se ve que las columnas de  $V$  son los vectores propios de  $A^T A$

De manera analoga:

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

Dado que  $V^T V = \mathbb{I}$ , se observa que  $U$  diagonaliza a  $AA^T$ , y en esta descomposición se ve que las columnas de  $U$  son los vectores propios de  $AA^T$ .

## 11.6. Construcción y Prueba de Ortogonalidad

Buscamos vectores que cumplan la relación  $Av_i = \sigma_i u_i$ .

1. Se eligen los  $v_i$  como los autovectores de  $A^T A$ .
2. Se calculan los  $\sigma_i$  como la raíz cuadrada de los autovalores de  $A^T A$ .
3. Se definen los vectores de salida como  $u_i = \frac{Av_i}{\sigma_i}$ .

Para demostrar que los  $u_i$  generados son ortogonales entre sí ( $u_i^T u_j = 0$ ), utilizamos la propiedad de los autovectores de  $A^T A$ :

$$u_i^T u_j = \left( \frac{Av_i}{\sigma_i} \right)^T \left( \frac{Av_j}{\sigma_j} \right) = \frac{v_i^T (A^T Av_j)}{\sigma_i \sigma_j} = \frac{v_i^T (\sigma_j^2 v_j)}{\sigma_i \sigma_j} = \frac{\sigma_j}{\sigma_i} (v_i^T v_j)$$

Como los  $v_i$  son ortonormales,  $v_i^T v_j = 0$ , lo que garantiza la ortogonalidad de los  $u_i$ .

## 11.7. Observaciones y Propiedades

### Conteo de Parámetros

La SVD describe cualquier transformación lineal como una secuencia de **Rotación** ( $V^T$ )  $\rightarrow$  **Estiramiento** ( $\Sigma$ )  $\rightarrow$  **Rotación** ( $U$ ).

El número de grados de libertad en la matriz original debe coincidir con los de su forma SVD:

- **Caso 2x2 (4 parámetros):** 2 valores singulares en  $\Sigma$ , 1 ángulo de rotación  $\theta$  para  $V$  y 1 ángulo para  $U$ .
- **Caso 3x3 (9 parámetros):** 3 valores singulares en  $\Sigma$ , 3 parámetros de rotación en 3D (los tres ángulos posibles, conocidos en ingeniería como *roll*, *pitch* y *yaw*) para  $V$ , y otros 3 para  $U$ .

### Igualdad de $U$ y $V$

Los vectores singulares a izquierda y derecha coinciden ( $U = V$ ) únicamente cuando la matriz  $A$  es cuadrada, simétrica y definida positiva. (Hernández 2026, 4 Ej 21)

### Autovalores Repetidos

Si existe un autovalor doble, los vectores singulares no son únicos, sino que generan un **plano singular** de direcciones posibles

## Determinantes

El determinante de una matriz ortogonal es siempre 1 (o  $-1$  si incluye reflexión) (Strang 2018 Lec. 3 '13). Para una matriz cuadrada, el producto de los valores singulares  $\prod \sigma_i$  es igual al valor absoluto del determinante.

## 11.8. La Pseudo-inversa ( $A^+$ )

La SVD proporciona una forma robusta de definir la inversa para matrices no invertibles o rectangulares (Strang 2018 Lec. 9 '06; Strang 2019, I.7):

$$A^+ = V\Sigma^+U^\top$$

Donde  $\Sigma^+$  se obtiene reemplazando cada  $\sigma_i > 0$  por  $1/\sigma_i$ .

Si  $A$  tiene columnas independientes,  $A^+ = (A^\top A)^{-1}A^\top$ , lo que conecta directamente con la solución de mínimos cuadrados.

### 11.8.1. Deducción

Cuando una matriz  $A \in \mathbb{R}^{m \times n}$  tiene sus columnas linealmente independientes ( $m \geq n$  y  $\text{rank}(A) = n$ ), el sistema  $Ax = b$  es generalmente sobredeterminado y no tiene una solución exacta (Strang 2018 Lec. 9; Armentano 2026 Clase 17). En este escenario, la pseudoinversa  $A^+$  surge naturalmente de la necesidad de encontrar la “mejor solución” posible.

La solución que minimiza el error cuadrático  $\|Ax - b\|_2^2$  debe satisfacer las **ecuaciones normales**:

$$A^\top A\hat{x} = A^\top b$$

Esta condición geométrica asegura que el residuo  $e = b - A\hat{x}$  sea ortogonal al espacio columna de  $A$ .

Si  $A$  tiene columnas independientes, entonces la matriz cuadrada  $A^\top A$  es simétrica y **estrictamente definida positiva** (Strang 2018 Lec. 5; Armentano 2026 Clase 9; Armentano 2026 Clase 17). Esto garantiza que  $A^\top A$  es invertible, permitiéndonos despejar el vector de coeficientes óptimo  $\hat{x}$ :

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

La pseudoinversa se define como el operador lineal que, al actuar sobre el vector de datos  $b$ , produce directamente la solución de mínimos cuadrados  $\hat{x}$ . Por inspección de la fórmula anterior, identificamos:

$$A^+ = (A^\top A)^{-1} A^\top$$

### 11.8.2. Propiedades

Esta expresión para  $A^+$  es conocida como la **inversa por la izquierda**, ya que posee una propiedad fundamental si se premultiplica por  $A$  (Strang 2018 Lec. 9):

$$A^+ A = ((A^\top A)^{-1} A^\top) A = (A^\top A)^{-1} (A^\top A) = I_{n \times n}$$

Esto significa que  $A^+$  deshace la acción de  $A$  perfectamente para cualquier vector que ya se encuentre en el espacio fila. Sin embargo, a diferencia de una inversa verdadera,  $AA^+ \neq I$  (a menos que  $m = n$ ), sino que representa la **matriz de proyección ortogonal** sobre el espacio columna de  $A$ .

Observar :Esto solo vale si  $A^\top A$  es invertible, es decir, si  $A$  tiene columnas l.i.. En este caso  $AA^\top$  representa una proyección ortogonal en  $C(A)$  y únicamente cuando  $m = n$  se cumple que  $AA^\top = \mathbb{1}$  (Hernández 2026, 6, ej3)

# 12 Normas

§

§

En el universo del álgebra lineal contemporánea, medir la magnitud de los objetos estructurados exige trascender la intuición geométrica clásica de la recta real. Al adentrarnos en espacios multidimensionales, las normas vectoriales emergen no solo como herramientas analíticas para definir distancias y esculpir nuevas geometrías, sino como el lenguaje fundamental para cuantificar el error y garantizar la convergencia algorítmica.

## 12.1. Definición

Una norma es una función que asigna un número real no negativo a un objeto  $v$  para cuantificar su magnitud. Para que una medida sea considerada matemáticamente una norma, debe cumplir con tres propiedades fundamentales: (Strang 2018 Lec. 8; Strang 2019, I.11)

**Positividad**  $\|v\| \geq 0$ ,  $\|v\| = 0 \implies v = 0$

**Homogeneidad**  $\|cv\| = |c| \cdot \|v\|$

**Desigualdad del triángulo**  $\|v + w\| \leq \|v\| + \|w\|$

### 12.1.1. La Identidad de Polarización y la Ley del Paralelogramo

La **Identidad de Polarización** es un resultado fundamental que establece un vínculo directo entre el producto interno de un espacio y la norma que este induce (Hernández 2026, Pr5 Ej4). Mientras que una norma nos dice qué tan “largo” es un vector, el producto interno nos da información sobre la relación angular entre dos vectores; esta identidad demuestra que, bajo ciertas condiciones, el producto interno puede recuperarse por completo conociendo solo las magnitudes (Hernández 2026, Pr5 Ej4).

La identidad se deriva expandiendo la norma al cuadrado de la suma (o resta) de dos vectores utilizando las propiedades del producto interno (Hernández 2026, Pr5 Ej4):

$$\begin{aligned}\|v + w\|^2 &= \langle v + w, v + w \rangle = \|v\|^2 + \|w\|^2 + 2\langle v, w \rangle \\ \|v - w\|^2 &= \langle v - w, v - w \rangle = \|v\|^2 + \|w\|^2 - 2\langle v, w \rangle\end{aligned}\tag{12.1}$$

Al despejar el producto interno de la primer ecuación, obtenemos la primera forma de la identidad (Hernández 2026, Pr5 Ej4):

$$\langle v, w \rangle = \frac{1}{2}(\|v + w\|^2 - \|v\|^2 - \|w\|^2)$$

Si en cambio restamos ambas ecuaciones 12.1, llegamos a la forma más utilizada (Hernández 2026, Pr5 Ej4):

$$\langle v, w \rangle = \frac{1}{4}(\|v + w\|^2 - \|v - w\|^2) \quad (12.2)$$

Si en lugar de sumar las ecuaciones, las restamos, obtenemos una regla equivalente para normas inducidas: la **Ley del Paralelogramo**. Geométricamente, establece que en un paralelogramo, la suma de los cuadrados de las longitudes de las diagonales es igual a la suma de los cuadrados de las longitudes de los cuatro lados (Hernández 2026, Pr5 Ej4).

$$\|v + w\|^2 + \|v - w\|^2 = 2\|v\|^2 + 2\|w\|^2 \quad (12.3)$$

### 12.1.2. Normas Inducidas por Producto Interno

Una norma es **inducida por un producto interno** si existe una función  $\langle v, w \rangle$  tal que  $\|v\| = \sqrt{\langle v, v \rangle}$ .

Estas normas son las únicas que satisfacen la Identidad de Polarización Hernández (2026).

Si una norma no cumple esta igualdad (como la norma  $\ell_1$  o  $\ell_\infty$ ), no puede provenir de un producto interno y, por lo tanto, no define una noción de “ángulo” o “proyección ortogonal” en el sentido euclidiano habitual.

Esta propiedad es la que permite que espacios como el de las matrices con la norma de Frobenius se consideren **Espacios de Hilbert**. Gracias a la polarización, podemos hablar de “ángulos” u “ortogonalidad” entre matrices utilizando el producto interno de Hilbert-Schmidt.

## 12.2. Normas Vectoriales

Las normas más utilizadas en el análisis de datos pertenecen a la familia  $\ell_p$

$$\|v\|_p = \left( \sum_{i=1}^n |v_i|^p \right)^{1/p}$$

Dentro de esta familia, destacan cuatro casos fundamentales:

**Norma  $\ell_2$  (Euclidiana)** Representa la distancia más corta entre dos puntos. Es la única norma de esta familia inducida por el producto punto estándar  $v^\top v$ .

$$\|v\|_2 = \sqrt{\sum_{i=1}^n v_i^2} = (v_1^2 + v_2^2 + \dots + v_n^2)^{\frac{1}{2}}$$

**Norma  $\ell_1$  (Manhattan)** Se define como la suma de las magnitudes absolutas. <sup>1</sup>.

$$\|v\|_1 = \sum_{i=1}^n |v_i| = |v_1| + |v_2| + \dots + |v_n|$$

**Norma  $\ell_\infty$  (Máximo)** Resulta de llevar  $p$  al infinito, lo que hace que el componente de mayor magnitud domine sobre los demás.

$$\|v\|_\infty = \max_{1 \leq i \leq n} |v_i|$$

### 12.2.1. Geometría de las Normas

La geometría de una norma se visualiza a través de su “bola unitaria” (el conjunto de vectores con norma igual a 1). La forma de esta bola determina el comportamiento en problemas de optimización:

La geometría de una norma se visualiza a través de su “bola unitaria”, definida como el conjunto de vectores con norma igual a 1.

$$\mathcal{B} = \{v \in \mathbb{R}^n : \|v\|_p \leq 1\}$$

La forma de esta bola determina cómo la norma penaliza las componentes del vector y es crucial en problemas de optimización con restricciones. A medida que el parámetro  $p$  varía, la forma de la bola unitaria experimenta una transformación geométrica continua:

**Norma  $\ell_1$  (Diamante)** En  $\mathbb{R}^2$ , la bola es un diamante con vértices en los ejes  $(\pm 1, 0)$  y  $(0, \pm 1)$ .

Geoméricamente, los vértices “puntiagudos” en los ejes explican por qué la minimización de  $\|v\|_1$  bajo restricciones lineales tiende a encontrar soluciones dispersas (*sparse*), ya que es más probable que la restricción toque primero un vértice.

**Norma  $\ell_2$  (Círculo/Esfera)** Representa el caso clásico donde la bola es un círculo perfecto.

No favorece ninguna dirección en particular, lo que resulta en soluciones donde muchas componentes son pequeñas pero no nulas.

**Norma  $\ell_\infty$  (Cuadrado/Cubo)** La bola unitaria es un cuadrado con lados paralelos a los ejes, definidos por las rectas  $x = \pm 1$  y  $y = \pm 1$ . En este caso, el tamaño del vector solo depende de su componente más grande.

---

<sup>1</sup>Su minimización es el motor del *compressed sensing*, pues tiende a producir soluciones ralas (con componentes nulos)

**Familia  $\ell_p$  Genérica** Para  $1 < p < 2$ , la bola es una forma intermedia entre el diamante y el círculo. Para  $p > 2$ , la bola se “infla” hacia el cuadrado (*squircle*). Una propiedad crítica es que para cualquier  $p \geq 1$ , la bola unitaria es un **conjunto convexo**, lo que garantiza que cualquier mínimo local en un problema de optimización sea también un mínimo global. Las normas verdaderas siempre definen conjuntos convexos. Cuando  $p < 1$ , la medida pierde la convexidad y deja de ser una norma válida.

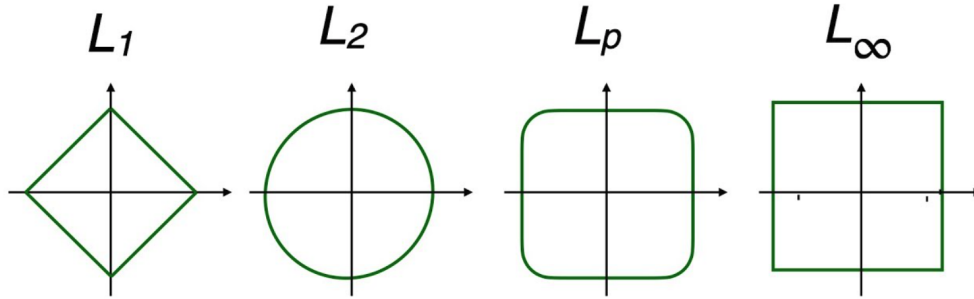


Figura 12.1: Familia de normas  $L_p$  Fuente: [Dr Will Wood](#)

### 12.2.2. El Caso Especial de $\ell_0$

Aunque se le llama norma  $\ell_0$  en la práctica, formalmente no es una norma por incumplir la homogeneidad escalar. Representa la cantidad de elementos distintos de cero en el vector.

$$\|v\|_0 = \#\{i : v_i \neq 0\}$$

Tiene una geometría degenerada. Su “bola unitaria” no abarca un área, sino que consiste únicamente en los ejes coordenados (donde solo una componente es no nula). Al no ser un conjunto convexo (y no cumplir con la homogeneidad escalar),  $\ell_0$  no es una norma verdadera. De forma general, la desigualdad triangular falla cuando  $p < 1$ <sup>2</sup>.

### 12.2.3. Normas $S$

La norma  $S$ , frecuentemente denominada norma de energía, permite adaptar la medición de magnitud a la estructura de un problema específico mediante una matriz simétrica definida positiva  $S$ .

Para un vector  $v$ , la norma  $S$  se define como:

$$\|v\|_S = \sqrt{v^T S v}$$

---

<sup>2</sup>Minimizar  $\|v\|_0$  es un problema de combinatoria NP-duro. La importancia de la norma  $\ell_1$  radica en que es la “envoltura convexa” de  $\ell_0$ , lo que permite resolver problemas de dispersión mediante técnicas de optimización convexa eficientes (Strang 2019, I.11).

Para que esta expresión satisfaga formalmente los axiomas de una norma, especialmente la positividad ( $\|v\| > 0$  para  $v \neq 0$ ), la matriz  $S$  debe ser estrictamente definida positiva. Si  $S$  es la matriz identidad ( $I$ ), la norma  $S$  colapsa a la norma euclidiana estándar  $\ell_2$ .

### Interpretación Geométrica

A diferencia de la norma  $\ell_2$ , cuya bola unitaria es una esfera perfecta, la bola unitaria de la norma  $S$  (definida por el conjunto  $\{v : v^\top S v \leq 1\}$ ) describe un elipsoide en  $\mathbb{R}^n$ .

Las direcciones de los ejes de este elipsoide coinciden con los vectores propios  $q_i$  de la matriz  $S$ . La extensión del elipsoide en cada dirección es inversamente proporcional a la raíz cuadrada del valor propio correspondiente, siendo la longitud del  $i$ -ésimo semieje igual a  $1/\sqrt{\lambda_i}$ .<sup>3</sup>

## 12.3. Normas en Espacios de Funciones

La generalización del concepto de norma permite medir la magnitud no solo de vectores discretos, sino también de objetos continuos como las funciones. Esta transición es fundamental para el análisis funcional y el procesamiento de señales (Strang 2018 Lec. 8).

### 12.3.1. Norma $L^p$ Funcional

En un espacio de funciones definidas sobre un intervalo  $[a, b]$ , la norma  $L^p$  sustituye la sumatoria por una integral de la magnitud de la función elevada a la potencia  $p$  (Hernández 2026, Pr1 Ej11):

$$\|f\|_p = \left( \int_a^b |f(x)|^p dx \right)^{1/p}$$

### La Norma $L^2$ y la Energía

La norma más utilizada es la **norma**  $L^2$ , también conocida como la norma de energía de la función. Al igual que en el caso vectorial, es la única de la familia  $L^p$  que es inducida por un producto interno (Strang 2018 Lec. 8; Armentano 2026 Clase 15):

$$\langle f, g \rangle = \int_a^b f(x)g(x) dx \implies \|f\|_2 = \sqrt{\int_a^b |f(x)|^2 dx}$$

---

<sup>3</sup>Esta norma es fundamental en algoritmos de optimización y aprendizaje profundo, donde se utiliza para definir funciones de pérdida que penalizan de manera distinta diferentes direcciones del espacio de parámetros.

Esta estructura convierte al conjunto de funciones de cuadrado integrable en un **Espacio de Hilbert**, permitiendo extender conceptos geométricos como la ortogonalidad. Por ejemplo, las funciones seno y coseno son ortogonales bajo este producto interno, lo que constituye la base de las series de Fourier.

## 12.4. Cauchy - Schwarz

La desigualdad de Cauchy-Schwarz es uno de los pilares fundamentales del análisis matemático y el álgebra lineal aplicada. Establece un límite superior estricto para el producto interno de dos vectores en términos del producto de sus normas Hernández (2026).

$$\forall v, w \in \mathbb{R}^n \quad |v^\top w| \leq \|v\|_2 \|w\|_2$$

La importancia de esta relación radica en los extremos de su cumplimiento:

**Igualdad** ( $|v^\top w| = \|v\|_2 \|w\|_2$ ) Ocurre si y solo si los vectores son linealmente dependientes (colineales), es decir,  $v = \alpha w$  para algún escalar  $\alpha$ .

**Ortogonalidad** ( $v^\top w = 0$ ) Representa el límite inferior del valor absoluto, donde los vectores son perpendiculares y no comparten ninguna dirección común.

Esta desigualdad garantiza que la noción geométrica del coseno de un ángulo ( $\cos \theta = \frac{v^\top w}{\|v\| \|w\|}$ ) siempre esté bien definida dentro del intervalo  $[-1, 1]$ , permitiendo que la geometría euclidiana sea consistente en cualquier número de dimensiones.

### 12.4.1. Deducción Algebraica y Geométrica

Para deducir la desigualdad, consideramos la proyección ortogonal de un vector  $v$  sobre la dirección de un vector  $w$ . Definimos el vector de proyección  $p$  como Strang (2018):

$$p = \frac{v^\top w}{\|w\|_2^2} w$$

El residuo o error de esta aproximación es el vector  $e = v - p$ , el cual es, por construcción, perpendicular a  $w$ .

Dado que la norma al cuadrado de cualquier vector debe ser mayor o igual a cero, aplicamos esta propiedad al vector de error  $e$ :

$$0 \leq \|e\|_2^2 = \langle v - p, v - p \rangle$$

Al expandir el producto interno y sustituir la definición de  $p$ , obtenemos:

$$0 \leq v^\top v - \frac{(v^\top w)^2}{w^\top w}$$

Reorganizando los términos, llegamos a la forma cuadrática de la desigualdad:

$$(v^\top w)^2 \leq (v^\top v)(w^\top w)$$

Al extraer la raíz cuadrada en ambos lados, se obtiene el enunciado clásico:  $|v^\top w| \leq \|v\|_2 \|w\|_2$ .

### Significado Geométrico

Esta deducción revela que la desigualdad es en realidad una afirmación sobre la “sombra” de un vector: el producto interno (la proyección escalonada) nunca puede superar el producto de las longitudes totales de los vectores originales. El residuo  $e$  mide qué tan lejos están los vectores de ser perfectamente colineales; cuando el residuo es cero, se alcanza la igualdad perfecta (Hernández 2026, Pr3 Ej2).

## 12.5. Desigualdad Triangular

La desigualdad de Cauchy-Schwarz es el eslabón matemático que permite validar a la norma euclidiana como una métrica legítima, pues sin ella no sería posible demostrar la **desigualdad triangular**. Esta propiedad es uno de los tres requisitos fundamentales para que una función sea considerada una norma: establece que el camino directo entre dos puntos es siempre menor o igual a la suma de los tramos de cualquier camino indirecto Hernández (2026).

$$\|v + w\|_2 \leq \|v\|_2 + \|w\|_2$$

Para demostrarlo, partimos del cuadrado de la norma de la suma, expandiéndola mediante el producto interno:

$$\|v + w\|_2^2 = \langle v + w, v + w \rangle = \|v\|_2^2 + \|w\|_2^2 + 2\langle v, w \rangle$$

En este punto, la desigualdad de Cauchy-Schwarz nos permite acotar el término del producto interno:  $2\langle v, w \rangle \leq 2|\langle v, w \rangle| \leq 2\|v\|_2 \|w\|_2$ .

$$\|v + w\|_2^2 \leq \|v\|_2^2 + \|w\|_2^2 + 2\|v\|_2 \|w\|_2$$

Observamos que el lado derecho es un cuadrado perfecto:

$$\|v + w\|_2^2 \leq (\|v\|_2 + \|w\|_2)^2$$

Al extraer la raíz cuadrada, obtenemos la desigualdad triangular.

## Consecuencias en el Espacio Vectorial

Sin Cauchy-Schwarz, la geometría del espacio  $\mathbb{R}^n$  colapsaría. Esta relación garantiza que:

- **La noción de distancia sea coherente:** Asegura que las longitudes medidas por la norma  $L_2$  se comporten según nuestra intuición física en cualquier dimensión.
- **Convexidad del conjunto unitario:** La desigualdad triangular implica que el conjunto de vectores con norma menor o igual a 1 es **convexo** (un círculo, esfera o hiper-esfera), lo cual es crítico para los algoritmos de optimización que buscan mínimos globales.
- **Consistencia con ángulos:** Permite que la definición del coseno entre dos vectores esté siempre entre -1 y 1, vinculando finalmente la magnitud con la dirección.

# 13 Normas Matriciales

§

§

Las normas matriciales extienden el concepto de magnitud a las transformaciones lineales. Al igual que con los vectores, existen diversas formas de medir una matriz  $A$ , y muchas de las más importantes dependen exclusivamente de sus valores singulares  $(\sigma_i)$ .

## 13.1. Producto Hilbert-Schmidt

Para extender los conceptos geométricos de ángulos, ortogonalidad y proyecciones al espacio de las matrices, es necesario definir un producto interno. El estándar para matrices reales es el **producto interno de Hilbert-Schmidt**. (Hernández 2026, Pr5 Ej6; Strang 2019, I.11)

**Definición** Dadas dos matrices  $A, B \in \mathbb{R}^{m \times n}$ , su producto interno se define como:

$$\langle A, B \rangle_{HS} = \text{tr}(A^\top B)$$

Este producto equivale a realizar el producto punto de todos los elementos correspondientes de ambas matrices y sumarlos, tratándolas como si fueran vectores de dimensión  $mn$ .

Para comprender por qué, analizamos la sumatoria de sus componentes: (Strang 2019, I.11; Hernández 2026, Pr5 Ej6)

En el producto matricial  $M = A^\top B$ , cada entrada en la diagonal principal  $M_{jj}$  es el resultado de multiplicar la fila  $j$  de  $A^\top$  (que es la columna  $j$  de  $A$ ) por la columna  $j$  de  $B$ . Esto es el producto punto de las dos columnas  $j$ -ésimas (Strang 2018 Lec. 2 '39):  $M_{jj} = \sum_{i=1}^m a_{ij}b_{ij}$ .

La traza de la matriz resultante es la suma de estos elementos diagonales. Al expandir la sumatoria sobre todas las columnas  $j = 1, \dots, n$ , el producto interno se transforma en una suma doble que abarca cada entrada de ambas matrices:

$$\langle A, B \rangle_{HS} = \text{tr}(A^\top B) = \sum_{j=1}^n \sum_{i=1}^m a_{ij}b_{ij}$$

Si definimos la operación  $\text{vec}(A)$  como el proceso de “estirar” la matriz para convertirla en un vector largo de dimensión  $mn$ , la sumatoria anterior coincide exactamente con el producto punto estándar entre vectores:

$$\langle A, B \rangle_{HS} = \text{vec}(A) \cdot \text{vec}(B)$$

El producto interno Hilbert-Schmidt cumple con las propiedades fundamentales:

- **Simetría:**  $\langle A, B \rangle = \langle B, A \rangle$ .
- **Linealidad:**  $\langle cA, B \rangle = c\langle A, B \rangle$ .
- **Definición Positiva:**  $\langle A, A \rangle \geq 0$ ,  $A \neq 0$ .

El uso de este producto interno convierte al espacio de las matrices  $\mathbb{R}^{m \times n}$  en un **espacio de Hilbert**, lo que permite aplicar resultados potentes del análisis funcional a problemas matriciales.

## 13.2. Norma Nuclear $\|A\|_N$

También conocida como **norma de traza**, se define como:

$$\|A\|_N = \sigma_1 + \dots + \sigma_r$$

Esta medida representa la extensión natural de la norma vectorial  $\ell_1$  al espacio de las matrices, operando sobre el espectro de valores singulares en lugar de los componentes individuales.<sup>1</sup>

---

<sup>1</sup>Ha ganado gran relevancia en problemas de “completitud de matrices” (como el famoso concurso de Netflix) y en la obtención de soluciones de bajo rango en sistemas ruidosos como las imágenes por resonancia magnética (MRI) (Strang 2018 Lec. 8 '47).

La utilidad de la norma nuclear en la ciencia de datos radica en su relación con el concepto de **bajo rango** y **dispersión** (*sparsity*):

Al igual que minimizar la norma  $\ell_1$  en vectores tiende a producir soluciones con muchos ceros, minimizar la norma nuclear en matrices tiende a producir soluciones donde muchos valores singulares son cero, resultando en una matriz de bajo rango.

Esta propiedad es el motor de los algoritmos de “completitud de matrices”, donde el objetivo es adivinar entradas faltantes en una matriz gigante basándose en la premisa de que los datos tienen una estructura subyacente simple.

Fue la técnica clave para resolver el concurso de Netflix, prediciendo las calificaciones de usuarios para películas que aún no habían visto.

En las resonancias magnéticas (MRI), permite reconstruir imágenes nítidas a partir de muestreos rápidos e incompletos, reduciendo el tiempo que el paciente debe permanecer en el escáner.

Ayuda a separar la señal estructural importante del ruido aleatorio, ya que el ruido tiende a dispersarse por todos los valores singulares, mientras que la información real se concentra en unos pocos.

Al ser una norma **ortogonalmente invariante**, la magnitud que mide no cambia si la matriz es rotada, lo que asegura que la recuperación de datos se base exclusivamente en la “energía” intrínseca de la información y no en su orientación en el sistema de coordenadas.

### 13.2.1. Caracterización Variacional

La norma nuclear posee una propiedad fundamental de **caracterización variacional** que permite su optimización sin necesidad de calcular explícitamente la SVD en cada iteración: Strang (2019)

$$\|A\|_N = \min_{A=UV} \|U\|_F \|V\|_F$$

Donde la minimización se realiza sobre todas las posibles factorizaciones de  $A$  en dos matrices  $U \in \mathbb{R}^{m \times k}$  y  $V \in \mathbb{R}^{k \times n}$ .

Demostración de alcanzabilidad mediante SVD

Para probar que el valor mínimo es efectivamente  $\|A\|_N$ , podemos proponer una factorización basada en la SVD de  $A = P\Sigma Q^\top$ . Definimos:  $U = P\Sigma^{\frac{1}{2}}$  y  $V = \Sigma^{\frac{1}{2}}Q^\top$

Es inmediato verificar que  $UV = P\Sigma^{\frac{1}{2}}\Sigma^{\frac{1}{2}}Q^\top = P\Sigma Q^\top = A$ . Calculando el cuadrado de la norma de Frobenius para  $U$  (Hernández 2026, Pr5 Ej6):

$$\|U\|_F^2 = \text{tr}(U^\top U) = \text{tr}(\Sigma^{\frac{1}{2}}P^\top P\Sigma^{\frac{1}{2}}) = \text{tr}(\Sigma) = \sum_{i=1}^r \sigma_i = \|A\|_N$$

Siguiendo el mismo procedimiento para  $V$ , obtenemos que  $\|V\|_F^2 = \|A\|_N$ . Por lo tanto, para esta elección:

$$\|U\|_F \|V\|_F = \sqrt{\|A\|_N} \cdot \sqrt{\|A\|_N} = \|A\|_N$$

Esta identidad es la base de los algoritmos que resolvieron el **concurso de Netflix** (Strang 2018 Lec. 7).<sup>2</sup>

### 13.3. Norma Operador $\|A\|$

La norma de operador mide el factor máximo de “estiramiento” que una matriz aplica a un vector.

---

<sup>2</sup>En lugar de resolver un problema de optimización sobre la matriz completa  $A$ , se optimizan directamente los factores  $U$  y  $V$ :

El término de penalización nuclear  $\|A\|_N$  se sustituye por  $\frac{1}{2}(\|U\|_F^2 + \|V\|_F^2)$ . Minimizar la norma de Frobenius de los factores es mucho más sencillo mediante métodos de descenso de gradiente estocástico (SGD) o mínimos cuadrados alternados, permitiendo procesar matrices con millones de entradas.

Esta propiedad vincula la “energía” de las matrices factorizadas con la estructura de bajo rango de la matriz original, siendo la herramienta clave para la recuperación de datos faltantes (Strang 2019, I.11).

$$\|A\| = \max_{x \neq 0} \left( \frac{\|Ax\|}{\|x\|} \right)$$

Existen otra definicion estándar para esta medida:

$$\|A\| = \max_{\|x\|=1} \|Ax\|$$

En esta expresión es facil ver que  $\|A\|$  representa la máxima elongación que sufre la “bola unitaria” bajo la transformación definida por  $A$ .<sup>3</sup>

Puede probarse facilmente que estas dos definiciones representan el mismo valor.

Consideremos un vector  $x$  cualquiera tal que  $x \neq 0$ . Este vector se puede normalizar para obtener un vector unitario  $u$  definiéndolo como  $u = \frac{x}{\|x\|}$ , de modo que  $\|u\| = 1$ .

Si analizamos la expresión de la primera definición, podemos aplicar la propiedad de **homogeneidad escalar** de las normas ( $\|cv\| = |c| \cdot \|v\|$ ):

$$\frac{\|Ax\|}{\|x\|} = \frac{\|A(\|x\| \frac{x}{\|x\|})\|}{\|x\|} = \frac{\|x\| \cdot \|A(\frac{x}{\|x\|})\|}{\|x\|} = \|Au\|$$

Este desarrollo muestra que el valor del cociente para cualquier vector  $x$  no nulo es idéntico al valor de la norma de  $Au$  para algún vector unitario  $u$ .

Esta norma es inducida por una norma vectorial en  $\mathbb{R}^n$ , por lo tanto la norma operador de una matriz cambia segun la norma elegida en el subespacio.

**Norma Operador Inducida por  $\ell_1$**

$$\|A\|_1 = \max_{1 \leq j \leq n} \sum_{i=1}^m |a_{ij}|$$

**Norma Operador Inducida por  $\ell_\infty$**

$$\|A\|_\infty = \max_{1 \leq i \leq m} \sum_{j=1}^n |a_{ij}|$$

**Norma Operador Inducida por  $\ell_2$  (tambien llamada Norma Espectral)**

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

---

<sup>3</sup>Esta equivalencia es la que permite simplificar problemas de optimización, restringiendo la búsqueda del máximo únicamente a la superficie de la bola unitaria en lugar de todo el espacio vectorial (Strang 2018 Lec. 8 '802; Hernández 2026, Pr5 E;9).

## 13.4. Norma de Frobenius $\|A\|_F$

Es la norma más intuitiva y utilizada en algoritmos de optimización para redes neuronales.

Su definición es equivalente a la norma  $\ell_2$  (norma Euclídea) de la matriz vista como un vector de  $mn$  componentes.

$$\|A\|_F = \left( \sum_{i=1}^m \sum_{j=1}^n |a_{ij}|^2 \right)^{\frac{1}{2}}$$

### Relación con el producto Hilbert-Schmidt

La **norma de Frobenius** es la norma inducida directamente por el producto interno Hilbert-Schmidt (Hernández 2026, Pr5 Ej6):

**Teorema** Sea  $A \in \mathbb{R}^{m \times n} \implies \|A\|_F = \sqrt{\langle A, A \rangle_{HS}} = \sqrt{\text{tr}(A^\top A)}$  (Strang 2018 Lec. 8 '39; Hernández 2026, Pr5 Ej6)

**Demostración** Consideremos el producto  $M = A^\top A$ . Por definición de multiplicación, la entrada en la fila  $j$  y columna  $j$  de  $M$  se obtiene multiplicando la fila  $j$  de  $A^\top$  (que es la columna  $j$  de  $A$ ) por la columna  $j$  de  $A$ , es decir, el producto punto de la columna  $j$  de  $A$  con ella misma:

$$M_{jj} = (A^\top A)_{jj} = \sum_{i=1}^m a_{ij}^2$$

La cantidad  $M_{jj}$  representa la suma de los cuadrados de todos los elementos presentes en la columna  $j$  de la matriz original  $A$ .

La traza de una matriz es la suma de sus elementos en la diagonal principal. Al sumar sobre todas las columnas  $j = 1, \dots, n$ , obtenemos (Strang 2018 Lec. 8 '808):

$$\text{tr}(A^\top A) = \sum_{j=1}^n M_{jj} = \sum_{j=1}^n \sum_{i=1}^m a_{ij}^2$$

La expresión resultante es una sumatoria doble que recorre cada entrada  $a_{ij}$  de la matriz, elevándola al cuadrado. Esto coincide exactamente con el cuadrado de la definición por componentes de la norma:

$$\text{tr}(A^\top A) = \sum_{i,j} a_{ij}^2 = \|A\|_F^2$$

Esta identidad permite transitar libremente entre el cálculo algorítmico (suma de entradas) y el análisis geométrico (traza), facilitando las pruebas de invarianza ortogonal necesarias para resultados avanzados como el Teorema de Eckart-Young.

Esta conexión es la que dota a la norma de Frobenius de su invarianza ortogonal y permite que el error en aproximaciones de bajo rango se calcule con la misma lógica que una distancia euclidiana.

La estructura de espacio de Hilbert es la que permite que la norma de Frobenius herede la fórmula de magnitud habitual de los vectores en  $\mathbb{R}^n$ , calculándose como la raíz cuadrada de la suma de los cuadrados de todas sus entradas (Strang 2018 Lec. 8 '39; Hernández 2026, Pr5 Ej7):

Esta identidad no es solo una curiosidad algebraica; es la razón por la cual la norma de Frobenius hereda todas las propiedades geométricas de los espacios euclidianos, incluyendo la noción de ángulo entre matrices y la proyección ortogonal.

Gracias a esta equivalencia, podemos definir formalmente la **ortogonalidad entre matrices**: decimos que  $A$  y  $B$  son ortogonales si su producto interno es nulo, es decir, si  $\text{tr}(A^\top B) = 0$  (Hernández 2026, Pr5 Ej6).

## 13.5. Propiedades de las normas matriciales

### 13.5.1. Regla de Submultiplicatividad para Normas de Matrices

Además de las propiedades derivadas de la definición de norma para vectores, las normas matriciales deben cumplir una propiedad extra: la submultiplicatividad. Ésta permite estimar el tamaño del producto de dos matrices a partir de las magnitudes de sus factores. Se expresa mediante la desigualdad (Hernández 2026, Pr5 Ej9c) <sup>4</sup>:

$$\|AB\| \leq \|A\| \cdot \|B\|$$

Prueba para Normas Inducidas (Normas de Operador)

Para cualquier norma matricial inducida por una norma vectorial (como la norma espectral  $\ell_2$ ), la demostración es una consecuencia directa de la definición de norma de operador (Hernández 2026, Pr5 Ej9d):

Por definición, para cualquier vector  $x$ , se cumple que  $\|Ax\| \leq \|A\| \cdot \|x\|$ . Consideremos el vector resultante de aplicar el producto  $AB$  a un vector  $x$ . Podemos agrupar los factores como  $A(Bx)$ :

$$\|ABx\| \leq \|A\| \cdot \|Bx\|$$

---

<sup>4</sup>Esta propiedad es vital en el análisis numérico, ya que garantiza que el error en una cadena de transformaciones lineales no crezca de manera incontrolada si las normas de las matrices involucradas están acotadas (Strang 2018 Lec. 8).

Aplicando nuevamente la propiedad al término  $\|Bx\|$ :

$$\|ABx\| \leq \|A\| \cdot (\|B\| \cdot \|x\|) = (\|A\| \cdot \|B\|) \cdot \|x\|$$

Tomando el supremo sobre todos los vectores unitarios ( $\|x\| = 1$ ), obtenemos la norma del producto:

$$\|AB\| = \sup_{\|x\|=1} \|ABx\| \leq \|A\| \cdot \|B\|$$

Prueba para la Norma de Frobenius

La norma de Frobenius también es submultiplicativa, aunque no sea una norma inducida. La demostración utiliza la desigualdad de Cauchy-Schwarz sobre las entradas de la matriz producto (Hernández 2026, Pr5 Ej5):

Sea  $C = AB$ . La entrada  $c_{ij}$  es el producto punto de la fila  $i$  de  $A$  por la columna  $j$  de  $B$ . Por Cauchy-Schwarz:

$$|c_{ij}|^2 = \left| \sum_k a_{ik} b_{kj} \right|^2 \leq \left( \sum_k |a_{ik}|^2 \right) \left( \sum_k |b_{kj}|^2 \right)$$

Observamos que el primer paréntesis es el cuadrado de la norma  $\ell_2$  de la fila  $i$  de  $A$  ( $\|a_{i*}\|_2^2$ ) y el segundo es el de la columna  $j$  de  $B$  ( $\|b_{*j}\|_2^2$ ).

Sumando sobre todas las entradas  $i, j$  para obtener  $\|AB\|_F^2$ :

$$\sum_{i,j} |c_{ij}|^2 \leq \sum_i \|a_{i*}\|_2^2 \sum_j \|b_{*j}\|_2^2$$

Dado que la suma de las normas de las filas es  $\|A\|_F^2$  y la de las columnas es  $\|B\|_F^2$ , al extraer la raíz cuadrada llegamos al resultado:

$$\|AB\|_F \leq \|A\|_F \cdot \|B\|_F$$

Prueba alternativa para la Norma de Frobenius

(Strang 2019, I.11)

En el caso particular donde  $A$  y  $B$  son vectores (en otras palabras,  $AB$  es una matriz de rango uno  $ab^\top$ ), la desigualdad en norma es justamente una igualdad:

$$\|ab^\top\|_F^2 = |a_1|^2(|b_1|^2 + \dots + |b_p|^2) + \dots + |a_m|^2(|b_1|^2 + \dots + |b_p|^2) = \|a\|_F^2 \|b\|_F^2$$

En el caso generico, donde  $AB$  es una suma de matrices de rango 1 obtenida multiplicando las filas de  $A$  por las columnas de  $B$ :

$$\|AB\|_F = \|a_1 b_1^\top + \dots + a_n b_n^\top\|_F$$

Aquí podemos aplicar la desigualdad triangular:

$$\|AB\|_F = \|a_1 b_1^\top\|_F + \dots + \|a_n b_n^\top\|_F$$

Esto es una suma de normas de matrices de rango 1, que ahora sabemos que cumplen la submultiplicatividad:

$$\|AB\|_F = \|a_1\|_F \|b_1\|_F + \dots + \|a_n\|_F \|b_n\|_F$$

Y aquí podemos aplicar Cauchy-Schwarz

$$\|AB\|_F \leq (\|a_1\|_F^2 + \dots + \|a_n\|_F^2)^{\frac{1}{2}} (\|b_1\|_F^2 + \dots + \|b_n\|_F^2)^{\frac{1}{2}}$$

Y esta es la definición de norma de Frobenius, así que

$$\|AB\|_F = \|A\|_F \|B\|_F$$

### 13.5.2. Invarianza Ortogonal

No todas las medidas de magnitud se comportan igual ante transformaciones geométricas. Entender qué hace “especial” a una norma requiere analizar su relación con los productos internos y las rotaciones.

Una norma matricial es **ortogonalmente invariante** si multiplicar una matriz  $A$  por una matriz ortogonal  $Q$  (una rotación o reflexión) no altera su norma (Hernández 2026, Pr5 Ej8). Algebraicamente:  $\|QA\| = \|A\| = \|AQ\|$

Como las matrices ortogonales no cambian los valores singulares de  $A$  (solo rotan los ejes de la elipse de transformación sin alterar sus semiejes), las normas que dependen exclusivamente de los valores singulares (Frobenius, Operador y Nuclear) son invariantes.<sup>5</sup>

Las normas inducidas por  $\ell_1$  o  $\ell_\infty$  en cambio, NO son invariantes. Geométricamente, esto se debe a que estas normas dependen de la orientación de los ejes coordenados (la “bola unitaria”

---

<sup>5</sup>Las tres normas son invariantes ante multiplicaciones por matrices ortogonales ( $Q$  y  $P$ ), ya que  $\|QAP\| = \|A\|$ . Esto sucede porque las transformaciones ortogonales no alteran los valores singulares de la matriz original (Hernández 2026, Pr5 Ej8).

Esta perspectiva permite elegir la norma adecuada según el problema: la norma espectral para el análisis de estabilidad, la de Frobenius para medir distancias euclidianas entre matrices, y la nuclear para la recuperación de datos faltantes.

es un diamante o un cuadrado); al rotar la matriz, los vértices dejan de estar alineados con los ejes, cambiando la suma de los componentes.

La norma Frobenius es invariante ante rotaciones (multiplicación por matrices ortogonales  $Q$ ), lo que simplifica enormemente los cálculos en el álgebra lineal numérica.

$$\|AQ\|_F = \|A\|_F$$

**Prueba** Para una matriz ortogonal  $Q$ , donde  $QQ^\top = I$ :  $\|AQ\|_F^2 = \text{tr}((AQ)^\top(AQ)) = \text{tr}(Q^\top A^\top A Q)$

Usando la propiedad cíclica de la traza ( $\text{tr}(ABC) = \text{tr}(BCA)$ ):  $\text{tr}(A^\top A Q Q^\top) = \text{tr}(A^\top A I) = \text{tr}(A^\top A) = \|A\|_F^2$

### 13.5.3. Vinculación con la SVD

La Descomposición en Valores Singulares (SVD) proporciona un marco unificado para calcular y comparar las normas matriciales más importantes. Dado que las normas ortogonalmente invariantes dependen exclusivamente del espectro de valores singulares ( $\sigma_1, \sigma_2, \dots, \sigma_r$ ), la SVD transforma problemas matriciales complejos en cálculos vectoriales simples sobre el vector  $\sigma$  (Strang 2018 Lec. 8; Hernández 2026, Pr5 Ej9).

Sean  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_r > 0$  los valores singulares de una matriz  $A$ . Las tres normas fundamentales se expresan en función de estos.

**Norma Operador/Espectral** Es el valor singular más grande. Representa el factor de ampliación máximo que la matriz puede aplicar a cualquier vector:

$$\|A\|_2 = \sigma_1$$

Ver prueba

Probaremos que el límite superior es el mayor valor singular, y que el límite es alcanzable.

**El límite superior** Por definición, la norma espectral es el máximo del cociente de Rayleigh para la magnitud de la salida:

$$\|A\|_2 = \max_{x \neq 0} \frac{\|Ax\|_2}{\|x\|_2}$$

Si analizamos el cuadrado del cociente utilizando la matriz de Gram  $A^\top A$ :

$$\frac{\|Ax\|_2^2}{\|x\|_2^2} = \frac{(Ax)^\top(Ax)}{x^\top x} = \frac{x^\top(A^\top A)x}{x^\top x}$$

Dado que  $A^\top A$  es una matriz simétrica y semidefinida positiva, sus valores propios son los cuadrados de los valores singulares ( $\sigma_i^2$ ). El cociente de Rayleigh garantiza que este valor no puede superar al valor propio más grande de la matriz de Gram [3, 4]:

$$\frac{x^\top (A^\top A)x}{x^\top x} \leq \sigma_1^2 \implies \frac{\|Ax\|_2}{\|x\|_2} \leq \sigma_1$$

**El máximo** Para probar que este límite es alcanzable, seleccionamos como entrada el primer vector singular derecho  $v_1$  de la SVD de  $A$ . Por las propiedades de la descomposición, sabemos que  $Av_1 = \sigma_1 u_1$ , donde  $u_1$  es un vector unitario. Al calcular el cociente para este vector específico:

$$\frac{\|Av_1\|_2}{\|v_1\|_2} = \frac{\|\sigma_1 u_1\|_2}{\|v_1\|_2} = \sigma_1 \frac{\|u_1\|_2}{\|v_1\|_2}$$

Como tanto  $u_1$  como  $v_1$  tienen norma igual a 1 por construcción, el resultado es exactamente  $\sigma_1$ . Esto demuestra que el valor singular máximo es el valor exacto de la norma de operador.

**Norma de Frobenius** Es la raíz cuadrada de la suma de los cuadrados de todos los valores singulares. Equivale a la “energía” total de la matriz y es la norma inducida por el producto interno de Hilbert-Schmidt

$$\|A\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2}$$

Ver prueba

(Strang 2018 Lec. 8 '40) Expresando  $A$  en su descomposición SVD, tenemos que:  $\|A\|_F = \|U\Sigma V^\top\|_F$

Gracias a la invarianza ortogonal de la norma Frobenius, y dado que  $U$  y  $V^\top$  son ortogonales, llegamos a:  $\|A\|_F = \|U\Sigma V^\top\|_F = \|\Sigma\|_F = \sqrt{\sigma_1^2 + \sigma_2^2 + \dots + \sigma_r^2}$

Ver prueba alternativa

(Strang 2019, I.11)

$$\|A\|_F^2 = \text{traza}(A^\top A) = \sum \lambda_i = \sum \sigma_i^2$$

**Norma Nuclear** Es la suma de todos los valores singulares. Es el análogo matricial de la norma  $\ell_1$  vectorial y se utiliza para promover soluciones de bajo rango en problemas de optimización:

$$\|A\|_N = \sum_{i=1}^r \sigma_i$$

A partir de estas definiciones, se puede establecer una jerarquía clara basada en cómo cada norma acumula la información de los valores singulares. Para cualquier matriz  $A$ , se cumple que:

$$\|A\|_2 \leq \|A\|_F \leq \|A\|_N$$

# 14 Teorema de Eckart-Young - Aproximación de Bajo Rango

§

§

El Teorema de Eckart-Young es uno de los resultados más potentes del álgebra lineal numérica. Establece que la mejor aproximación de una matriz  $A$  por otra matriz de menor rango se obtiene simplemente truncando su Descomposición en Valores Singulares (SVD). (Strang 2018 Lec. 7 '03; Strang 2019, I.7) En otras palabras, si  $B$  es cualquier matriz de rango a lo sumo  $k$ , entonces la distancia a la matriz original  $A$  en norma de operador ( $L_2$ ) no puede ser menor que el valor singular  $\sigma_{k+1}$ .

## 14.0.1. Enunciado

Sea  $A$  una matriz de rango  $r$  y sea su SVD completa  $A = \sum_{i=1}^r \sigma_i u_i v_i^\top$ .

Para cualquier  $k < r$ , la matriz de rango  $k$  que minimiza la distancia a  $A$  es:

$$A_k = \sum_{i=1}^k \sigma_i u_i v_i^\top$$

Es decir, para cualquier otra matriz  $B$  con  $\text{rango}(B) \leq k$ , se cumple que:

$$\|A - A_k\| \leq \|A - B\|$$

## 14.0.2. Medida del Error de Aproximación

La calidad de esta aproximación depende de la norma utilizada para medir el error. El teorema es válido para las normas ortogonalmente invariantes más importantes (Strang 2018 Lec. 7 '16; Armentano 2026 Clase 12): <sup>1</sup>

---

<sup>1</sup>En ciencia de datos, la SVD permite comprimir información eliminando los componentes menos significativos (ruido)

**Norma Espectral** El error es exactamente el primer valor singular descartado.

$$\|A - A_k\|_2 = \sigma_{k+1}$$

**Norma de Frobenius** El error es la raíz cuadrada de la suma de los cuadrados de todos los valores singulares descartados.

$$\|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \dots + \sigma_r^2}$$

### 14.0.3. Argumento de la Demostración (para la norma $L_2$ )

La prueba para la norma espectral se basa en demostrar que cualquier matriz  $B$  de rango  $k$  “falla” al aproximar al menos una dirección en el subespacio generado por los primeros  $k + 1$  vectores singulares. (Armentano 2026 Clase 12)

Comenzamos eligiendo un vector  $x$  que esté en el espacio generado por  $\{v_1, \dots, v_{k+1}\}$  y que simultáneamente esté en el núcleo de  $B$ ; lo cual es posible ya que  $\dim(\mathcal{N}(B)) \geq n - k$ . Para este vector unitario  $x$ , se tiene que  $Bx = 0$ , por lo tanto:

$$\|(A - B)x\|_2 = \|Ax\|_2 = \left\| \sum_{i=1}^{k+1} \sigma_i c_i u_i \right\|_2 \geq \sigma_{k+1}$$

Esto garantiza que el error máximo (la norma del operador) no puede ser menor que  $\sigma_{k+1}$ .

### 14.0.4. Demostración para Norma $\ell_2$

(Armentano 2026 Clase 12) Sabemos por la definición de la norma  $\ell_2$  que la aproximación truncada  $A_k = \sum_{i=1}^k \sigma_i u_i v_i^\top$  deja un error residual igual al valor singular más grande de la parte descartada:

$$\|A - A_k\|_2 = \left\| \sum_{i=k+1}^r \sigma_i u_i v_i^\top \right\|_2 = \sigma_{k+1}$$

Para probar que ninguna otra matriz  $B$  de rango  $k$  es mejor que  $A_k$ , consideramos el núcleo de  $B$ , que tiene dimensión al menos  $n - k$ . Al mismo tiempo, tomamos el subespacio generado por los primeros  $k + 1$  vectores singulares derechos de  $A$ , denotado como  $\mathcal{V}_{k+1} = [\{v_1, \dots, v_{k+1}\}]$ , cuya dimensión es  $k + 1$ .

**Existencia de un vector crítico** Dado que la suma de sus dimensiones  $(n - k) + (k + 1) = n + 1$  supera la dimensión total del espacio, debe existir un vector unitario  $x$  que pertenezca a ambos subespacios:

- $Bx = 0$  (por estar en el núcleo de  $B$ ).

- $x = \sum_{i=1}^{k+1} c_i v_i$  con  $\|x\|_2 = 1$  (por estar en  $\mathcal{V}_{k+1}$ ).

**Cálculo de la cota inferior** Evaluamos la norma del error  $(A - B)$  actuando sobre este vector especial  $x$ :

$$\|(A - B)x\|_2^2 = \|Ax - Bx\|_2^2 = \|Ax\|_2^2$$

Sustituyendo la expansión de  $x$  en términos de vectores singulares y recordando que  $Av_i = \sigma_i u_i$ :

$$\|Ax\|_2^2 = \left\| \sum_{i=1}^{k+1} c_i \sigma_i u_i \right\|_2^2 = \sum_{i=1}^{k+1} c_i^2 \sigma_i^2$$

Como los valores singulares están ordenados ( $\sigma_i \geq \sigma_{k+1}$  para todo  $i \leq k + 1$ ), podemos acotar la suma:

$$\sum_{i=1}^{k+1} c_i^2 \sigma_i^2 \geq \sigma_{k+1}^2 \sum_{i=1}^{k+1} c_i^2 = \sigma_{k+1}^2 \|x\|_2^2 = \sigma_{k+1}^2$$

**Conclusión** Puesto que la norma matricial  $\|A - B\|_2$  se define como el máximo valor que puede alcanzar el cociente  $\frac{\|(A-B)x\|}{\|x\|}$ , y acabamos de encontrar un vector  $x$  donde ese cociente es al menos  $\sigma_{k+1}$ , concluimos que:

$$\|A - B\|_2 \geq \sigma_{k+1}$$

Esto demuestra que ninguna matriz de rango  $k$  puede estar más cerca de  $A$  que la aproximación  $A_k$  proporcionada por la SVD.

### 14.0.5. Extensión a la Norma de Frobenius

El Teorema de Eckart-Young también es válido para la **norma de Frobenius**. Esta generalización es fundamental en ciencia de datos, ya que la norma de Frobenius es la métrica estándar para medir el error de reconstrucción en aproximaciones matriciales.

A diferencia de la norma  $L_2$  (que solo considera el valor singular más grande del residuo), la norma de Frobenius depende de todo el espectro de valores singulares descartados. Para la aproximación truncada  $A_k$ , el error es:

$$\|A - A_k\|_F = \sqrt{\sigma_{k+1}^2 + \sigma_{k+2}^2 + \dots + \sigma_r^2}$$

Este valor representa la raíz cuadrada de la “energía” de la señal que se pierde al descartar las componentes menos significativas.

**Invarianza Ortogonal** Multiplicar una matriz por matrices ortogonales ( $Q$  o  $P$ ) no altera su norma de Frobenius, es decir,  $\|QAP\|_F = \|A\|_F$ .

Utilizando la SVD de  $A = U\Sigma V^\top$ , podemos rotar el problema de aproximación hacia las direcciones principales de la matriz. Minimizar  $\|A - B\|_F$  para una matriz  $B$  de rango  $k$  equivale a minimizar  $\|\Sigma - U^\top B V\|_F$ . Dado que  $\Sigma$  es una matriz diagonal con los valores singulares

ordenados, la solución óptima es mantener los  $k$  valores más grandes y descartar el resto, lo que nos devuelve exactamente a la definición de  $A_k$ .<sup>2</sup>

---

<sup>2</sup>Esta propiedad garantiza que la SVD truncada no solo minimiza el error máximo (norma  $L_2$ ), sino que también minimiza el **error cuadrático medio total** de la aproximación. En aplicaciones como la compresión de imágenes, esto asegura que  $A_k$  sea la representación que mejor preserva la varianza y los detalles visuales globales de la imagen original para un presupuesto de rango  $k$  determinado.

# 15 PCA (Análisis de Componentes Principales)

\$

\$

El Análisis de Componentes Principales es una de las herramientas más potentes de la estadística para entender matrices de datos masivos. Su objetivo central es reducir la dimensionalidad de los datos, extrayendo la información más importante mediante la identificación de las direcciones de mayor varianza (Strang 2018 Lec. 7; Armentano 2026 Clase 15).

La potencia del PCA reside en su capacidad para encontrar la “mejor línea” que resume una nube de puntos en un espacio multidimensional. Esta línea no es la misma que se obtiene mediante una regresión por mínimos cuadrados ordinarios, y su distinción es puramente geométrica.

- **Mínimos Cuadrados:** Busca minimizar el error vertical (la distancia en el eje de la variable dependiente) entre el dato y la línea.
- **PCA:** Busca la línea que minimiza la **distancia perpendicular** de cada punto a dicha recta. Es la dirección que maximiza la varianza de las proyecciones de los datos.

En el marco de la SVD, esta “mejor línea” es precisamente la dirección definida por el primer vector singular derecho,  $v_1$ . Al proyectar la matriz de datos centrada  $X$  sobre esta dirección, obtenemos la matriz de rango 1,  $\tilde{X}_1$ , que es la sombra más informativa de los datos originales.

**Centrado de Datos** Antes de aplicar PCA, es imperativo que los datos tengan **media cero**.

Si partimos de una matriz de datos  $A$  donde las filas representan muestras y las columnas variables, restamos a cada fila el promedio de todas las filas. Este paso centra la nube de puntos en el origen, permitiendo que las componentes principales capturen la estructura de la varianza intrínseca y no el desplazamiento del conjunto.

**SVD** Una vez centrados los datos en una matriz  $X \in \mathbb{R}^{n \times d}$  (donde  $n$  es el número de individuos y  $d$  el de variables), utilizamos la Descomposición en Valores Singulares (SVD):  $X = U\Sigma V^T$ .

**Componentes Principales** Los vectores de la base ortonormal  $V = [v_1, v_2, \dots, v_d]$  son las componentes principales. El vector  $v_1$  apunta en la dirección de la “mejor línea”, aquella que minimiza la suma de los cuadrados de las distancias perpendiculares desde los puntos a la línea.

**Varianza y Valores Singulares** La importancia de cada componente está determinada por su valor singular  $\sigma_i$ . La dirección  $v_1$  asociada al valor singular más grande  $\sigma_1$  es la que explica la mayor cantidad de varianza de los datos.

### 15.0.1. Ejemplo

Para ilustrar el funcionamiento del PCA, consideramos una matriz de datos  $X \in \mathbb{R}^{n \times d}$  con  $d = 2$  variables (por ejemplo, altura y peso) y  $n$  individuos (Armentano 2026 Clase 15).

$$X = \begin{pmatrix} X_{11} & X_{12} \\ X_{21} & X_{22} \\ \vdots & \vdots \\ X_{n1} & X_{n2} \end{pmatrix}$$

Tras centrar los datos, aplicamos la Descomposición en Valores Singulares (SVD), lo que nos permite descomponer la matriz original en una suma de matrices de rango 1:

$$X = \tilde{X}_1 + \tilde{X}_2 = \sigma_1 u_1 v_1^\top + \sigma_2 u_2 v_2^\top$$

Cada una de estas matrices  $\tilde{X}_i$  representa una “capa” de la información total del conjunto de datos, organizada por orden de importancia según la magnitud de sus valores singulares ( $\sigma_1 \geq \sigma_2$ ).

$\tilde{X}_1$  (Componente Principal 1) Es la matriz que captura la mayor varianza. Geométricamente, todos los puntos (filas) de  $\tilde{X}_1$  se encuentran perfectamente alineados sobre la recta definida por el primer vector singular derecho  $v_1$ .

$\tilde{X}_2$  (Componente Principal 2) Captura la varianza remanente en la dirección ortogonal a  $v_1$ , definida por  $v_2$ .

La relación entre la matriz original  $X$  y su aproximación  $\tilde{X}_1$  se define mediante la **proyección ortogonal**. Si denotamos como  $X_i$  a la fila  $i$ -ésima de la matriz de datos, su proyección sobre la dirección del vector  $v_1$  es:

$$\tilde{X}_{i,1} = P_{v_1}(X_i) = (X_i \cdot v_1) v_1^\top$$

Dado que  $v_1$  es un vector singular, esta expresión vale  $\sigma_1 u_{i,1} v_1^\top$  que a su vez corresponde a  $\tilde{X}_i$

Donde  $u_{i,1}$  es la coordenada del  $i$ -ésimo individuo en la escala del primer componente principal.

Esto significa que cada fila de  $\tilde{X}_1$  es el punto más cercano sobre la “mejor línea” ( $v_1$ ) al dato original  $X_i$ . En otras palabras,  $\tilde{X}_1$  no es simplemente una versión “reducida” de los datos, sino la sombra geométrica más fiel de la nube de puntos sobre su dirección principal.

# 16 Minimos Cuadrados

§

§

En el análisis de datos, a menudo nos enfrentamos a la tarea de representar un fenómeno mediante un modelo matemático. La diferencia entre el fenómeno real y nuestro modelo se denomina error de representación.

Supongamos que tenemos  $m$  puntos de datos  $(x_i, y_i)$ . Si buscamos un polinomio de grado  $m - 1$  que pase exactamente por todos los puntos, estamos realizando una **interpolación** (Armentano 2026 clase 17). Esto requiere resolver un sistema lineal  $Vc = y$ , donde  $V$  es la matriz de Vandermonde:

$$V = \begin{pmatrix} 1 & x_1 & \dots & x_1^{m-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_m & \dots & x_m^{m-1} \end{pmatrix}$$

Si los valores de  $x_i$  son distintos,  $V$  es invertible y existe una solución única. Sin embargo, la interpolación exacta tiene riesgos: si los datos tienen ruido, el polinomio puede presentar oscilaciones salvajes (como el fenómeno de Runge), lo que invalida su capacidad predictiva.

Cuando el número de datos  $m$  es mucho mayor que el número de parámetros  $n$  del modelo (sistema sobredeterminado), generalmente no existe una solución exacta para  $Ax = b$  (Strang 2018 Lec. 9). En lugar de forzar un ajuste exacto en pocos puntos, buscamos una solución aproximada  $\hat{x}$  que sea “lo más cercana posible” a todos los datos simultáneamente.

El criterio estándar, propuesto originalmente por Gauss, consiste en **minimizar la norma  $\ell_2$  del residuo**:

$$\min \|Ax - b\|_2^2 = \min \sum_{i=1}^m (ax_i - b_i)^2$$

Este cambio de paradigma transforma un problema sin solución (en el sentido estricto) en un problema de optimización con una solución robusta y única, sentando las bases de la regresión lineal y el aprendizaje automático moderno.

## 16.1. Las Ecuaciones Normales

Cuando un sistema  $Ax = b$  es sobredeterminado ( $m > n$ ), el vector  $b$  generalmente no pertenece al espacio columna de  $A$ , por lo que no existe una solución exacta. El objetivo de los mínimos cuadrados es encontrar un vector  $\hat{x}$  que minimice el error cuadrático total:

$$\hat{x} = \operatorname{argmin} \|Ax - b\|_2^2$$

Geoméricamente, la distancia más corta desde el vector  $b$  hasta el subespacio  $C(A)$  se encuentra mediante una **proyección ortogonal**. Para que el error  $e = b - A\hat{x}$  sea mínimo, este debe ser perpendicular a todas las columnas de  $A$ :

La condición de ortogonalidad establece que el producto punto de cada columna de  $A$  con el error debe ser cero:  $A^\top \cdot e = 0$ . Entonces  $A^\top \cdot (b - A\hat{x}) = 0$ . Distribuyendo los términos, obtenemos las **Ecuaciones Normales**:

$$A^\top A\hat{x} = A^\top b \tag{16.1}$$

La matriz  $A^\top A$  es fundamental en este proceso. Es una matriz cuadrada ( $n \times n$ ), simétrica y semidefinida positiva.

Si las columnas de  $A$  son linealmente independientes, entonces la matriz  $A^\top A$  es invertible y la solución  $\hat{x}$  es única:

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

Si las columnas son dependientes, existen infinitas soluciones que minimizan el error, y suele seleccionarse la de norma mínima mediante la pseudoinversa.

Este método transforma un problema de aproximación en un sistema lineal cuadrado que puede resolverse mediante algoritmos estándar como la eliminación gaussiana.

## 16.2. Proyecciones y Ortogonalidad

La solución del problema de mínimos cuadrados tiene una interpretación geométrica elegante: estamos buscando el punto en el subespacio definido por el espacio columna de  $A$  que se encuentra a la distancia mínima del vector de datos  $b$  (Strang 2018 Lec. 9; Strang 2019, II.2; Armentano 2026 Clase 17).

Dado que  $b$  generalmente no está en  $C(A)$ , no existe un vector  $x$  tal que  $Ax = b$ . La mejor aproximación es el vector  $p = A\hat{x}$ , que es la **proyección ortogonal** de  $b$  sobre el plano (o hiperplano) formado por las columnas de  $A$ .

El vector error (o residuo)  $e = b - A\hat{x}$  representa la parte de los datos que el modelo no puede explicar. Para que la distancia  $\|e\|$  sea mínima, el vector de error debe ser perpendicular a todas las columnas de  $A$ . Esto significa que  $e$  pertenece al espacio nulo de  $A^\top$  ( $\mathcal{N}(A^\top)$ ).

Podemos expresar la proyección  $p$  directamente a partir de  $b$  mediante una matriz  $P$  tal que  $p = Pb$ . Al sustituir la solución de las ecuaciones normales (Ecuación 16.1) en  $p = A\hat{x}$ , obtenemos:

$$P = A(A^\top A)^{-1}A^\top$$

**Propiedades de  $P$ :**

1. **Simetría:**  $P^\top = P$
2. **Idempotencia:**  $P^2 = P$
3. **Caso Ortonormal:** Si las columnas de  $A$  son ortonormales ( $A = Q$ ), la matriz de proyección se simplifica drásticamente a  $P = QQ^\top$  (Armentano 2026 Clase 5; Strang 2018 Lec. 3).

### 16.3. Solución vía SVD La Pseudoinversa de Moore-Penrose $A^+$

La Descomposición en Valores Singulares proporciona la herramienta definitiva para resolver el problema de mínimos cuadrados, especialmente en casos donde la matriz  $A$  es singular o tiene columnas dependientes. En estos escenarios, las ecuaciones normales fallan porque  $A^\top A$  no es invertible. La solución es la **pseudoinversa**  $A^+$ .

Si conocemos la SVD de una matriz,  $A = U\Sigma V^\top$ , su pseudoinversa se construye invirtiendo los componentes de la descomposición:

$$A^+ = V\Sigma^+U^\top$$

Donde  $\Sigma^+$  es una matriz de dimensiones  $n \times m$  que contiene los recíprocos de los valores singulares no nulos ( $1/\sigma_i$ ) en su diagonal, y ceros en todas las demás entradas.

La importancia de la pseudoinversa radica en que siempre proporciona una solución válida al problema mín  $\|Ax - b\|_2^2$ , denotada como  $x^+ = A^+b$ :

Si las columnas de  $A$  son independientes,  $A^+$  coincide con el operador de las ecuaciones normales:  $A^+ = (A^\top A)^{-1}A^\top$ . En este caso,  $x^+$  es la única solución de mínimos cuadrados.

Si  $A$  tiene columnas dependientes, existen infinitos vectores  $x$  que minimizan el error. La pseudoinversa selecciona automáticamente el **vector de norma mínima** (mín  $\|x\|_2$ ).

Geoméricamente,  $A^+$  actúa como un puente entre los cuatro subespacios fundamentales:

- Lleva el espacio columna  $C(A)$  de vuelta al espacio fila  $C(A^\top)$  de forma perfecta.
- Cualquier componente del vector  $b$  que se encuentre en el espacio nulo de  $A^\top$  ( $\mathcal{N}(A^\top)$ ) es enviado directamente a cero por  $A^+$ , eliminando el ruido que no puede ser explicado por el modelo.

## 16.4. Ortogonalización: El método de Gram-Schmidt

Una tercera vía para resolver el problema de mínimos cuadrados consiste en transformar las columnas de la matriz  $A$  en una base ortonormal. Este proceso, conocido como **Gram-Schmidt**, permite factorizar la matriz como  $A = QR$ . (Strang 2018 Lec. 11; Armentano 2026 Clase 17)

<sup>1</sup>

- $Q$  Es una matriz de Stiefel  $m \times n$  (columnas ortonormales entre sí, es decir,  $Q^\top Q = I$ ). Geométricamente,  $Q$  abarca el mismo espacio columna que  $A$ , pero sin redundancias ni problemas de escala.
- $R$ : Es una matriz triangular superior que contiene los coeficientes de las combinaciones lineales que vinculan a  $A$  con  $Q$ . Al ser triangular, facilita enormemente la resolución de sistemas mediante sustitución hacia atrás.

Sustituyendo  $A = QR$  en las ecuaciones normales ( $A^\top A \hat{x} = A^\top b$ ):

$$\begin{aligned}(QR)^\top(QR)\hat{x} &= (QR)^\top b \\ R^\top(Q^\top Q)R\hat{x} &= R^\top Q^\top b\end{aligned}$$

Como  $Q^\top Q = I$  y asumiendo que  $R$  es invertible (columnas de  $A$  independientes), podemos simplificar el sistema a:

$$R\hat{x} = Q^\top b$$

---

<sup>1</sup>A diferencia de las ecuaciones normales, que requieren calcular  $A^\top A$  (lo cual puede duplicar el error de redondeo si la matriz está mal condicionada), el método  $QR$  es numéricamente más estable. Al operar directamente sobre  $Q$  y  $R$ , se evita la “pérdida de precisión” asociada al cuadrado de los valores singulares, convirtiéndolo en el estándar de la industria para sistemas sobredeterminados de tamaño moderado.

# Referencias

- Armentano, Diego. 2026. *Notas de clase del curso Geometría y Álgebra Lineal para Aprendizaje Estadístico*. Facultad de Ciencias Económicas y Administración, Universidad de la República. <https://github.com/diegoax/ALNAE-2026/tree/main/Notebooks>.
- Diego Armentano, Manuel Hernández. 2025. *Pruebas parciales de Álgebra Lineal para Aprendizaje Estadístico*. Facultad de Ciencias Económicas y Administración, Universidad de la República. <https://github.com/diegoax/ALNAE-2026/tree/main>.
- Hernández, Manuel. 2026. *Ejercicios practicos del curso Geometría y Álgebra Lineal para Aprendizaje Estadístico*. Facultad de Ciencias Económicas y Administración, Universidad de la República. <https://github.com/diegoax/ALNAE-2026/tree/main/practicos>.
- Strang, Gilbert. 2018. «Matrix Methods in Data Analysis, Signal Processing, and Machine Learning». Massachusetts Institute of Technology: MIT OpenCourseWare. <https://ocw.mit.edu/courses/18-065-matrix-methods-in-data-analysis-signal-processing-and-machine-learning-spring-2018/>.
- Strang, Gilbert. 2019. *Linear Algebra and Learning from Data*. Wellesley-Cambridge Press. <https://math.mit.edu/~gs/learningfromdata/>.

# Espectro de Matrices Aleatorias

El estudio de matrices con coeficientes aleatorios permite entender cuán “comunes” son ciertas propiedades estructurales en el álgebra lineal. Si generamos una matriz  $A \in \mathbb{R}^{n \times n}$  con entradas distribuidas continuamente (por ejemplo, mediante una distribución normal), podemos predecir su comportamiento espectral. (Strang 2019, I.6; Armentano 2026 clase 7)

## Existencia de Autovectores en el Plano Complejo

Desde un punto de vista algebraico, la probabilidad de que una matriz cuadrada tenga autovectores es del 100% siempre que consideremos el campo de los números complejos  $\mathbb{C}$ . Esto se debe a que el polinomio característico  $\chi_A(\lambda) = \det(A - \lambda I)$  es de grado  $n$  y, por el Teorema Fundamental del Álgebra, posee exactamente  $n$  raíces.

## La Probabilidad de Autovalores Reales

Si generamos una matriz  $A \in \mathbb{R}^{n \times n}$  con entradas aleatorias (distribuidas de forma continua), la probabilidad de que sus autovalores sean reales depende fundamentalmente de la dimensión  $n$  y de las raíces de su polinomio característico.

## Perspectiva del Polinomio Característico

Los autovalores son las raíces de  $\chi_A(\lambda) = \det(A - \lambda I) = 0$ .

- **Dimensión ( $n$ ) Impar:** La probabilidad de tener al menos un autovalor real es **1**. Esto ocurre porque cualquier polinomio de grado impar con coeficientes reales tiende a  $+\infty$  y  $-\infty$  en sus extremos, lo que garantiza que cruce el eje real al menos una vez.
- **Dimensión ( $n$ ) Par:** La probabilidad es **menor a 1**. Existe una probabilidad no nula de que todas las raíces sean pares de complejos conjugados, lo que resultaría en una matriz sin autovalores reales.

## Perspectiva del Determinante

El determinante es el producto de los autovalores:  $\det(A) = \lambda_1 \lambda_2 \dots \lambda_n$ . Para matrices aleatorias, la probabilidad de que  $\det(A) = 0$  es cero. Esto implica que la probabilidad de que el cero sea un autovalor es nula, asegurando que los autovalores reales existentes sean casi seguramente no nulos.

## Diagonalizabilidad casi segura

Para una matriz aleatoria, la probabilidad de que el polinomio característico tenga raíces múltiples es cero. Por lo tanto, si los autovalores son reales, la matriz es **diagonalizable casi con seguridad**. Esto tiene implicaciones fundamentales para la estructura de la matriz:

- **Autovalores distintos:** Casi con seguridad, la matriz tendrá  $n$  autovalores diferentes.
- **Independencia Lineal:** Una matriz con  $n$  autovalores distintos garantiza que sus autovectores asociados son linealmente independientes.

### **i** Nota

Una matriz aleatoria es **diagonalizable casi con seguridad** (con probabilidad 1), permitiendo siempre la factorización  $A = X\Lambda X^{-1}$ .

## El Determinante y la Invertibilidad

La probabilidad de que una matriz aleatoria sea singular es cero. En términos espectrales, esto significa que el valor  $\lambda = 0$  casi nunca es una raíz del polinomio característico  $\det(A - \lambda I) = 0$ .

Esta propiedad asegura que los algoritmos numéricos que dependen de la invertibilidad o del cálculo de potencias  $A^k$  (con  $k$  negativo o positivo) sean estables frente a perturbaciones aleatorias.

## Resumen de Probabilidades Espectrales

| Propiedad                                 | Probabilidad  | Razón  |
|---|---------------|--|
| Diagonalizable                            | <b>1</b>      | Los autovalores son casi siempre distintos           |
| Invertible $\det(A) \neq 0$               | <b>1</b>      | El conjunto de matrices singulares tiene medida cero |
| Autovectores en $\mathbb{R}^n$ ( $n$ par) | <b>&lt; 1</b> | Posibilidad de autovalores complejos conjugados      |
| Autovectores en $\mathbb{C}^n$            | <b>1</b>      | Teorema Fundamental del Álgebra                      |

## Ejemplos Conceptuales

**Matrices de Rotación** Una matriz de rotación pura en  $\mathbb{R}^2$  tiene la forma:  $Q = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix}$ .

Sus autovalores son complejos ( $e^{\pm i\theta}$ ), ilustrando por qué en dimensiones pares la probabilidad de autovalores reales no es la unidad.

**Matrices Simétricas** Si restringimos la matriz aleatoria para que sea simétrica ( $A = A^T$ ), la probabilidad de tener todos sus autovalores reales es **1**, independientemente de  $n$  (Armentano 2026 clase 7).

**Matrices de Markov** En una matriz aleatoria de Markov (columnas suman 1), siempre existe el autovalor real  $\lambda = 1$ .

## Factorización $LU$

La eliminación gaussiana es el algoritmo fundamental para la resolución de sistemas lineales  $Ax = b$ . Desde la perspectiva del álgebra lineal numérica, este proceso se interpreta como la factorización de la matriz original en dos componentes triangulares, permitiendo una comprensión profunda de la propagación de la información dentro del sistema.

Un paradigma moderno y sumamente útil consiste en visualizar cada etapa de la eliminación gaussiana como la resta de una **matriz de rango 1** a la matriz de trabajo actual.

Sea  $A \in \mathbb{R}^{n \times n}$ . La matriz puede expresarse como la suma de  $n$  matrices de rango 1, donde cada término representa el aporte de un paso de eliminación:

$$A = \sum_{i=1}^n \ell_i u_i^T = \ell_1 u_1^T + \ell_2 u_2^T + \dots + \ell_n u_n^T$$

donde:

- $\ell_i$  es un vector columna que contiene los multiplicadores aplicados debajo del  $i$ -ésimo pivote.
- $u_i^T$  es el vector fila correspondiente a la fila pivote original en esa etapa.

Este enfoque revela que la eliminación reduce sistemáticamente la complejidad del sistema, “extrayendo” una dirección independiente en cada paso hasta agotar el rango de la matriz.

## Factorización $A = LU$

Si el proceso de eliminación se completa sin necesidad de intercambios de filas, la matriz  $A$  admite una factorización única de la forma:

$$A = LU$$

### Matriz $L$ (Lower):

Es una matriz **triangular inferior** con unidades en la diagonal principal ( $L_{ii} = 1$ ). Sus entradas por debajo de la diagonal son precisamente los multiplicadores utilizados para eliminar las variables.

### Matriz $U$ (Upper):

Es una matriz **triangular superior** que contiene los coeficientes resultantes tras completar la eliminación. Sus entradas diagonales son los **pivotes**.

La matriz  $L$  se construye agrupando los vectores columna  $\ell_i$ , y la matriz  $U$  agrupando los vectores fila  $u_i^T$ :

$$L = [\ell_1, \ell_2, \dots, \ell_n], \quad U = \begin{bmatrix} u_1^T \\ u_2^T \\ \vdots \\ u_n^T \end{bmatrix}$$

Es vital notar que para que  $L$  tenga unos en la diagonal, los vectores  $\ell_i$  deben estar normalizados respecto al valor del pivote en ese paso.

## Factorización $PA = LU$

En la práctica numérica, la eliminación puede fallar si se encuentra un cero en la posición del pivote, o ser inestable si el pivote es muy pequeño. Para solventar esto, se recurre al **pivoteo parcial**.

### Teorema de Existencia General:

Para toda matriz  $A$  invertible, existe una **matriz de permutación**  $P$  tal que la matriz reordenada admite una factorización  $LU$ :

$$PA = LU$$

La matriz  $P$  simplemente registra los intercambios de filas necesarios para asegurar que los pivotes sean los máximos posibles en valor absoluto, minimizando el error de redondeo.

## Complejidad Computacional

La multiplicación de matrices y su factorización involucran un conteo de operaciones fundamental para el análisis de algoritmos. Para matrices de tamaño  $n \times n$ :

1. **Eliminación:** El costo es de aproximadamente  $\frac{1}{3}n^3$  multiplicaciones y sumas.
2. **Resolución:** Una vez obtenida la forma  $LU$ , resolver  $Ly = b$  (sustitución hacia adelante) y  $Ux = y$  (sustitución hacia atrás) requiere solo  $n^2$  operaciones.

Esta asimetría en el costo subraya la importancia de la factorización: una vez que  $A$  es factorizada, resolver para múltiples vectores  $b$  es extremadamente eficiente.

# Diagonalización

Una matriz cuadrada  $A$  de tamaño  $n \times n$  se define como diagonalizable si puede descomponerse en la forma  $A = X\Lambda X^{-1}$ . En esta estructura,  $X$  es la matriz que contiene los autovectores en sus columnas y  $\Lambda$  es la matriz diagonal donde se ubican los autovalores correspondientes.

Para que esta factorización exista, la matriz debe cumplir con una condición necesaria: tener un conjunto completo de  $n$  autovectores linealmente independientes que formen una base del espacio.

## Definiciones Previas

**Multiplicidad Algebraica** Es el número de veces que  $\lambda$  aparece como raíz del polinomio característico.

**Multiplicidad Geométrica** Es la dimensión del espacio propio (el kernel o núcleo de la matriz  $A - \lambda I$ ). Representa cuántos vectores propios independientes “nacen” de ese valor propio.

## Criterio de Diagonalización

Para que una matriz sea diagonalizable, cada uno de sus valores propios debe aportar tantos vectores propios linealmente independientes como su repetición en el polinomio característico indique. Si algún valor propio “falla” en aportar suficientes vectores, la matriz es defectuosa y no puede diagonalizarse.

Una matriz  $n \times n$  es diagonalizable si y solo si la suma de las multiplicidades geométricas de todos sus valores propios es igual a  $n$ .

En términos técnicos: Una matriz es diagonalizable si, para cada valor propio  $\lambda_i$ , se cumple que:

$$m_g(\lambda_i) = m_a(\lambda_i)$$

## Explicación Intuitiva: La Falta de “Direcciones”

Para diagonalizar una matriz  $3 \times 3$ , necesitas “estirar” el espacio en 3 direcciones independientes (los 3 vectores propios). La matriz de paso  $P$  en la ecuación  $A = PDP^{-1}$  se construye colocando esos vectores en las columnas. Si  $m_g = 1$ , la matriz solo tiene una dirección de estiramiento. Al intentar construir  $P$ , te faltarían dos columnas para que la matriz sea invertible. Matemáticamente, estás intentando representar una transformación 3D usando la información de una sola dimensión; la información restante está “atrapada” en bloques de Jordan y no se puede separar en una diagonal limpia.

**¿Por qué ocurre esto?** Cuando  $m_g < m_a$ , la matriz no solo escala vectores (que es lo que hace una matriz diagonalizable), sino que también realiza **cizallamientos** (*shears*). El cizallamiento es una operación lineal que no puede reducirse a una simple base de vectores propios, lo que impide la estructura diagonal.

### ¿ Puede una matriz 3x3 no nula tener sus 3 valores propios iguales a 0 ?

Este fenómeno ocurre específicamente en las matrices nilpotentes. Para que esto suceda, la matriz debe ser defectuosa, lo que significa que no posee un conjunto completo de vectores propios linealmente independientes y, por lo tanto, no es diagonalizable.

El ejemplo más sencillo es una matriz en forma normal de Jordan que consista en un único bloque de Jordan asociado al valor propio cero. Al ser una matriz triangular superior, sus valores propios son los elementos de la diagonal principal, todos los cuales son 0.

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

Para que una matriz no nula tenga todos sus valores propios en cero, debe haber una discrepancia entre  $m_a$  y  $m_g$ : En este caso,  $(-\lambda)^3 = 0$  implica que el 0 se repite 3 veces ( $m_a = 3$ ).

Si  $m_g < m_a$ , la matriz no puede ser la matriz nula (la cual tiene  $m_g = 3$ ). En el ejemplo, el rango de la matriz es 2, por lo que la dimensión del núcleo (el espacio propio para  $\lambda = 0$ ) es  $3 - 2 = 1$ . Como  $1 < 3$ , la matriz es necesariamente no nula y no diagonalizable.

Como  $m_g = 1$  y  $m_a = 3$ , no hay suficientes vectores propios para construir una base de  $\mathbb{R}^3$ .

# Sintesis

\$

\$

## Vectores Especiales

|                         |                                      |   |
|-------------------------|--------------------------------------|---|
| $v, u \in \mathbb{R}^n$ | $\in$                                |   |
| $vv^T$                  | $\mathcal{M}_{n \times n}$ (rango 1) | $P_v$   |
| $v^T v$                 | $\mathbb{R}$                         | $\langle v, v \rangle = \ v\ ^2$  |
| $vu^T$                  | $\mathcal{M}_{n \times n}$ (rango 1) | descomposicion RC   |
| $v^T u$                 | $\mathbb{R}$                         | $\chi_A(\lambda) = (-\lambda)^{n-1}(v^T u - \lambda)$<br>$\langle v, u \rangle = \ v\  \ u\  \cos \theta$ |

## Matrices Especiales

|           |   |  |
|-----------|---|--|
| Stiefel   | $Q \in \mathcal{M}_{m \times n}$<br>$Q^T Q = I$           | <ul style="list-style-type: none"> <li>▪ Isometría</li> <li>▪ Columnas ortonormales</li> <li>▪ <math>QQ^T = P_Q</math></li> <li>▪ Idempotente, Invariante en <math>Q</math></li> </ul>   |
| Ortogonal | $Q \in \mathcal{M}_{n \times n}$                          | <ul style="list-style-type: none"> <li>▪ Caso particular de Stiefel, ademas cuadrada</li> <li>▪ Autovectores <math>\in \mathbb{C}^n</math> ortogonales (<math>v^* w = 0</math>)</li> <li>▪ <math>Q^{-1} = Q^T</math></li> <li>▪ <math>\lambda_i \in \mathbb{C},  \lambda_i  = 1</math></li> <li>▪ Isometria (preserva normas y angulos)</li> <li>▪ si <math>\lambda \in \mathbb{R}</math> entonces <math>\lambda = \pm 1</math></li> <li>▪ Rotacion <math>\det(Q) = 1</math>, o Reflexion <math>\det(Q) = -1</math></li> </ul> |
| Simétrica | $S \in \mathcal{M}(\mathbb{R})_{n \times n}$<br>$S = S^T$ | <ul style="list-style-type: none"> <li>▪ Autovectores ortogonales</li> <li>▪ <math>\lambda \in \mathbb{R}</math></li> <li>▪ <math>S = Q \Lambda Q^T</math></li> </ul>  |

|                       |   |  |
|-----------------------|---|--|
| Simétrica y Ortogonal | $S \in \mathcal{M}(\mathbb{R})_{n \times n}$<br>$S = S^\top$<br>$S^{-1} = S^\top$ | <ul style="list-style-type: none"> <li>▪ <math>\lambda = \pm 1</math></li> </ul>   |
| Hermítica             | $S \in \mathcal{M}(\mathbb{C})_{n \times n}$<br>$S = S^* = \overline{S}^\top$     | <ul style="list-style-type: none"> <li>▪ Autovectores ortogonales (en <math>\mathbb{C}</math>, <math>v^*w = 0</math>)</li> <li>▪ <math>\lambda \in \mathbb{C}</math></li> <li>▪ <math>S = Q\Lambda Q^\top</math></li> <li>▪ <math>\lambda</math> imaginario puro o cero</li> </ul> |
| Antisimétrica         | $A$<br>$A^\top = -A$  | <ul style="list-style-type: none"> <li>▪ <math>\lambda</math> imaginario puro o cero</li> </ul>  |
| Householder           | $H$   | <ul style="list-style-type: none"> <li>▪ <math>H = (I - 2vv^\top)u</math></li> <li>▪ Simétrica (proyección ortogonal)</li> <li>▪ Involutiva (<math>H^2 = \mathbb{I}</math>)</li> <li>▪ Ortogonal (isometría)</li> </ul>  |
| Proyección            | $QQ^\top$   | <ul style="list-style-type: none"> <li>▪ <math>\geq 0</math></li> <li>▪ Simétrica</li> <li>▪ Idempotente</li> <li>▪ No Invertible, No Ortogonal</li> </ul>   |

## Proyecciones y Reflexiones

|  |                                |
|--|--------------------------------|
| Proyección de $u$ sobre $v$                  | $u_v = (vv^\top)u$             |
| Proyección de $u$ sobre $v^\perp$            | $u_{v^\perp} = (I - vv^\top)u$ |
| Proyección de $u$ sobre $Q$                  | $QQ^\top u$                    |
| Reflección de $u$ sobre hiperplano $v^\perp$ | $H = (I - 2vv^\top)u$          |

## Factorizaciones

|                       | Hipótesis  | Origen            |
|-----------------------|--|-------------------|
| $A = CR$              |  |                   |
| $A = X\Lambda X^{-1}$ | $A \in \mathcal{M}_{n \times n}$ con $n$ Autovectores l.i. | $AX = X\Lambda$   |
| $S = Q\Lambda Q^\top$ | $S = S^\top$ , $S \in \mathcal{M}_{n \times n}$            | Teorema Espectral |
| $A = U\Sigma V^\top$  |  |                   |

## Normas

$$\|A\|_N = \sigma_1 + \dots + \sigma_r$$

$$\|A\|_F = (\sigma_1^2 + \dots + \sigma_r^2)^{\frac{1}{2}}$$

$$\|A\|_2 = \sigma_1$$

---

---

# Problemas de Optimización

§

§

En problemas de gran escala o con datos ruidosos, la matriz  $A$  puede estar mal condicionada o ser casi singular, lo que provoca que las soluciones de mínimos cuadrados convencionales sean inestables o demasiado complejas. La regularización introduce una penalización por la magnitud de los coeficientes para obtener modelos más robustos y parsimoniosos.

En el análisis de datos, a menudo buscamos soluciones a sistemas  $Ax = b$  que no solo minimicen el error, sino que posean estructuras específicas como la **dispersión** (*sparsity*). Aquí es donde la elección de la norma en el problema de optimización redefine el resultado (Strang 2018 Lec. 7; Strang 2018 Lec. 8; Strang 2018 Lec. 11).

La ventaja fundamental de estos métodos sobre los mínimos cuadrados de Gauss es la **interpretabilidad**. Mientras que una solución con muchos valores pequeños es difícil de explicar, una solución con pocos valores distintos de cero identifica directamente los factores clave que explican el fenómeno.

## Basis Pursuit

El problema de **Basis Pursuit** busca encontrar el vector con la menor norma  $\ell_1$  que satisfaga **exactamente** el sistema de ecuaciones lineales:

$$\text{mín } \|x\|_1 \quad \text{sujeto a } Ax = b$$

A diferencia de la solución de norma mínima  $\ell_2$  (producida por la pseudo-inversa), que tiende a distribuir el “peso” entre todas las componentes, Basis Pursuit “empuja” la solución hacia los ejes coordenados. Geométricamente, esto ocurre porque la bola unitaria de  $\ell_1$  es un diamante; al expandirse, es mucho más probable que toque el hiperplano de restricción  $Ax = b$  en un vértice (donde muchas componentes son cero).

## Lasso

En situaciones con ruido donde  $Ax = b$  no puede cumplirse exactamente, se utiliza el **Lasso** (*Least Absolute Shrinkage and Selection Operator*). Este método añade una penalización  $\ell_1$  al problema clásico de mínimos cuadrados:

$$\text{mín } \|Ax - b\|_2^2 + \lambda \|x\|_1$$

En esta expresión:  $\|Ax - b\|_2^2$  Mide el ajuste a los datos (fidelidad).  $\lambda \|x\|_1$ : Promueve la simplicidad del modelo (dispersión).

El parámetro  $\lambda$  controla el equilibrio entre el ajuste y la cantidad de ceros en la solución. Al aumentar  $\lambda$ , el modelo se vuelve más “parsimonioso”, seleccionando solo las variables (columnas de  $A$ ) más significativas.

- **Promoción de la Dispersión (*Sparsity*):** A diferencia de la bola  $\ell_2$ , que es redonda, la geometría de la norma  $L_1$  es la de un diamante con vértices sobre los ejes. Geométricamente, al expandir esta bola de penalización, es mucho más probable que toque el espacio de soluciones en un vértice, lo que fuerza a que muchos coeficientes sean exactamente cero.
- **Interpretabilidad:** Al producir vectores con pocos componentes no nulos, Lasso actúa como un selector de variables automático. Esto es crucial en ciencia de datos, ya que permite identificar los factores clave que explican un fenómeno en lugar de tener una mezcla ininteligible de miles de pequeños coeficientes.

## Ridge Regression

Este método, también conocido como regresión de cresta o penalización de Tikhonov, añade el cuadrado de la norma  $L_2$  al error residual:

$$\text{mín } \|Ax - b\|_2^2 + \delta^2 \|x\|_2^2$$

- **Estabilidad Numérica:** El término  $\delta^2 I$  asegura que la matriz del sistema ( $A^\top A + \delta^2 I$ ) sea estrictamente definida positiva e invertible, incluso si  $A^\top A$  es singular. Esto protege el cálculo contra errores de redondeo masivos cuando los valores singulares de  $A$  son cercanos a cero.
- **Relación con la Pseudoinversa:** Un resultado matemático profundo establece que, a medida que el parámetro de penalización  $\delta$  tiende a cero, la solución de Ridge converge exactamente a la solución de norma mínima proporcionada por la pseudoinversa Moore-Penrose:  $x^+ = A^+ b$ .