MA–460: ALGEBRA LINEAL II

Joseph C. Várilly

Escuela de Matemática, Universidad de Costa Rica

I Ciclo Lectivo de 2020

Introducción

El álgebra lineal comprende el estudio de los espacios vectoriales y las aplicaciones lineales entre ellos. Un espacio vectorial finitodimensional tiene una estructura sencilla: todos sus vectores son combinaciones lineales de un número finito de vectores básicos. Una aplicación lineal de un espacio vectorial en otro está determinado por sus valores sobre una base del primero, expresados por una matriz de coeficientes en una base del segundo. En consecuencia, hay una estrecha relación entre las propiedades estructurales de las aplicaciones lineales y los algoritmos para manipular matrices.

Este es un segundo curso de álgebra lineal. En el curso anterior, los espacios vectoriales y las matrices fueron introducidos en el contexto, clásico y fundamental, de la resolución de sistemas de ecuaciones de primer grado en varias variables. También se adquirió familiaridad con los conceptos esenciales de base y dimensión de un espacio vectorial, núcleo e imagen de una aplicación lineal, espacio vectorial dual y el teorema de rango y nulidad.

El estudio del álgebra lineal comprende aspectos tanto *estructurales* como *algorítmicos*. En este segundo curso, el énfasis recaerá sobre las estructuras, sean ellas de las aplicaciones lineales, de los espacios vectoriales dotados de un producto escalar, o de las formas bilineales y cuadráticos. Aun así, para entender bien esta teoría, hay que prestar la debida atención a su presentación en algoritmos y a los métodos explícitos de cálculo.

Al inicio, se hará un breve repaso de los temas del curso anterior: vectores, aplicaciones lineales, matrices, determinantes. Luego, se abordará la búsqueda de los autovalores y autovectores de una matriz (o de una aplicación lineal), con el afán de transformar una matriz dada a una forma diagonal. Esto no siempre se puede lograr: es necesario examinar en detalle la estructura de una aplicación lineal, para determinar cuando esta "diagonalización" es factible y qué hacer en el caso contrario.

Un concepto fundamental es el de *ortogonalidad*. En presencia de un producto escalar sobre un espacio vectorial real o complejo, las aplicaciones lineales se reparten en diversos clases: ortogonales o unitarios, simétricos o hermíticos, definidas positivas. Tales clasificaciones dan lugar a diversos factorizaciones de matrices, entre ellas la llamada descomposición polar, cuyos factores admiten una descripción en términos de sus autovalores y autoespacios mediante el *teorema espectral*.

Las *formas bilineales* sobre un espacio vectorial real o complejo se clasifican de diversas maneras. Las formas antisimétricas se caracterizan por su rango; las formas simétricas, por su rango y signatura. Las aplicaciones lineales que conservan una determinada forma bilineal forman grupos notables de matrices invertibles.

Muchas aplicaciones concretas del álgebra lineal dependen de propiedades especiales de varias clases de matrices. En la última parte del curso, se examinará algunas ejemplos de este fenómeno, conocido hoy en día como la *teoría de matrices*.

Programa de materias

1 Fundamentos del álgebra lineal: un repaso

Espacios vectoriales, independencia lineal, bases, dimensión. Aplicaciones lineales, núcleo e imagen, rango y nulidad, espacio dual. Ecuaciones lineales y matrices, operaciones de fila, eliminación gaussiana. Determinantes y su evaluación, regla de Cramer.

2 Estructura de aplicaciones lineales

Autovalores de una aplicación lineal o matriz, autovectores. Aplicaciones cíclicas y matrices diagonalizables. Formas canónicas de una matriz. Polinomio característico de una matriz, el teorema de Cayley y Hamilton. Polinomio mínimo de una aplicación lineal. Forma normal de Jordan de una matriz.

3 Ortogonalidad y teoría espectral

Productos escalares reales y complejos, bases ortonormales, el algoritmo de Gram y Schmidt. Aplicaciones y matrices ortogonales y unitarias. Matrices simétricas y hermíticas. Aplicaciones y matrices positivas, descomposición polar. El teorema espectral.

4 Formas bilineales

Formas bilineales simétricas, congruencia de matrices, rango y signatura. Formas cuadráticas y sus signaturas. Formas bilineales alternantes, bases canónicas. Aplicaciones ortogonales y simplécticas, estructuras complejas.

5 Tópicos adicionales sobre matrices

Descomposición de matrices según sus valores singulares. Determinantes y pfaffianos de matrices antisimétricas. Fórmulas minimax para autovalores de matrices hermíticas. Matrices con entradas no negativas, el teorema de Frobenius y Perron.

Bibliografía

En la página del curso, se colocará semanalmente unos apuntes detallados. En la biblioteca se encuentra un gran surtido de libros sobre álgebra lineal; algunos son textos básicos con muchos ejercicios de rutina. A continuación se ofrece una lista de libros recomendables, que tratan la materia con mayor profundidad.

1 Sheldon Axler, *Linear Algebra Done Right*, 3a edición, Undergraduate Texts in Mathematics, Springer, Cham, Suiza, 2015.

2 Harry Dym, *Linear Algebra in Action*, AMS, Providence, RI, 2007.

3 Feliks R. Gantmacher, *The Theory of Matrices 1*, Chelsea, New York, NY, 1959.

4 Lidia I. Goloviná, *Algebra lineal y algunas de sus aplicaciones*, Mir, Moscú, 1974.

5 Paul R. Halmos, *Espacios vectoriales de dimensión finita*, Continental, México, 1971.

6 Roger A. Horn y Charles R. Johnson, *Matrix Analysis*, 2a edición, Cambridge University Press, Cambridge, 2013.

7 Yitzhak Katznelson y Yonatan R. Katznelson, *A (Terse) Introduction to Linear Algebra*, Student Mathematical Library **44**, AMS, Providence, RI, 2008.

8 Anatoly I. Maltsev, *Fundamentos de álgebra lineal*, Mir, Moscú, 1972.

9 Denis Serre, *Matrices: Theory and Applications*, Graduate Texts in Mathematics **216**, Springer, New York, 2002.

10 Helene Shapiro, *Linear Algebra and Matrices*, Pure and Applied Undergraduate Texts **24**, AMS, Providence, RI, 2015.

11 Georgi E. Shilov, *Linear Algebra*, Dover Books, Mineola, NY, 1977.

12 Xingshi Zhan, *Matrix Theory*, Graduate Studies in Mathematics **147**, AMS, Providence, RI, 2013.

Índice general

1 Fundamentos del álgebra lineal

The beginner [. . .] should not be discouraged if, on first reading of section 0, he finds that he does not have the prerequisites for reading the prerequisites.

— Paul Halmos[1]

Antes de abordar el estudio de aplicaciones lineales en general, conviene hacer un breve repaso de los conceptos fundamentales de los espacios vectoriales y las matrices, ya vistos en el curso anterior. El objeto de este resumen es fijar los conceptos y las notaciones que serán usados más adelante. Por lo tanto, se dejan las proposiciones sin demostración en este capítulo inicial.

1.1. Espacios vectoriales

En el álgebra lineal se emplean *escalares*, *vectores* y *matrices*. Los escalares forman un **cuerpo**,[2] es decir, un conjunto dotado con operaciones conmutativos de suma y producto, en donde cada elemento a tiene un negativo $-a$; cada elemento no cero a tiene un inverso multiplicativo $a^{-1} = 1/a$; y la ley distributiva $a(b+c) = ab + ac$ se cumple. Cada cuerpo contiene al menos dos elementos distintos: el cero aditivo 0, y la unidad multiplicativa 1.

Tres cuerpos ya son bien conocidos: $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$ denotarán los números *racionales*, *reales* y *complejos*, respectivamente. En lo sucesivo, $\mathbb{Z}$ denotará los números *enteros* y $\mathbb{N} = \{0, 1, 2, 3, \dots\}$ denotará los números enteros no negativos, también llamados números *naturales*.[3] Fíjese que $\mathbb{N}$ y $\mathbb{Z}$ no son cuerpos.

Hay cuerpos con un número finito de elementos, entre ellos $\mathbb{F}_p = \mathbb{Z}/p\mathbb{Z}$, los residuos bajo división por un entero primo p. Obsérvese que $\mathbb{F}_2 = \{0, 1\}$ tiene la mínima cantidad posible de elementos.

Notación. La letra $\mathbb{F}$ denotará un cuerpo *cualquiera*. Sus elementos se llamarán **escalares**.

Definición 1.1. Un **espacio vectorial** sobre un cuerpo $\mathbb{F}$ (a veces: un *espacio $\mathbb{F}$-vectorial*) es un conjunto V, cuyos elementos se llaman **vectores**, dotado con dos operaciones: suma de vectores y multiplicación escalar. La suma es asociativa y conmutativa, el cero para la suma se escribe $\mathbf{0} \in V$ y la multiplicación escalar cumple las relaciones:

$$1\boldsymbol{x} = \boldsymbol{x}; \quad a(\boldsymbol{x}+\boldsymbol{y}) = a\boldsymbol{x} + a\boldsymbol{y}; \quad (a+b)\boldsymbol{x} = a\boldsymbol{x} + b\boldsymbol{x}; \quad \text{para} \quad a, b \in \mathbb{F}, \ \boldsymbol{x}, \boldsymbol{y} \in V.$$

[1]Tomado del prólogo de su libro *Measure Theory* (Van Nostrand, Princeton, 1950).

[2]El nombre viene del alemán *Körper*, un término introducido por Richard Dedekind en 1871; se llama *corps* en francés, *cuerpo* en español, *corp* en rumano, etc., pero en inglés se llama *field*. En español, no debe usarse la traducción secundaria "campo", reservada para campos vectoriales, campos magnéticos, etc.

[3]Conviene incluir 0 como número natural, aunque este costumbre no es universal. Los autores franceses lo siguen, empleando la notación $\mathbb{N}^* = \{1, 2, 3, \dots\}$. En cambio, los autores alemanes a veces escriben $\mathbb{N} = \{1, 2, 3, \dots\}$ y $\mathbb{N}_0 = \{0, 1, 2, 3, \dots\}$, sin previo aviso. *Caveat lector.*

La totalidad de las “n-tuplas” $\boldsymbol{x} = (x_1, \dots, x_n)$, con cada $x_k \in \mathbb{F}$, es un espacio vectorial sobre $\mathbb{F}$ denotado por $\underline{\mathbb{F}^n}$. Las sumas y los múltiplos de n-tuplas se definen “entrada por entrada”. $\lozenge$

Ejemplo 1.2. La totalidad de *polinomios* con coeficientes en $\mathbb{F}$ se denota por $\underline{\mathbb{F}[t]}$, en donde la letra t es un *indeterminado*. Sus elementos son los $p(t) = a_0 + a_1 t + a_2 t^2 + \cdots + a_n t^n$ con cada $a_k \in \mathbb{F}$, $a_n \neq 0$. El número natural n es el *grado* del polinomio $p(t)$. Con la suma de polinomios y su multiplicación por escalares, $\mathbb{F}[t]$ es un espacio vectorial sobre $\mathbb{F}$. $\lozenge$

Definición 1.3. Sea V un espacio vectorial sobre $\mathbb{F}$. Un **subespacio** de V es una parte $W \subseteq V$ tal que W sea también un espacio $\mathbb{F}$-vectorial bajo las mismas operaciones de suma y multiplicación escalar. Dicho de otro modo, W es un subespacio de V si $W \subseteq V$ y si para $\boldsymbol{x}, \boldsymbol{y} \in W$, $c \in \mathbb{F}$, valen $\boldsymbol{x} + \boldsymbol{y} \in W$, $c\boldsymbol{x} \in W$.

La notación $\underline{W \leqslant V}$ significa que W es un subespacio de V. $\lozenge$

Ejemplo 1.4. Si $V = \mathbb{F}[t]$, y si $n \in \mathbb{N}$, sea $\mathbb{F}_n[t]$ la colección de polinomios de grado *no mayor que n*. Los polinomios constantes a_0, de grado 0, pertenecen a $\mathbb{F}_n[t]$. Es evidente que $\mathbb{F}_n[t]$ es subespacio de $\mathbb{F}[t]$ y que $\mathbb{F}_n[t] \leqslant \mathbb{F}_{n+1}[t]$ para cada $n \in \mathbb{N}$. $\lozenge$

Definición 1.5. Sea V un espacio vectorial sobre $\mathbb{F}$. Un vector en V de la forma

$$\boxed{\boldsymbol{x} = a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_m\boldsymbol{x}_m,} \tag{1.1}$$

donde $a_k \in \mathbb{F}$, $\boldsymbol{x}_k \in V$ para $k = 1, 2, \dots, m$, se llama una **combinación lineal** de los vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m$, con **coeficientes** $a_1, \dots, a_m$.

Se dice que la colección de vectores $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es **linealmente dependiente** si hay un juego de coeficientes $a_1, \dots, a_m$, *no todos cero*, tal que

$$a_1\boldsymbol{x}_1 + a_2\boldsymbol{x}_2 + \cdots + a_m\boldsymbol{x}_m = \boldsymbol{0}. \tag{1.2}$$

En cambio, si el único juego de coeficientes que hace cumplir (1.2) es $a_1 = \cdots = a_m = 0$, se dice que los vectores $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ son **linealmente independientes**.

Un conjunto $X \subset V$, posiblemente infinito, se dice linealmente independiente si cada parte finita de X es linealmente independiente; es decir, X es linealmente independiente si la ecuación (1.2) admite solamente la solución trivial $a_1 = \cdots = a_m = 0$ toda vez que $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in X$. $\lozenge$

Definición 1.6. El **subespacio generado** por una colección de vectores $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ en V es el menor subespacio $W \leqslant V$ que incluye esta colección. Es evidente que W es la totalidad de las combinaciones lineales posibles de la forma (1.1). Se usará la notación

$$\underline{\operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_m \rangle} := \{\, a_1\boldsymbol{x}_1 + \cdots + a_m\boldsymbol{x}_m : a_1, \dots, a_m \in \mathbb{F} \,\}$$

para denotar este subespacio.

Si X es una parte de V, el subespacio $\underline{\operatorname{lin}\langle X\rangle}$ generado por X es el menor subespacio de V que incluye X; esto es la totalidad de las combinaciones lineales de vectores en X. ◊

Definición 1.7. Si V es un espacio vectorial sobre $\mathbb{F}$, una **base** de V es una parte $\mathcal{B} \subset V$ tal que: (a) $\mathcal{B}$ es *linealmente independiente*; (b) $\mathcal{B}$ *genera* V, es decir, $\operatorname{lin}\langle\mathcal{B}\rangle = V$.

$\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V si y solo si cada vector $\boldsymbol{x} \in V$ puede expresarse como una combinación lineal (1.1) *de manera única*. Para $\boldsymbol{x} \in V$, los coeficientes $c_1, \dots, c_n \in \mathbb{F}$ tales que $\boldsymbol{x} = c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n$ forman un vector $\boldsymbol{c} = (c_1, \dots, c_n) \in \mathbb{F}^n$. Este vector

$$\underline{[\boldsymbol{x}]^{\mathcal{B}}} := \boldsymbol{c} \in \mathbb{F}^n \tag{1.3}$$

es el **representante** de $\boldsymbol{x} \in V$ en el espacio $\mathbb{F}^n$ *con respecto a la base* $\mathcal{B}$.

La *base estándar* de $\mathbb{F}^n$ es $\underline{\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}}$, donde

$$\boldsymbol{e}_1 := (1, 0, \dots, 0), \quad \boldsymbol{e}_2 := (0, 1, \dots, 0), \quad \dots, \quad \boldsymbol{e}_n := (0, 0, \dots, 1). \tag{1.4}$$

Nótese que $[\boldsymbol{c}]^{\mathcal{E}} = \boldsymbol{c}$, para todo $\boldsymbol{c} \in \mathbb{F}^n$. ◊

El origen $\{\mathbf{0}\}$ es un subespacio de cualquier espacio vectorial. Obsérvese que el *conjunto* $\{\mathbf{0}\}$ es linealmente *dependiente*, porque $1\,\mathbf{0} = \mathbf{0}$ es una solución no trivial de (1.2). Por otro lado, el conjunto vacío $\emptyset$ es linealmente independiente; y el menor subespacio que lo incluye es $\{\mathbf{0}\}$. Por lo tanto $\emptyset$ es una base de $\{\mathbf{0}\}$, con cero elementos.

Definición 1.8. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base de un espacio vectorial V y sean $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in V$. Si $m > n$, no es difícil verificar que $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ son linealmente dependientes. En consecuencia, si $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ es otra base de V, entonces $m \leqslant n$ y vice versa, así que $m = n$. Si V posee una base finita, el número n de sus elementos es la **dimensión** de V; en símbolos, $\underline{n = \dim V}$. (Se dice que V es *finitodimensional* en este caso.) En particular, vale $\dim \mathbb{F}^n = n$. ◊

El conjunto de *monomios* $\{1, t, t^2, \dots, t^n, \dots\}$ es una base del espacio vectorial $\mathbb{F}[t]$ de todos los polinomios sobre $\mathbb{F}$. De este modo se ve que $\mathbb{F}[t]$ es *infinitodimensional*.

► Para construir una base para un espacio vectorial dado, es útil saber que siempre se puede *completar una base parcial*, es decir, es posible prolongar una base de un subespacio en una base del espacio de marras, en vista del siguiente escolio.[4]

Escolio 1.9. *Sea V un espacio vectorial sobre $\mathbb{F}$ con $\dim V = n$, y sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\} \subset V$ un conjunto linealmente independiente de vectores, con $m < n$. Siempre es posible hallar otros vectores $\boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n \in V$ tales que $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m, \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n\}$ sea una base de V.* ⊟

[4] Un **escolio** es una proposición, usualmente no muy difícil, que el *escolar* debe demostrar por cuenta propia.

Definición 1.10. Sean V y W espacios vectoriales sobre el mismo cuerpo $\mathbb{F}$. Su **suma directa** $V \oplus W$ es el producto cartesiano $V \times W$, dotado de las siguientes operaciones de suma y multiplicación escalar:

$$\begin{aligned} (\boldsymbol{x}, \boldsymbol{y}) + (\boldsymbol{x}', \boldsymbol{y}') &:= (\boldsymbol{x} + \boldsymbol{x}', \boldsymbol{y} + \boldsymbol{y}'), &&\text{para} \quad \boldsymbol{x}, \boldsymbol{x}' \in V, \quad \boldsymbol{y}, \boldsymbol{y}' \in W, \\ c(\boldsymbol{x}, \boldsymbol{y}) &:= (c\boldsymbol{x}, c\boldsymbol{y}), &&\text{para} \quad \boldsymbol{x} \in V, \quad \boldsymbol{y} \in W, \quad c \in \mathbb{F}. \end{aligned}$$

Si V y W son finitodimensionales, entonces $\dim(V \oplus W) = \dim V + \dim W$. ◊

Definición 1.11. Sea V un espacio vectorial sobre $\mathbb{F}$ y sean U, W dos subespacios de V. Su **suma** es el subespacio

$$U + W := \{\, \boldsymbol{x} + \boldsymbol{y} : \boldsymbol{x} \in U,\ \boldsymbol{y} \in W \,\} \leqslant V.$$

En general, $\dim(U + W) \leqslant \dim U + \dim W$, con igualdad si y sólo si $U \cap W = \{\boldsymbol{0}\}$. En el caso de que $U \cap W = \{\boldsymbol{0}\}$, se identifica esta suma $U + W$ con la suma directa $U \oplus W$, pues tienen la misma dimensión.

El subespacio W es un **suplemento** de U en V si $U \cap W = \{\boldsymbol{0}\}$ y $U \oplus W = V$. Cada subespacio de V posee varios suplementos: si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es una base de U que se prolonga en una base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, entonces $\operatorname{lin}\langle \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n \rangle$ es un suplemento de U en V. ◊

Definición 1.12. Sea V un espacio vectorial sobre $\mathbb{F}$ y sea W un subespacio de V. El **espacio vectorial cociente** $\underline{V/W}$ es el conjunto de los *traslados* $\underline{\boldsymbol{x} + W} := \{\, \boldsymbol{x} + \boldsymbol{w} : \boldsymbol{w} \in W \,\}$ del subespacio W, dotado de las siguientes operaciones de suma y multiplicación escalar:

$$(\boldsymbol{x} + W) + (\boldsymbol{y} + W) := (\boldsymbol{x} + \boldsymbol{y}) + W, \qquad c(\boldsymbol{x} + W) := (c\boldsymbol{x}) + W,$$

para $\boldsymbol{x}, \boldsymbol{y} \in V$, $c \in \mathbb{F}$. El cero de V/W es el propio W, considerado como traslado trivial del subespacio W de V. Si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V tal que $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ sea una base de W, entonces $\{\boldsymbol{x}_{k+1} + W, \dots, \boldsymbol{x}_n + W\}$ es una base de V/W. Por ende, $\dim(V/W) = \dim V - \dim W$. ◊

1.2. Aplicaciones lineales

Definición 1.13. Sean V y W dos espacios vectoriales sobre un mismo cuerpo $\mathbb{F}$. Se dice que una función $T\colon V \to W$ es una **aplicación lineal** (o bien, *aplicación* $\mathbb{F}$*-lineal*) si T cumple

$$\boxed{\begin{aligned} T(\boldsymbol{x} + \boldsymbol{y}) &= T(\boldsymbol{x}) + T(\boldsymbol{y}), \\ T(c\boldsymbol{x}) &= cT(\boldsymbol{x}), \end{aligned}} \qquad \begin{aligned} &\text{para} \quad \boldsymbol{x}, \boldsymbol{y} \in V, \\ &\text{para} \quad \boldsymbol{x} \in V,\ c \in \mathbb{F}. \end{aligned}$$

Se denota la totalidad de aplicaciones lineales $T\colon V \to W$ por $\underline{\mathcal{L}(V, W)}$.

Este $\mathcal{L}(V, W)$ es un espacio vectorial sobre $\mathbb{F}$, bajo las operaciones:

$$\underline{T+S} : \boldsymbol{x} \mapsto T(\boldsymbol{x}) + S(\boldsymbol{x}),$$
$$\underline{c\,T} : \boldsymbol{x} \mapsto c\,T(\boldsymbol{x}).$$

Si Z es otro espacio vectorial sobre $\mathbb{F}$, y si $T \in \mathcal{L}(V, W)$, $S \in \mathcal{L}(W, Z)$, su *composición*[5] es la aplicación lineal $ST \in \mathcal{L}(V, Z)$ dado por $\underline{ST} : \boldsymbol{x} \mapsto S(T(\boldsymbol{x}))$. ◊

Una aplicación lineal *está determinada por sus valores sobre los elementos de una base* de su dominio.

Escolio 1.14. *Sean V y W dos espacios $\mathbb{F}$-vectoriales y sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base de V. Si $\boldsymbol{y}_1, \dots, \boldsymbol{y}_n$ son n vectores cualesquiera en W (no necesariamente distintos ni independientes), hay una única aplicación lineal $T \in \mathcal{L}(V, W)$ tal que $T(\boldsymbol{x}_k) = \boldsymbol{y}_k$ para $k = 1, 2, \dots, n$.* ⊟

Definición 1.15. Sea V un espacio vectorial sobre $\mathbb{F}$. Una **forma lineal** sobre V es una aplicación lineal $f : V \to \mathbb{F}$. El **espacio dual** de V es el espacio vectorial $\underline{V^*}$ de todas las formas lineales, esto es, $V^* := \underline{\mathcal{L}(V, \mathbb{F})}$. Si $\dim V$ es finito, entonces $\dim V^* = \dim V$. ◊

Definición 1.16. Si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V, la correspondiente **base dual** $\underline{\{f_1, \dots, f_n\}}$ de V^* se define por

$$f_k(c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_n\boldsymbol{x}_n) := c_k, \tag{1.5}$$

para $k = 1, 2, \dots, n$. Estas formas lineales f_k cumplen $f_k(\boldsymbol{x}_k) = 1$ y $f_k(\boldsymbol{x}_r) = 0$ si $k \neq r$. ◊

Definición 1.17 (Corchete de Iverson)**.** Vale la pena introducir usar el siguiente convenio de notación.[6] Si $R(x)$ es una relación lógica cualquiera que involucra un parámetro x, la notación $[\![R(x)]\!]$ denota la siguiente función booleana:

$$[\![R(x)]\!] := \begin{cases} 1, & \text{si } R(x) \text{ es CIERTO;} \\ 0, & \text{si } R(x) \text{ es FALSO.} \end{cases}$$

La conocida *delta de Kronecker* δ_{ij}, que es la función de dos índices i, j, que vale 1 cuando $i = j$ y vale 0 cuando $i \neq j$, resulta ser

$$\delta_{ij} = [\![i = j]\!].$$

De igual modo, la *función indicatriz* de un conjunto A es $\underline{\mathbf{1}_A(x)} := [\![x \in A]\!]$, y la *función de signo* sobre $\mathbb{R}$, que vale 1, 0 ó -1 cuando t es respectivamente positivo, cero o negativo, se escribe como $\underline{\operatorname{signo}(t)} := [\![t > 0]\!] - [\![t < 0]\!]$.

Con esta notación, la base dual de V^* queda determinada por $f_k(\boldsymbol{x}_r) := [\![k = r]\!]$. ◊

[5]La composición de las funciones T y S se denota también por $S \circ T$. Sin embargo, es usual abreviarlo a ST cuando se trata de *aplicaciones lineales*.

[6]Esta notación fue introducido en 1962 por Kenneth E. Iverson, en su libro *A Programming Language*. Para una evaluación de los usos y las ventajas de esta notación, véase: Donald E. Knuth, "Two notes on notation", *American Mathematical Monthly* **99** (1992), 403–422.

Cada vector $\boldsymbol{x} \in V$ da lugar a una forma lineal sobre V^*, a saber, la **evaluación** $f \mapsto f(\boldsymbol{x})$. No se distinguirá entre el vector $\boldsymbol{x} \in V$ y este miembro de $\underline{V^{**}} = (V^*)^*$; de esta manera, V se identifica con un subespacio del **espacio bidual** V^{**}; como $\dim V^{**} = \dim V^* = \dim V$, este subespacio es todo V^{**} cuando V es finitodimensional. Así, el espacio dual de V^* coincide con el espacio original V. Además, la base dual a $\{f_1, \dots, f_n\}$ es la base original $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V. (Esta reciprocidad entre V y V^* justifica el empleo de la palabra "dual".)

Definición 1.18. Sean V, W dos espacios vectoriales sobre $\mathbb{F}$. La **aplicación transpuesta** de $S \in \mathcal{L}(V, W)$ es la aplicación lineal $\underline{S^{\mathrm{t}}} \in \mathcal{L}(W^*, V^*)$ dada por

$$\boxed{S^{\mathrm{t}}(g) := g \circ S} \quad \text{para todo} \quad g \in W^*. \tag{1.6}$$

En otras palabras, si $g \in W^*$, $\boldsymbol{x} \in V$, entonces $S^{\mathrm{t}}(g) : \boldsymbol{x} \mapsto g(S(\boldsymbol{x}))$.

Si $S \in \mathcal{L}(V, W)$ y $R \in \mathcal{L}(W, Z)$, resulta que $\underline{(RS)^{\mathrm{t}} = S^{\mathrm{t}} R^{\mathrm{t}}}$, porque si $h \in Z^*$, entonces

$$S^{\mathrm{t}} R^{\mathrm{t}}(h) = S^{\mathrm{t}}(R^{\mathrm{t}}(h)) = S^{\mathrm{t}}(h \circ R) = (h \circ R) \circ S = h \circ (RS) = (RS)^{\mathrm{t}}(h). \qquad \lozenge$$

Definición 1.19. Sean V, W dos espacios vectoriales sobre $\mathbb{F}$ y sea $T \in \mathcal{L}(V, W)$. El **núcleo** de T es el subespacio $\underline{\ker T}$ de V dado por

$$\boxed{\ker T := \{\, \boldsymbol{x} \in V : T(\boldsymbol{x}) = \boldsymbol{0} \,\}} \leqslant V.$$

La **imagen** de T es el subespacio $\underline{T(V)}$ de W:

$$\boxed{T(V) := \{\, T(\boldsymbol{x}) : \boldsymbol{x} \in V \,\}} \leqslant W.$$

La **nulidad** de T es $\underline{n(T)} := \dim(\ker T)$. El **rango** de T es $\underline{r(T)} := \dim(T(V))$. Obsérvese que $n(T) \leqslant \dim V$ y $r(T) \leqslant \dim W$. ◊

Escolio 1.20. *Sea $T \in \mathcal{L}(V, W)$; entonces:*

(a) *T es inyectivo si y solo si $\ker T = \{\boldsymbol{0}\}$, si y solo si $n(T) = 0$;*

(b) *T es sobreyectivo si y solo si $T(V) = W$, si y solo si $r(T) = \dim W$.* ⊟

Definición 1.21. Sea V un espacio vectorial sobre $\mathbb{F}$ con $\dim V = n$. Considérese dos subespacios $M \leqslant V$ y $N \leqslant V^*$. El **anulador de** M es el subespacio $\underline{M^\perp} \leqslant V^*$ dado por

$$\boxed{M^\perp := \{\, f \in V^* : f(\boldsymbol{x}) = 0 \text{ para todo } \boldsymbol{x} \in M \,\}.}$$

El **anulado** de N es el subespacio $\underline{{}^\perp N} \leqslant V$ dado por

$$\boxed{{}^\perp N := \{\, \boldsymbol{x} \in V : f(\boldsymbol{x}) = 0 \text{ para todo } f \in N \,\}.}$$

Resulta que $\dim(M^\perp) = n - \dim M$; y $\dim({}^\perp N) = n - \dim N$. ◊

Escolio 1.22. *Sean V, W dos espacios vectoriales sobre $\mathbb{F}$ y sea $T \in \mathcal{L}(V,W)$. Entonces*

$$T(V)^\perp = \ker T^{\mathrm{t}} \quad \textit{y} \quad (\ker T)^\perp = T^{\mathrm{t}}(W^*).$$

Recíprocamente, $^\perp T^{\mathrm{t}}(W^*) = \ker T$; y $^\perp(\ker T^{\mathrm{t}}) = T(V)$.

Por lo tanto $n(T^{\mathrm{t}}) = \dim W - r(T)$ y además $r(T^{\mathrm{t}}) = \dim V - n(T)$.

Teorema 1.23 (Rango y nulidad)**.** *Para cualquier $T \in \mathcal{L}(V,W)$, valen:*

(a) $\boxed{r(T) = r(T^{\mathrm{t}}),}$

(b) $\boxed{r(T) + n(T) = \dim V.}$ ⊟

Demostración. Como $r(T^{\mathrm{t}}) = \dim V - n(T)$, las relaciones (a) y (b) son equivalentes. Basta demostrar la parte (b).

Si $n(T) = k$, sea $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k\}$ una base de $\ker T$. Por el Escolio 1.9, hay vectores $\{\boldsymbol{y}_1,\dots,\boldsymbol{y}_r\}$, donde $r = \dim V - k$, tales que $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k,\boldsymbol{y}_1,\dots,\boldsymbol{y}_r\}$ es una base de V.

Afirmación: $\{T\boldsymbol{y}_1,\dots,T\boldsymbol{y}_r\}$ es una base de $T(V)$.

Si $\boldsymbol{z} \in V$, entonces $\boldsymbol{z} = a_1\boldsymbol{x}_1 + \cdots + a_k\boldsymbol{x}_k + b_1\boldsymbol{y}_1 + \cdots + b_r\boldsymbol{y}_r$ y por lo tanto $T\boldsymbol{z} = T(b_1\boldsymbol{y}_1 + \cdots + b_r\boldsymbol{y}_r) = b_1\,T\boldsymbol{y}_1 + \cdots + b_r\,T\boldsymbol{y}_r$ porque $T(a_1\boldsymbol{x}_1 + \cdots + a_k\boldsymbol{x}_k) = \boldsymbol{0}$. Esto dice que $\{T\boldsymbol{y}_1,\dots,T\boldsymbol{y}_r\}$ genera $T(V)$.

Por otro lado, si $c_1\,T\boldsymbol{y}_1 + \cdots + c_r\,T\boldsymbol{y}_r = \boldsymbol{0}$, entonces $T(c_1\,\boldsymbol{y}_1 + \cdots + c_r\,\boldsymbol{y}_r) = \boldsymbol{0}$ así que $c_1\,\boldsymbol{y}_1 + \cdots + c_r\,\boldsymbol{y}_r \in \ker T$. Entonces hay escalares $d_1,\dots,d_k \in \mathbb{F}$ tales que

$$c_1\,\boldsymbol{y}_1 + \cdots + c_r\,\boldsymbol{y}_r = d_1\boldsymbol{x}_1 + \cdots + d_k\boldsymbol{x}_k.$$

Por ser $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k,\boldsymbol{y}_1,\dots,\boldsymbol{y}_r\}$ linealmente independiente en V, esto solo es posible si cada $c_i = 0$ (y cada $d_j = 0$). Se ha mostrado que $\{T\boldsymbol{y}_1,\dots,T\boldsymbol{y}_r\}$ es linealmente independiente.

Se concluye que $r(T) = \dim T(V) = r := \dim V - k$, así que $k + r = \dim V$, lo cual demuestra la relación (b). □

1.3. Matrices

Definición 1.24. Una **matriz** $m \times n$ con entradas en un cuerpo $\mathbb{F}$ es un arreglo rectangular de elementos de $\mathbb{F}$, con m filas o renglones y n columnas. Para abreviar, se escribe $\underline{A = [a_{ij}]}$, donde se sobreentiende que $i = 1,2,\dots,m$ y $j = 1,2,\dots,n$.

La totalidad de tales matrices $m \times n$ es un espacio $\mathbb{F}$-vectorial de dimensión mn. En el caso $m = n$, se habla de *matrices cuadradas*. $\underline{M_n(\mathbb{F})}$ denota el espacio vectorial de matrices cuadradas $n \times n$ con entradas en $\mathbb{F}$. (Otra notación usada es $\mathbb{F}^{n\times n} = M_n(\mathbb{F})$. Así, el espacio vectorial de matrices rectangulares $m \times n$ puede denotarse por $\underline{\mathbb{F}^{m\times n}}$.) ◊

Las **columnas** de una matriz $A \in \mathbb{F}^{m\times n}$ son *vectores en* $\mathbb{F}^m$. La columna número j es

$$\underline{\boldsymbol{a}_j} = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix} \in \mathbb{F}^m \quad \text{para} \quad j = 1, 2, \dots, n.$$

De este modo, $A = [\boldsymbol{a}_1\ \boldsymbol{a}_2\ \cdots\ \boldsymbol{a}_n]$ es una lista ordenada de **vectores de columna**.

Las **filas** de A serán *vectores en el espacio dual* $(\mathbb{F}^n)^* \simeq \mathbb{F}^n$. La fila número i es[7]

$$\underline{\boldsymbol{a}^i} = [a_{i1}\ a_{i2}\ \cdots\ a_{in}] \in (\mathbb{F}^n)^* \quad \text{para} \quad i = 1, 2, \dots, m.$$

La fila $\boldsymbol{a}^i$ corresponde a la siguiente forma lineal en $(\mathbb{F}^n)^*$:

$$\boldsymbol{x} \mapsto \boldsymbol{a}^i \cdot \boldsymbol{x} = a_{i1}x_1 + a_{i2}x_2 + \cdots + a_{in}x_n,$$

En general, la notación $\underline{\boldsymbol{x} \cdot \boldsymbol{y}}$ denotará el *producto punto* de dos vectores en $\mathbb{F}^n$, esto es, $\boldsymbol{x} \cdot \boldsymbol{y} := x_1y_1 + x_2y_2 + \cdots + x_ny_n$.

Definición 1.25. La **transpuesta** A^{t} de una matriz $A \in \mathbb{F}^{m\times n}$ es una matriz $n \times m$, cuya entrada (i, j) es la entrada (j, i) de A; en símbolos, $\underline{A^{\mathrm{t}}} := [a_{ji}]$. Una matriz cuadrada $A \in M_n(\mathbb{F})$ se llama **simétrica** si $\underline{A^{\mathrm{t}} = A}$.

Así también, la matriz A es una columna ordenada de *vectores de fila*: $A = [\boldsymbol{a}^1\ \boldsymbol{a}^2\ \cdots\ \boldsymbol{a}^m]^{\mathrm{t}}$.

Una matriz cuadrada $A \in M_n(\mathbb{F})$ es **triangular inferior** si $a_{ij} = 0$ para $i < j$; **triangular superior** si $a_{ij} = 0$ para $i > j$; y **diagonal** si $a_{ij} = 0$ para $i \neq j$. ◊

Definición 1.26. El **producto de matrices** de dos matrices $A \in \mathbb{F}^{m\times n}$ y $B \in \mathbb{F}^{n\times r}$ se define como la matriz $C = \underline{AB} \in \mathbb{F}^{m\times r}$ cuya entrada (i, j) es el producto punto de la fila i de A con la columna j de B; es decir,

$$\boxed{c_{ij} := \boldsymbol{a}^i \cdot \boldsymbol{b}_j = \textstyle\sum_{k=1}^n a_{ik}b_{kj}.} \tag{1.7}$$

Para que esta definición sea factible, el número n de columnas de A debe ser igual al número de filas de B; entonces $\boldsymbol{a}^i \in (\mathbb{F}^n)^*$, $\boldsymbol{b}_j \in \mathbb{F}^n$ y el producto punto tiene sentido.

El producto matricial obedece las leyes algebraicas:

$$A(BC) = (AB)C; \quad (A + D)B = AB + DB; \quad A(B + E) = AB + AE. \tag{1.8}$$

y además $(AB)^{\mathrm{t}} = B^{\mathrm{t}}A^{\mathrm{t}}$. Sin embargo, este producto no es conmutativo, porque $AB \neq BA$ en general.

[7]Para distinguir entre vectores de fila y vectores de columna, conviene usar exponentes o “superíndices” para etiquetar aquellos.

Si $A, B \in M_n(\mathbb{F})$, su producto AB también pertenece a $M_n(\mathbb{F})$. Este producto es asociativo y distributivo sobre la suma de matrices, en vista de (1.8), Entonces $M_n(\mathbb{F})$ es a la vez un *espacio vectorial* sobre $\mathbb{F}$ y un *anillo* (no conmutativo): dícese que que $M_n(\mathbb{F})$ es un **álgebra** sobre $\mathbb{F}$ (o una $\mathbb{F}$-*álgebra*).[8] El álgebra $M_n(\mathbb{F})$ tiene una identidad multiplicativa, la **matriz identidad** $1_n \in M_n(\mathbb{F})$, que cumple $A1_n = 1_nA = A$ para todo $A \in M_n(\mathbb{F})$.

Concretamente, $\underline{1_n} := [\delta_{ij}]$, donde $\delta_{ij} = [\![i = j]\!]$ es la "delta de Kronecker". Esta es una matriz diagonal:[9]

$$1_n = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix}.$$

En un contexto en donde el tamaño n es fijo, se suele abreviar $\underline{1 := 1_n}$. ◊

Definición 1.27. Un elemento $A \in M_n(\mathbb{F})$ es una **matriz invertible** o *matriz no singular* si hay otra matriz $C \in M_n(\mathbb{F})$ tal que $AC = CA = 1_n$. Si A es invertible, su **matriz inversa** C es única y se denota $C =: \underline{A^{-1}}$. Fíjese que $(AB)^{-1} = B^{-1}A^{-1}$ cuando A, B son matrices invertibles.[10] ◊

Definición 1.28. Sea $T : V \to W$ una aplicación lineal entre dos espacios vectoriales finitodimensionales sobre $\mathbb{F}$. Dadas dos bases, $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ para V y $\mathcal{C} = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ para W, la **matriz de** T con respecto a estas bases es la matriz $A = [a_{ij}] \in \mathbb{F}^{m\times n}$ dada por[11]

$$\boxed{T(\boldsymbol{x}_j) =: \textstyle\sum_{i=1}^m a_{ij}\, \boldsymbol{y}_i.} \tag{1.9}$$

Para exhibir la dependencia de la matriz A tanto de T como de las bases $\mathcal{B}$ y $\mathcal{C}$, se escribe

$$A = [a_{ij}] =: \underline{[T]^{\mathcal{C}}_{\mathcal{B}}}.$$
◊

Si $\boldsymbol{x} = c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n \in V$ y si $T(\boldsymbol{x}) = b_1\boldsymbol{y}_1 + \cdots + b_m\boldsymbol{y}_m \in W$, entonces

$$\sum_{i=1}^m b_i\, \boldsymbol{y}_i = T(\boldsymbol{x}) = T\biggl(\sum_{j=1}^n c_j\, \boldsymbol{x}_j\biggr) = \sum_{j=1}^n c_j\, T(\boldsymbol{x}_j) = \sum_{j=1}^n \sum_{i=1}^m a_{ij} c_j\, \boldsymbol{y}_i,$$

[8] Una $\mathbb{F}$-álgebra es una estructura algebraica con tres operaciones compatibles: suma, producto y multiplicación escalar por elementos de $\mathbb{F}$. Los polinomios $\mathbb{F}[t]$ es otro ejemplo (conmutativo) de una $\mathbb{F}$-álgebra.

[9] En los libros de texto sobre álgebra lineal, es común denotar la matriz identidad por I_n o I en vez de 1_n o 1. Sin embargo, la experiencia aconseja denotar todos los elementos cero por el símbolo 0 y todos los elementos identidad por 1, hasta donde sea posible. Así, la matriz con entradas nulas se denota por 0 y no por O; el operador lineal identidad en $\mathcal{L}(V, V)$ se denota por 1_V y no por id_V; etcétera.

[10] Los autores argentinos escriben *inversible*, lo cual tiene cierta base etimológica. Sin embargo, la forma *invertible* es más común y es esa que se encuentra en el Diccionario de la Real Academia.

[11] Fíjese bien en la forma "enrevesada" de combinar los índices i, j al lado derecho de esta ecuación.

donde se ha usado la linealidad de la aplicación T. En vista de la independencia lineal de los vectores $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$, se concluye que

$$b_i = \sum_{j=1}^{n} a_{ij} c_j \quad \text{para cada } i, \quad \text{o bien:} \quad \underline{\boldsymbol{b} = A\boldsymbol{c}}.$$

En otras palabras, $[T(\boldsymbol{x})]^{\mathcal{C}} = A\,[\boldsymbol{x}]^{\mathcal{B}}$, de tal manera que

$$[T(\boldsymbol{x})]^{\mathcal{C}} = [T]^{\mathcal{C}}_{\mathcal{B}}\,[\boldsymbol{x}]^{\mathcal{B}}. \tag{1.10}$$

Fíjese que $T \mapsto A = [T]^{\mathcal{C}}_{\mathcal{B}}$ es un **isomorfismo lineal**, es decir, una aplicación lineal biyectiva, entre los espacios vectoriales $\mathcal{L}(V, W)$ y $\mathbb{F}^{m\times n}$; por ende

$$\dim \mathcal{L}(V, W) = \dim \mathbb{F}^{m\times n} = mn = (\dim W)(\dim V).$$

En efecto, el Escolio 1.14 afirma que $T \mapsto A$ es inyectiva, y la aplicación inversa es $A \mapsto T_A$, donde

$$\underline{T_A(\boldsymbol{x}) := A\boldsymbol{x}} \quad \text{para todo} \quad \boldsymbol{x} \in \mathbb{F}^n.$$

Es fácil verificar que las biyecciones $T \mapsto A$, $A \mapsto T_A$ son lineales.

También es posible comprobar que la correspondencia $T \leftrightarrow A = [T]^{\mathcal{C}}_{\mathcal{B}}$ preserva las otras operaciones algebraicas, como sigue.

(a) Si $\mathcal{D}$ es una base del espacio vectorial Z y si $B = [S]^{\mathcal{D}}_{\mathcal{C}}$ es la matriz de una aplicación lineal $S \in \mathcal{L}(W, Z)$, entonces la matriz de la composición ST es el producto matricial BA, es decir,

$$[ST]^{\mathcal{D}}_{\mathcal{B}} = BA = [S]^{\mathcal{D}}_{\mathcal{C}}[T]^{\mathcal{C}}_{\mathcal{B}}.$$

(b) Si $[T]^{\mathcal{C}}_{\mathcal{B}} = A$ y si T es biyectivo, entonces $[T^{-1}]^{\mathcal{B}}_{\mathcal{C}} = A^{-1}$.

(c) Si $\mathcal{B}^* = \{f_1, \dots, f_n\} \subset V^*$ y $\mathcal{C}^* = \{g_1, \dots, g_m\} \subset W^*$ son las bases duales de $\mathcal{B}$ y $\mathcal{C}$ respectivamente, la matriz correspondiente a la aplicación transpuesta $T^{\mathrm{t}} \in \mathcal{L}(W^*, V^*)$ es la matriz transpuesta A^{t}:

$$[T]^{\mathcal{C}}_{\mathcal{B}} = A \implies [T^{\mathrm{t}}]^{\mathcal{B}^*}_{\mathcal{C}^*} = A^{\mathrm{t}}.$$

Proposición 1.29. *Si $A \in \mathbb{F}^{m\times n}$, la imagen de la aplicación correspondiente $T_A \colon \mathbb{F}^n \to \mathbb{F}^m$ es el subespacio de $\mathbb{F}^m$ generado por las columnas de la matriz A.*

Demostración. Si $\boldsymbol{x} \in \mathbb{F}^n$, entonces $T_A(\boldsymbol{x}) = A\boldsymbol{x} \in \mathbb{F}^m$. En términos de la base estándar $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de $\mathbb{F}^n$, se ve que $\boldsymbol{x} = \sum_{j=1}^n x_j\,\boldsymbol{e}_j$. Es evidente que $A\boldsymbol{e}_j = \boldsymbol{a}_j$ para $j = 1, \dots, n$. Por la linealidad de T_A, se obtiene

$$\boxed{A\boldsymbol{x} = T_A(\boldsymbol{x}) = \sum_{j=1}^{n} x_j\,T_A(\boldsymbol{e}_j) = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n.} \tag{1.11}$$

Esto muestra que cada vector $A\boldsymbol{x}$ es una combinación lineal de las columnas $\boldsymbol{a}_1, \dots, \boldsymbol{a}_n$; e inversamente, cada combinación lineal de estas columnas es de la forma $A\boldsymbol{x}$ para algún $\boldsymbol{x} \in \mathbb{F}^n$. Se concluye que $T_A(\mathbb{F}^n) = \operatorname{lin}\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_n \rangle \leqslant \mathbb{F}^m$. $\square$

En general las n columnas de A no son linealmente independientes; de hecho, son independientes si y solo si $n = r(T_A)$. Esto motiva la siguiente definición.

Definición 1.30. Si $A \in \mathbb{F}^{m\times n}$, el **rango de la matriz** A se define como el rango de la aplicación lineal T_A, es decir, $\underline{r(A)} := r(T_A)$. Del mismo modo, la **nulidad** de A se define como la nulidad de T_A, esto es, $\underline{n(A)} := n(T_A)$.

Por lo tanto, $r(A)$ es *el máximo número de columnas linealmente independientes* de entre las columnas de A. Ahora el Teorema 1.23 implica que $\underline{r(A^{\mathsf{t}}) = r(A)}$. Por tanto, el rango de A es también el máximo número de *filas* linealmente independientes en A. (Fíjese que las filas de A son las columnas de A^{t}.) En consecuencia, vale $r(A) \leqslant \min\{m, n\}$. $\Diamond$

► Dadas dos espacios vectoriales finitodimensionales V y W, sean $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ y $\mathcal{B}' = \{\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_n\}$ dos bases de V y sean $\mathcal{C} = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ y $\mathcal{C}' = \{\boldsymbol{y}'_1, \dots, \boldsymbol{y}'_m\}$ dos bases de W. Una aplicación lineal $T \in \mathcal{L}(V, W)$ tiene dos matrices $A = [T]^{\mathcal{C}}_{\mathcal{B}}$ y $B = [T]^{\mathcal{C}'}_{\mathcal{B}'}$. La relación entre estas matrices es

$$[T]^{\mathcal{C}'}_{\mathcal{B}'} = [1_W]^{\mathcal{C}'}_{\mathcal{C}}\, [T]^{\mathcal{C}}_{\mathcal{B}}\, [1_V]^{\mathcal{B}}_{\mathcal{B}'}, \quad \text{o bien} \quad B = QAP,$$

donde $P = [1_V]^{\mathcal{B}}_{\mathcal{B}'}$ y $Q = [1_W]^{\mathcal{C}'}_{\mathcal{C}}$ son las **matrices de cambio de base** respectivas para V y W. Concretamente,

$$\boldsymbol{x}'_s =: \sum_{j=1}^{n} p_{js}\boldsymbol{x}_j, \qquad \boldsymbol{y}_i =: \sum_{r=1}^{m} q_{ri}\boldsymbol{y}'_r. \tag{1.12}$$

Obsérvese que P, por ser la matriz de una aplicación identidad 1_V respecto de ciertas bases, es una matriz cuadrada invertible. La matriz Q es invertible por la misma razón.

Definición 1.31. Dos **matrices** $A, B \in \mathbb{F}^{m\times n}$ se dicen **equivalentes** si hay un par de matrices *invertibles* $P \in M_n(\mathbb{F})$ y $Q \in M_m(\mathbb{F})$ tales que $\underline{B = QAP}$. $\Diamond$

Si A y B representan una aplicación lineal $T \in \mathcal{L}(V, W)$ respecto de dos pares de bases para V y W, entonces A y B son equivalentes. Inversamente, si A y B son equivalentes mediante la relación $B = QAP$, y si se cambia las bases estándares en $\mathbb{F}^n$ y $\mathbb{F}^m$ por (1.12), entonces B es la matriz de T_A con respecto a las nuevas bases. En particular, vale $\underline{r(A) = r(T) = r(B)}$.

Los cambios de base de mayor interés ocurren cuando $W = V$, $\mathcal{C} = \mathcal{B}$ y $\mathcal{C}' = \mathcal{B}'$, es decir, $\boldsymbol{y}_k = \boldsymbol{x}_k$ y $\boldsymbol{y}'_r = \boldsymbol{x}'_r$ para $k, r = 1, \dots, n$. En este caso, por inspección de (1.12), o bien por la reciprocidad entre $P = [1_V]^{\mathcal{B}}_{\mathcal{B}'}$ y $Q = [1_V]^{\mathcal{B}'}_{\mathcal{B}}$, se ve que $Q = P^{-1}$ en $M_n(\mathbb{F})$.

Definición 1.32. Dos **matrices cuadradas** $A, B \in M_n(\mathbb{F})$ se dicen **semejantes** si hay una matriz *invertible* $P \in M_n(\mathbb{F})$ tal que $\underline{B = P^{-1}AP}$. $\lozenge$

Si A y B representan una aplicación lineal $T \in \mathcal{L}(V, V)$ respecto de dos bases de V, entonces A y B son semejantes. Inversamente, si A y B son matrices semejantes mediante $B = P^{-1}AP$, y si se cambio la base estándar $\mathcal{E} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de $\mathbb{F}^n$ a la base $\mathcal{P} = \{\boldsymbol{p}_1, \dots, \boldsymbol{p}_n\}$ mediante (1.12), entonces $B = [T_A]^{\mathcal{P}}_{\mathcal{P}}$ es la matriz de T_A respecto de la nueva base $\mathcal{P}$.

1.4. Ecuaciones lineales y eliminación gaussiana

Un sistema de ecuaciones lineales tiene la siguiente forma:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\
\cdots\cdots\cdots\cdots &= \vdots \\
a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m
\end{aligned}
\tag{1.13a}
$$

donde los coeficientes a_{ij} y b_i pertenecen al cuerpo $\mathbb{F}$.

Para la resolución de este sistema, se procede a *eliminar* x_1 de la segunda ecuación y las siguientes, al restar de la ecuación número i un múltiplo apropiado de la primera ecuación: por un factor $-a_{i1}/a_{11}$, si $a_{11} \neq 0$. En seguida, se elimina x_2 de la tercera ecuación y las siguientes, al restar de ellas ciertos múltiplos de la segunda; y así sucesivamente. En el k-ésimo paso, es posible continuar si el coeficiente actual de x_k en la ecuación número k no es cero. En cambio, si este coeficiente fuera cero, habría que intercambiar esta ecuación con otra más abajo cuyo coeficiente de x_k no es nula, antes de seguir con el proceso de eliminación.[12]

Considérese el caso en donde $m = n$. Si el proceso de eliminación resulta exitoso, se obtiene un **sistema triangular** de ecuaciones:

$$
\begin{aligned}
a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\
a_{22}^{(2)}x_2 + \cdots + a_{2n}^{(2)}x_n &= b_2^{(2)} \\
\cdots\cdots\cdots\cdots &= \vdots \\
a_{nn}^{(n)}x_n &= b_n^{(n)}
\end{aligned}
$$

[12]Este algoritmo recibe el nombre de **eliminación gaussiana**, en parte porque *Carl Friedrich Gauß* lo usó para resolver un sistema 6×6 durante sus investigaciones sobre la órbita del asteroide Pallas. Por supuesto, el método en sí es mucho más antiguo. Aparece en el manuscrito chino *Jiuzhang Suanshu* (*Nueve capítulos sobre el arte de calcular*), de autoría desconocida en la época de la dinastía Han (~150 a.C.). Su octavo capítulo, *Fang cheng* (arreglo cuadrilongo), trata el método de eliminación.

con $a_{11} \neq 0, a_{22}^{(2)} \neq 0, \dots, a_{nn}^{(n)} \neq 0$. Entonces se puede despejar las variables $x_n, x_{n-1}, \dots, x_1$ por **sustitución regresiva**, para obtener una solución única del sistema. En cambio, si alguno de los **pivotes** $a_{kk}^{(k)}$ resulta ser 0, entrarían otras posibilidades, como la inexistencia de soluciones o la existencia de varias soluciones.

Al resumir la metodología de manipular sistemas de ecuaciones, se ve que los cálculos admisibles son combinaciones de las siguientes tres **operaciones elementales**:

(a) multiplicar una ecuación por una constante $c \neq 0$;

(b) sustraer de una ecuación un múltiplo de cualquier otra ecuación;

(c) intercambiar dos ecuaciones de la lista.

► El sistema de ecuaciones lineales (1.13) puede escribirse como una sola ecuación matricial:

$$A\boldsymbol{x} = \boldsymbol{b}, \quad \text{con} \quad A \in \mathbb{F}^{m\times n},\ \boldsymbol{b} \in \mathbb{F}^m. \tag{1.13b}$$

Las **incógnitas** $x_1, \dots, x_n$ forman un vector de columna $\boldsymbol{x} \in \mathbb{F}^n$. Alternativamente, el sistema puede expresarse como una sola ecuación vectorial en $\mathbb{F}^m$:

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}, \tag{1.13c}$$

en donde los vectores $\boldsymbol{a}_j \in \mathbb{F}^m$ son las *columnas* de la matriz A. Esta última forma muestra que un sistema de ecuaciones lineales expresa un vector dado $\boldsymbol{b}$ como una *combinación lineal de otros vectores dados* $\boldsymbol{a}_j$, y que los *coeficientes desconocidos* de esa combinación lineal corresponden a la solución del sistema de ecuaciones. De (1.13c) se ve que hay al menos una solución $\boldsymbol{x}$ al sistema si y solo si el vector $\boldsymbol{b}$ pertenece al subespacio $\operatorname{lin}\langle \boldsymbol{a}_1, \dots, \boldsymbol{a}_n \rangle = T_A(\mathbb{F}^n)$.

► La **matriz aumentada** de la ecuación $A\boldsymbol{x} = \boldsymbol{b}$ es la matriz $[A \mid \boldsymbol{b}] \in \mathbb{F}^{m\times(n+1)}$, donde se agrega $\boldsymbol{b}$ como columna suplementaria a la matriz A. Hay tres tipos de **operaciones de fila elementales** sobre matrices aumentadas que corresponden a las operaciones elementales sobre sistemas de ecuaciones:

(a) *multiplicar una fila por una constante* $c \neq 0$;

(b) *sustraer de una fila un múltiplo de cualquier otra fila*;

(c) *intercambiar dos filas de la matriz aumentada.*

Si $[A \mid \boldsymbol{b}]$ es equivalente a $[A' \mid \boldsymbol{b}']$ por operaciones de fila, entonces $A\boldsymbol{x} = \boldsymbol{b}$ si y solo si $A'\boldsymbol{x} = \boldsymbol{b}'$, es decir, los sistemas de ecuaciones asociados tienen las mismas soluciones $\boldsymbol{x}$.

Escolio 1.33. *Las operaciones de fila no cambian el rango de una matriz.* ⊟

Proposición 1.34. *Cualquier operación de fila elemental sobre una matriz $B \in \mathbb{F}^{m\times n}$ se efectúa por* premultiplicación[13] $B \mapsto AB$ *de esa matriz por una matriz cuadrada* $A \in M_m(\mathbb{F})$.

Demostración. Ad (a): Se multiplica la fila i de B por $c \neq 0$ con $B \mapsto M_i(c)B$, donde

$$M_i(c) = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & c & \dots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix}. \tag{1.14a}$$

Aquí $M_i(c)$ es la matriz diagonal con entradas diagonales $m_{ii} = c$ y $m_{jj} = 1$ para $j \neq i$.

Ad (b): Se sustrae de la fila $\boldsymbol{b}^k$ unas c veces la fila $\boldsymbol{b}^i$ con $B \mapsto R_{ik}(c)B$, donde

$$R_{ik}(c) = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & -c & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix} \quad \text{si} \quad k > i. \tag{1.14b}$$

Si $k \neq i$, $R_{ik}(c)$ tiene entradas $r_{ik} = -c$, $r_{jj} = 1$ para todo j; sus otras entradas son ceros.

Ad (c): Se intercambian las filas i y k de B con $B \mapsto P_{ik}B$, donde

$$P_{ik} = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix}. \tag{1.14c}$$

Las entradas de P_{ik} son: $p_{ik} = p_{ki} = 1$, $p_{jj} = 1$ si $j \neq i$, $j \neq k$; y otras entradas ceros. □

[13] Como el producto de matrices no es conmutativo, se debe distinguir entre los procesos de *pre*multiplicación $B \mapsto AB$ y *pos*multiplicación $C \mapsto CA$ por una matriz A.

Las matrices en las fórmulas (1.14) son invertibles: por cálculo directo, se ve que

$$M_i(c)^{-1} = M_i(1/c), \qquad R_{ik}(c)^{-1} = R_{ik}(-c), \qquad P_{ik}^{-1} = P_{ik}.$$

Por lo tanto, las operaciones de fila elementales pueden deshacerse por otras operaciones de fila elementales, ejecutadas mediante premultiplicación por otras matrices de los tipos (1.14).

Ejemplo 1.35. Considérese este sistema de 3 ecuaciones lineales, con 3 incógnitas:

$$\begin{aligned} x_1 + 3x_2 - x_3 &= 1, \\ 3x_1 + 4x_2 - 4x_3 &= 7, \\ 3x_1 + 6x_2 + 2x_3 &= -3. \end{aligned}$$

Entonces $A \in M_3(\mathbb{F})$, $[A \mid \boldsymbol{b}] \in \mathbb{F}^{3\times 4}$, y la siguiente eliminación gaussiana:

$$\left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 3 & 4 & -4 & 7 \\ 3 & 6 & 2 & -3 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 0 & -5 & -1 & 4 \\ 3 & 6 & 2 & -3 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 0 & -5 & -1 & 4 \\ 0 & -3 & 5 & -6 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} 1 & 3 & -1 & 1 \\ 0 & -5 & -1 & 4 \\ 0 & 0 & \frac{28}{5} & -\frac{42}{5} \end{array}\right]$$

es equivalente a la composición de tres premultiplicaciones,

$$[A \mid \boldsymbol{b}] \longmapsto R_{23}(\tfrac{3}{5})\, R_{13}(3)\, R_{12}(3)\, [A \mid \boldsymbol{b}] =: \underline{[V \mid \boldsymbol{c}]},$$

en donde V es una matriz triangular superior. Para revertir el proceso, fíjese que

$$\begin{aligned} [A \mid \boldsymbol{b}] &= \bigl(R_{23}(\tfrac{3}{5})\, R_{13}(3)\, R_{12}(3)\bigr)^{-1} [V \mid \boldsymbol{c}] \\ &= R_{12}(-3)\, R_{13}(-3)\, R_{23}(-\tfrac{3}{5}) [V \mid \boldsymbol{c}] =: L\, [V \mid \boldsymbol{c}]. \end{aligned}$$

Esta matriz L es una matriz triangular inferior:

$$L = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 3 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & \frac{3}{5} & 1 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & \frac{3}{5} & 1 \end{bmatrix}.$$

El resultado que este proceso de "pivoteo", en donde se ha empleado únicamente operaciones de fila del segundo tipo,[14] es una **factorización** de la matriz A como $\underline{A = LV}$, donde L es una matriz triangular inferior y V es una matriz triangular superior:

$$\begin{bmatrix} 1 & 3 & -1 \\ 3 & 4 & -4 \\ 3 & 6 & 2 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 3 & 1 & 0 \\ 3 & \frac{3}{5} & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 & -1 \\ 0 & -5 & -1 \\ 0 & 0 & \frac{28}{5} \end{bmatrix}. \qquad \Diamond$$

[14] El verbo *pivotear*, aun no reconocido por la Real Academia Española, viene del francés *pivoter*: girar en torno de un punto de apoyo (*pivot*). En la práctica de la programación lineal, un "pivote" es un elemento no cero de una matriz que sirve de marcador para luego convertir las demás entradas de su columna en ceros mediante operaciones de fila del segundo tipo.

Obsérvese que las entradas subdiagonales de L son los múltiplos de filas escogidos en el proceso de eliminación del sistema de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$. Al guardar cuenta de dichos múltiplos (efectuados en el orden apropiado), se escribe la matriz triangular inferior L sin necesidad de hacer más cálculos. El sistema $V\boldsymbol{x} = \boldsymbol{c}$ es el resultado de la fase de eliminación. En resumen, el proceso de eliminación gaussiana no solo resuelve el sistema de ecuaciones, sino que además proporciona la factorización $A = LV$.

La matriz triangular L es *unipotente*, es decir, sus entradas diagonales son iguales a 1. Además, las entradas diagonales $a_{kk}^{(k)}$ de V no son ceros. Por tanto, V puede factorizarse en un producto $\underline{V = DU}$, donde D es la matriz diagonal con $d_{kk} = a_{kk}^{(k)}$; y U es la matriz triangular superior unipotente obtenida al dividir cada fila de V por su entrada diagonal. De este modo se obtiene la factorización $\underline{A = LDU}$, en donde D es diagonal y L, U son triangulares unipotentes.

► Hay matrices que no admiten este tipo de factorización. La eliminación gaussiana de un sistema $A\boldsymbol{x} = \boldsymbol{b}$ es *simple* si se reduce a un sistema triangular $V\boldsymbol{x} = \boldsymbol{c}$ con operaciones de fila del segundo tipo solamente. En tal caso, se obtiene $A = LV$ donde L es el producto de matrices de tipo $R_{ik}(-c)$ con $i < k$, $c \in \mathbb{F}$ y las entradas diagonales de V no son ceros.

En el caso contrario, se obtiene $a_{kk}^{(k)} = 0$ para algún k y es necesario intercambiar algunas filas para continuar con la eliminación. Si (y solo si) la matriz original A es invertible, se llegará eventualmente al deseado sistema triangular $V\boldsymbol{x} = \boldsymbol{c}$. El algoritmo produce una factorización más general de tipo $\underline{A = PLV}$ para matrices invertibles, donde P es un producto de matrices de tipo P_{ik}.

Para obtener esta factorización, se subdivide el algoritmo en *pasos*: el paso número k se compone de operaciones de fila con el fin de reemplazar con ceros todas las entradas de la columna k debajo de la diagonal. Si después de $(k-1)$ pasos, la entrada (k, k) también es cero, se busca una fila j con $j > k$ cuya entrada (k, j) no sea cero;[15] luego, se intercambian las filas j y k y se guarda cuenta de la matriz P_{kj} que registra el cambio de filas. El nuevo elemento no cero en la posición (k, k) es el pivote $a_{kk}^{(k)}$, y se procede al paso número $(k+1)$. El factor P de A es el producto ordenado, de derecha a izquierda,[16] de todos los P_{kj} que ocurren durante el proceso. Es posible comnprobar que la matriz $P^{-1}A$ admite una eliminación gaussiana simple, de donde $P^{-1}A = LV$. En otras palabras, se puede empezar de nuevo con el sistema $A\boldsymbol{x} = \boldsymbol{b}$, ejecutando al inicio todos los intercambios de fila que serán eventualmente necesarias, para obtener un sistema $P^{-1}A\boldsymbol{x} = P^{-1}\boldsymbol{b}$; a partir de allí, se puede continuar con operaciones de fila del segundo tipo solamente.[17]

[15]Habrá al menos una fila j que cumple esta condición, si A es invertible. En la práctica, se aconseja elegir j tal que el valor absoluto de la entrada (k, j) sea el mayor posible.

[16]Cada P_{kj} es su propio inverso, $P_{kj}^{-1} = P_{kj}$. Por tanto, P^{-1} es el producto ordenado de los P_{kj} de izquierda a derecha. Por ejemplo, si $P = P_{68}P_{35}P_{23}$, entonces $P^{-1} = P_{23}P_{35}P_{68}$.

[17]Para un análisis detallado del método de eliminación gaussiana, véase: Gilbert W. Stewart, *Introduction to*

El algoritmo de eliminación termina sin éxito si en algún paso número k, la entrada diagonal (k, k) es cero y todas las entradas debajo de ésta en la columna k también son ceros. Si (y solo si) esto ocurre, el rango de la matriz A es menor que n y por ende A no es invertible.

► En el caso general, en donde A es una matriz rectangular $m \times n$, la existencia y unicidad de soluciones están dadas por la Proposición siguiente.

Proposición 1.36. *Si $A \in \mathbb{F}^{m\times n}$, el conjunto de soluciones del sistema* homogéneo *de ecuaciones lineales $A\boldsymbol{x} = \mathbf{0}$ es el subespacio* $\ker T_A \leqslant \mathbb{F}^n$*, de dimensión $n(A)$. Hay una solución única ($\boldsymbol{x} = \mathbf{0}$, por supuesto) si y solo si $n(A) = 0$, si y solo si $r(A) = n$.*

El sistema inhomogéneo *de ecuaciones lineales $A\boldsymbol{x} = \boldsymbol{b}$, con $\boldsymbol{b} \neq \mathbf{0}$, posee una solución si y solo si $\boldsymbol{b} \in T_A(\mathbb{F}^n)$, si y solo si el rango $r([A \mid \boldsymbol{b}])$ de la matriz aumentada es igual a $r(A)$. En ese caso, su conjunto de soluciones $\{\, \boldsymbol{x} : A\boldsymbol{x} = \boldsymbol{b} \,\}$ es el* subespacio afín $\boldsymbol{x}_0 + \ker T_A$*, donde $\boldsymbol{x}_0$ es alguna solución particular. Luego $A\boldsymbol{x} = \boldsymbol{b}$ posee solución única si y solo si $r([A \mid \boldsymbol{b}]) = r(A) = n$.* ⊟

Definición 1.37. Dícese que una matriz $A \in \mathbb{F}^{m\times n}$ está **en forma escalonada** si:

(a) hay algunas columnas iguales a los vectores iniciales de la base estándar de $\mathbb{F}^m$; es decir, $\boldsymbol{a}_{j_1} = \boldsymbol{e}_1, \boldsymbol{a}_{j_2} = \boldsymbol{e}_2, \dots, \boldsymbol{a}_{j_k} = \boldsymbol{e}_k$ para algún k;

(b) estas columnas aparecen en su orden natural: $j_1 < j_2 < \cdots < j_k$;

(c) si $j < j_1$, entonces $\boldsymbol{a}_j = \mathbf{0}$; si $j_r < j < j_{r+1}$, entonces los últimos $(m - r)$ elementos de $\boldsymbol{a}_j$ son ceros; y si $j > j_k$, entonces los últimos $(m - k)$ elementos de $\boldsymbol{a}_j$ son ceros.

Las columnas $\boldsymbol{a}_{j_1}, \dots, \boldsymbol{a}_{j_k}$ se llaman **columnas básicas**[18] de A. ◊

Una ejemplo de una matriz en forma escalonada es:

$$A = \begin{bmatrix} 0 & 1 & 2 & 0 & 1 & 7 \\ 0 & 0 & 0 & 1 & 5 & 2 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Aquí $m = 3$, $n = 6$, $k = 2$, $j_1 = 2$ y $j_2 = 4$. La tercera fila consta de ceros. De hecho, es una consecuencia de la Definición 1.37 que las últimas $(m - k)$ filas son nulas, si $k < m$.

Escolio 1.38. *Cualquier matriz $A \in \mathbb{F}^{m\times n}$ puede transformarse en una única forma escalonada mediante operaciones de fila.* ⊟

Matrix Computations, Academic Press, New York, 1973.

[18] Es evidente que dichas columnas forman una base del subespacio imagen $T_A(\mathbb{F}^n)$ generado por todas las columnas de A.

Escolio 1.39. *Si $A \in \mathbb{F}^{m \times n}$ es una matriz en forma escalonada con k columnas básicas, entonces $T_A(\mathbb{F}^n) = \mathrm{lin}\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_k \rangle$ y por ende $\underline{r(A) = k}$.* ⊟

Escolio 1.40. *Dos matrices $A, B \in \mathbb{F}^{m \times n}$ son equivalentes si y solo si se puede transformar A en B por operaciones de fila y de columna, si y solo si $\underline{r(A) = r(B)}$.* ⊟

Proposición 1.41. *Si la matriz $A \in \mathbb{F}^{m \times n}$ tiene rango k, hay matrices invertibles $Q \in M_m(\mathbb{F})$ y $P \in M_n(\mathbb{F})$ tales que*

$$QAP = \begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix},$$

en donde cada 0 *es un bloque rectangular de entradas nulas.*

Demostración. La matriz A puede reducirse a su forma escalonada B, que contiene k columnas básicas, mediante operaciones de fila solamente. La Proposición 1.34 muestra que hay una matriz invertible Q tal que $QA = B$. En seguida, se puede *intercambiar columnas* para colocar las columnas básicas de B en las primeras k posiciones, manteniendo el orden relativo de las columnas no básicas.

Una *operación de columna* sobre B se efectúa como sigue: (i) se transpone $B \mapsto B^{\mathrm{t}}$; (ii) se ejecuta la correspondiente operación de fila sobre B^{t} por una premultiplicación $B^{\mathrm{t}} \mapsto MB^{\mathrm{t}}$ donde M es una matriz de tipo (1.14); (iii) se transpone de nuevo, $MB^{\mathrm{t}} \mapsto BM^{\mathrm{t}}$. En resumen, una operación de columna elemental se efectúa al *posmultiplicar* B por cierta matriz invertible. Por tanto, una sucesión de operaciones de columna transforma B en BP', donde P' es una matriz invertible. El resultado de mover las columnas básicas de B a la izquierda es entonces

$$QAP' = \begin{bmatrix} 1_k & F \\ 0 & 0 \end{bmatrix}, \qquad (QAP')^{\mathrm{t}} = \begin{bmatrix} 1_k & 0 \\ F^{\mathrm{t}} & 0 \end{bmatrix},$$

en donde F es un bloque $k \times (n-k)$. Si $F \neq 0$, se aplica eliminación gaussiana *simple* a la matriz $(QAP')^{\mathrm{t}}$: esto reduce el bloque F^{t} a 0. Dicha eliminación se obtiene por una premultiplicación $(QAP')^{\mathrm{t}} \mapsto P''(QAP')^{\mathrm{t}}$ con P'' invertible. Sea $P := P'(P'')^{\mathrm{t}}$; entonces

$$QAP = (QAP')(P'')^{\mathrm{t}} = \big(P''(QAP')^{\mathrm{t}}\big)^{\mathrm{t}} = \begin{bmatrix} 1_k & 0 \\ 0 & 0 \end{bmatrix}.$$ □

1.5. Determinantes

El **determinante** de la matriz cuadrada $A := \begin{bmatrix} a & b \\ c & d \end{bmatrix} \in M_2(\mathbb{F})$ se define así:

$$\underline{\det A} \equiv \begin{vmatrix} a & b \\ c & d \end{vmatrix} := (ad - bc) \in \mathbb{F}. \tag{1.15}$$

Este es un escalar que *determina* si la matriz es invertible o no.

El criterio de invertibilidad es un valor no cero del determinante. En efecto, las identidades:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix} \begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} ad-bc & 0 \\ 0 & ad-bc \end{bmatrix}$$

muestran que A es invertible en $M_2(\mathbb{F})$ si y solo si $\det A \neq 0$, en cuyo caso

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad-bc} \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}. \tag{1.16}$$

El determinante de una matriz 3×3 se define "por expansión según la primera fila":

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &:= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}a_{22}a_{23} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{aligned} \tag{1.17a}$$

El determinante de una matriz cuadrada $A \in M_n(\mathbb{F})$ puede definirse por inducción, como sigue. Para $n = 1$, se define $\det [a_{11}] := a_{11} \in \mathbb{F}$. Ahora supóngase que el determinante ya está definido para matrices en $M_{n-1}(\mathbb{F})$. Si $A \in M_n(\mathbb{F})$, sea $\underline{A_{ij}}$ – con A mayúscula, nótese bien – la **submatriz** de A obtenida al borrar la fila i y la columna j de A. Cada A_{ij} es una matriz $(n-1) \times (n-1)$. El **menor** m_{ij} de la matriz A correspondiente a la entrada a_{ij} es el determinante de esa submatriz:

$$\underline{m_{ij}} := \det A_{ij}.$$

Obsérvese que la ecuación (1.17a) puede abreviarse con esta notación:

$$\det A = a_{11}\, m_{11} - a_{12}\, m_{12} + a_{13}\, m_{13}. \tag{1.17b}$$

Caso $\underline{n = 2}$: $m_{11} = a_{22}$ y $m_{12} = a_{21}$, luego $\det A = a_{11}a_{22} - a_{12}a_{21} = a_{11}m_{11} - a_{12}m_{12}$.

Definición 1.42. Si $A \in M_n(\mathbb{F})$; *se define* $\det A \in \mathbb{F}$ *por expansión según la primera fila*:

$$\underline{\det A} := a_{11}\, m_{11} - a_{12}\, m_{12} + \cdots + (-1)^{1+n} a_{1n}\, m_{1n} = \sum_{j=1}^{n} (-1)^{1+j} a_{1j}\, m_{1j}, \tag{1.18a}$$

o bien *por expansión según la fila* número i:

$$\det A := (-1)^{i+1} a_{i1}m_{i1} + (-1)^{i+2} a_{i2}m_{i2} + \cdots + (-1)^{i+n} a_{in}m_{in} = \sum_{j=1}^{n} (-1)^{i+j} a_{ij}\, m_{ij}, \tag{1.18b}$$

o bien *por expansión según la columna* número j:

$$\det A := (-1)^{1+j} a_{1j}m_{1j} + (-1)^{2+j} a_{2j}m_{2j} + \cdots + (-1)^{n+j} a_{nj}m_{nj} = \sum_{i=1}^{n} (-1)^{i+j} a_{ij}\, m_{ij}. \tag{1.18c}$$

El lema siguiente muestra que todas estas definiciones conducen al mismo resultado. ◊

Lema 1.43. *Si $A \in M_n(\mathbb{F})$, entonces $\underline{\det A}$, definido por cualquiera de las fórmulas en* (1.18)*, es igual a*[19]

$$\boxed{\det A = \sum_{\sigma \in S_n} (-1)^\sigma a_{1j_1} a_{2j_2} \dots a_{nj_n},} \tag{1.19}$$

donde la sumatoria recorre todas las $n!$ permutaciones $\sigma = (j_1, \dots, j_n)$ de $(1, 2, \dots, n)$ y el signo $\underline{(-1)^\sigma} = \pm 1$ vale $+1$ o -1 según sea la permutación σ par *o* impar.[20]

Demostración. Según la fórmula (1.18a), vale $\det A := \sum_{j_1=1}^n (-1)^{1+j_1} a_{1j_1} m_{1j_1}$, y además $m_{1j_1} = \det A_{1j_1}$ es una suma análoga de términos con $\pm a_{2j_2}$ multiplicado por menores correspondientes de la submatriz A_{1j_1}. Al repetir este argumento $(n-1)$ veces, $\det A$ queda expresado como suma de los productos que aparecen al lado derecho de (1.19), en donde cada producto contiene un factor tomado de cada fila y de columnas distintas. Hay n términos en (1.18a), $(n-1)$ términos en la expansión correspondiente a cada m_{1j_1}, etc., para un total de $n(n-1)(n-2)\cdots 3 \cdot 2 = n!$ términos en la expansión final. La suma de estos productos recorre todas las permutaciones posibles de $(1, 2, \dots, n)$.

Quedan por determinarse los signos en (1.19). En primer lugar, $(-1)^{1+j_1}$ es el signo de la transposición $1 \leftrightarrow j_1$. Por inducción sobre n, se comprueba que el producto de los diversos ± 1 que aparecen en la expansión iterativa de (1.18a) es efectivamente $+1$ si y solo si $(j_1, \dots, j_n)$ es una permutación par de $(1, 2, \dots, n)$.

El mismo argumento se aplica con cualquiera de las recetas (1.18b) ó (1.18c). En estos casos, la fila o columna de expansión se puede elegir arbitrariamente en cada iteración de la expansión. A lo sumo, podría ocurrir que el resultado final difiere del lado derecho de (1.19) por un múltiplo global de (± 1), que sería independiente de la matriz A. Un cálculo explícito muestra que todas las recetas en (1.18) dan $+1$ como el valor de $(\det 1_n)$, así que el desarrollo (1.19) es correcto en todos los casos. □

Proposición 1.44. *Si $A, B \in M_n(\mathbb{F})$, entonces* $\boxed{\det(AB) = (\det A)(\det B).}$

[19]La fórmula (1.19) se debe a *Gottfried Wilhelm Leibniz*, quien consideró la condición necesaria para resolver un sistema inhomogéneo de n ecuaciones de primer grado en $(n-1)$ variables. Esta condición es la anulación de la suma alternante de productos que aparece en (1.19). Así se expresó Leibniz en su carta del 28 de abril de 1693, dirigido al Marquis de l'Hôpital: *"Datis aequationibus quotcunque sufficientibus ad tollendas quantitates, quae simplicem gradum non egrediuntur, pro aequatione prodeunte primo sumendae sunt omnes combinationes possibiles, quas ingreditur una tantum coefficiens uniuscunque aequationis; secundo eae combinationes opposita habent signa, si in eodem prodeuntis aequationis latere ponantur, quae habent tot coefficentes communes, quot sunt unitates in numero quantitatum tollendarum unitate minuto; caeterae habent eadem signa"*. La notación de sumatoria de productos resume y clarifica esta descripción verbal.

[20]Una permutación σ es **par** si es el producto de un número par de transposiciones $i \leftrightarrow j$; σ es **impar** en el caso contrario. Si σ es el producto de k transposiciones, entonces $(-1)^\sigma := (-1)^k$ por definición.

Demostración. Sea $C := AB$. Por (1.19), $\det C$ es una suma de productos $\pm c_{1j_1}c_{2j_2}\cdots c_{nj_n}$. Cada c_{kj_k} es a su vez una suma de términos $\sum_{i_k=1}^{n} a_{ki_k}b_{i_kj_k}$ y por ende

$$\det(AB) = \sum (-1)^{\sigma} a_{1i_1}a_{2i_2}\dots a_{ni_n}\, b_{i_1j_1}b_{i_2j_2}\cdots_{i_nj_n}\,. \tag{1.20}$$

Esta suma extiende sobre las permutaciones $\sigma = (j_1,\dots,j_n)$ de $(1,2,\dots,n)$ y sobre toda posibilidad para $i_1,\dots,i_n$. Cuando dos de los i_k son iguales, la suma $\sum_\sigma (-1)^\sigma b_{i_1j_1}b_{i_2j_2}\cdots_{i_nj_n}$ se anula por cancelación de términos, así que aparecen en (1.20) solamente aquellos términos en donde $\tau = (i_1,\dots,i_n)$ es una permutación de $(1,2,\dots,n)$.

Sea $\rho = (r_1,\dots,r_n)$ la permutación de $(1,2,\dots,n)$ que transforma cada i_k en el j_k correspondiente. Entonces σ es la composición de las dos permutaciones τ y ρ, es decir, $\sigma = \rho\circ\tau$, así que $(-1)^\sigma = (-1)^\rho(-1)^\tau$. En conclusión,

$$\det(AB) = \Biggl(\sum_\tau (-1)^\tau a_{1i_1}a_{2i_2}\cdots a_{ni_n}\Biggr)\Biggl(\sum_\rho (-1)^\rho b_{1r_1}b_{2r_2}\cdots b_{nr_n}\Biggr) = (\det A)(\det B). \qquad \square$$

Proposición 1.45. *Las operaciones de fila elementales, aplicadas a una matriz $A \in M_n(\mathbb{F})$, cambian su determinante de las siguientes maneras:*

(a) *al multiplicar una fila de A por $c \neq 0$,* se multiplica $\det A$ por c *también;*

(b) *al sustraer de una fila de A un múltiplo de otra fila,* $\det A$ no cambia*;*

(c) *al intercambiar dos filas de A,* $\det A$ cambia de signo.

Demostración. Por las Proposiciones 1.34 y 1.44, basta verificar las igualdades:

$$\det M_i(c) = c, \qquad \det R_{ik}(c) = 1, \qquad \det P_{ik} = -1,$$

donde $M_i(c)$, $R_{ik}(c)$ y P_{ik} son las matrices definidas por (1.14).

Al expandir $\det A$ en cualquier fila i de una matriz A que tenga 1 en la diagonal y 0 en las demás entradas, la fórmula (1.18b) muestra que $\det A = (-1)^{i+i}m_{ii} = m_{ii}$. Por lo tanto, se puede eliminar la fila i y también la columna i de esa matriz A sin cambiar su determinante.

En las tres matrices definidas en (1.14), se puede eliminar así todas las filas y columnas excepto aquellas numeradas i y k – y para $M_i(c)$, se puede tomar cualquier k con $k \neq i$. La expansión según filas (1.18b) reduce el cálculo de estos determinantes al caso 2×2, en donde

$$\det M_i(c) = \begin{vmatrix} 1 & 0 \\ 0 & c \end{vmatrix} = c, \qquad \det R_{ik}(c) = \begin{vmatrix} 1 & 0 \\ -c & 1 \end{vmatrix} = 1, \qquad \det P_{ik} = \begin{vmatrix} 0 & 1 \\ 1 & 0 \end{vmatrix} = -1. \qquad \square$$

Proposición 1.46. *Si $A \in M_n(\mathbb{F})$, entonces* $\boxed{\det A^{\mathrm{t}} = \det A.}$

Demostración. La fórmula (1.19), aplicada a la transpuesta de A, da

$$\det A^{\mathrm{t}} = \sum_{\sigma \in S_n} (-1)^\sigma a_{j_1 1} a_{j_2 2} \dots a_{j_n n};$$

Sea $\pi \equiv (p_1, \dots, p_n) := \sigma^{-1}$ la permutación recíproca que lleva cada j_k en k. Si σ es el producto de m transposiciones, π es el producto de las mismas m transposiciones en el orden inverso: por lo tanto $(-1)^\pi = (-1)^\sigma$. Luego,

$$\det A^{\mathrm{t}} = \sum_{\pi \in S_n} (-1)^\pi a_{1p_1} a_{2p_2} \dots a_{np_n} = \det A. \qquad \square$$

Proposición 1.47. *Si $A \in M_n(\mathbb{F})$ es una matriz triangular, su determinante es el producto de los elementos diagonales de A, esto es,* $\det A = a_{11}a_{22} \cdots a_{nn}$. *En particular, vale* $\det 1_n = 1$.

Demostración. Supóngase que A es una matriz triangular inferior. La expansión (1.18a) según la primera fila da $\det A = a_{11}\, m_{11}$. La submatriz A_{11} también es triangular inferior, y su determinante es $m_{11} = a_{22}\, m_{12,12}$, donde $m_{12,12}$ es el menor correspondiente a la entrada a_{22}. Al repetir este argumento $(n-2)$ veces, se obtiene

$$\det A = a_{11}a_{22} \dots a_{n-2,n-2} \begin{vmatrix} a_{n-1,n-1} & 0 \\ a_{n,n-1} & a_{nn} \end{vmatrix} = a_{11}a_{22} \dots a_{nn}.$$

Si A es triangular superior, aplíquese el mismo argumento (o bien la Proposición 1.46). $\square$

Escolio 1.48. *Si $A \in M_n(\mathbb{F})$ tiene dos filas iguales, o dos columnas iguales; o bien si una fila o una columna de A es nula, entonces* $\det A = 0$. $\boxminus$

Proposición 1.49. $\det A = 0$ *si y solo si A es singular (es decir, no invertible).*

Demostración. Si A es invertible, entonces $1 = \det 1_n = (\det A)(\det A^{-1})$, así que $\det A$ no puede ser cero.

Considérese el efecto de aplicar a la matriz A unos k pasos del algoritmo de eliminación gaussiana, con intercambio de filas cuando sea necesario, usando las operaciones de fila de los tipos segundo y tercero. El resultado de este proceso es una matriz de la forma

$$A' = \begin{bmatrix} U' & X' \\ 0 & Y' \end{bmatrix}, \quad \text{con} \quad U' \in M_k(\mathbb{F}), \quad X' \in \mathbb{F}^{k \times (n-k)}, \quad Y' \in M_{n-k}(\mathbb{F});$$

donde U' es triangular superior y sus elementos diagonales son los *pivotes* $a_{11}, a_{22}^{(2)}, \dots, a_{kk}^{(k)}$; aquí 0 es un bloque rectangular de ceros. La Proposición 1.45 muestra que $\det A' = \pm \det A$. Al expandir $\det A'$ según la primera columna k veces, se obtiene

$$\det A' = a_{11}a_{22}^{(2)} \dots a_{kk}^{(k)} \det Y'.$$

Si A no es invertible, el algoritmo de eliminación gaussiana se detiene en el paso número k para algún $k \leqslant n$, porque $a_{kk}^{(k)} = 0$. Entonces $\det A' = 0$ y por ende $\det A = 0$.

Por otro lado, si A es invertible, se puede ejecutar n pasos de la eliminación, hasta llegar al último pivote $a_{nn}^{(n)} \neq 0$. Esto conduce a una importante *fórmula para el determinante* de una matriz invertible:

$$\det A = (-1)^r a_{11} a_{22}^{(2)} \dots a_{nn}^{(n)} \neq 0, \tag{1.21}$$

donde r es el número de intercambios de filas que ocurren en la eliminación. □

En términos de la factorización $A = PLV$ discutido anteriormente, se puede notar que $\det P = (-1)^r$, $\det L = 1$ por ser L una matriz triangular unipotente, $\det V = a_{11} a_{22}^{(2)} \dots a_{nn}^{(n)}$ por ser V triangular superior con los pivotes en la diagonal.

► Las matrices rectangulares que no son cuadradas no tienen determinantes. Sin embargo, vale la pena considerar los determinantes de sus *submatrices cuadradas*.

Proposición 1.50. *Sea $A \in \mathbb{F}^{m\times n}$. Su rango $r(A)$ es el mayor entero $\underline{k}$ tal que A posee una submatriz B de dimensiones $k \times k$ con* $\det B \neq 0$.

Demostración. Colóquese $k := r(A)$. Entonces hay k *columnas* linealmente independientes en A (aquellas que se convierten en columnas básicas al reducir A a su forma escalonada mediante operaciones de fila). Sea C la submatriz $m \times k$ de A que se obtiene al borrar las otras columnas de A.

Ahora $r(C^t) = r(C) = k$; luego C tiene k *filas* linealmente independientes. Sea B la submatriz $k \times k$ de C (y por ende de A) que se obtiene al borrar las otras filas de C. Entonces $B \in M_k(\mathbb{C})$ con $r(B) = k$, así que B es invertible y por tanto $\det B \neq 0$.

Si $k < \min\{m, n\}$, sea D cualquier submatriz $(k+1) \times (k+1)$ de A. Las columnas de D forman parte de $(k+1)$ columnas de A, que cumplen una relación de dependencia lineal. Luego las columnas de D cumplen la misma relación de dependencia lineal, así que $r(D) \leqslant k$; por lo tanto, D es singular y por ende $\det D = 0$. □

► La solución de un sistema de ecuaciones lineales, cuyo número de incógnitas es igual al número de ecuaciones, puede expresarse mediante determinantes. La fórmula correspondiente se llama **la regla de Cramer**.[21] En la práctica, es un método ineficiente para sistemas con más de tres variables; pero tiene importancia teórica. Entre otras cosas, muestra que un sistema de ecuaciones con coeficientes enteros tiene soluciones racionales.

Definición 1.51. Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada. El **cofactor** de su entrada a_{ij} es $(-1)^{i+j} m_{ji}$, donde el *menor* $m_{ji} = \det A_{ji}$ es el determinante de la submatriz A_{ji} obtenida al

[21] Esta regla, aparentemente independiente del trabajo anterior de Leibniz, aparece por primera vez en: Gabriel Cramer, *Introduction à l'analyse des lignes courbes algébriques*, Ginebra, 1750. El japonés Takakazu Seki, contemporáneo de Leibniz, ya había dado el caso 3×3 de la regla en 1683.

remover la fila j y la columna i de A. Fíjese que $m_{ij} = \det(A^{\mathrm{t}})_{ij}$ es un menor de la matriz transpuesta A^{t}.

La matriz **matriz adjugada**[22] de A es la matriz $\operatorname{adj} A \in M_n(\mathbb{F})$ cuya entrada (i, j) es el cofactor de a_{ij}. Para obtener $\operatorname{adj} A$, es cuestión de (i) reemplazar cada elemento a_{ij} de A por el menor m_{ij}; (ii) multiplicar cada entrada por el signo $(-1)^{i+j}$ que corresponde a su lugar en un "tablero de ajedrez"; (iii) tomar la transpuesta de la matriz resultante. ◊

Obsérvese, en particular, la fórmula: $\operatorname{adj}\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \begin{bmatrix} d & -b \\ -c & a \end{bmatrix}$.

Proposición 1.52. *Si $A \in M_n(\mathbb{F})$, entonces*

$$\boxed{A\,(\operatorname{adj} A) = (\operatorname{adj} A)\,A = (\det A)\,1_n.} \tag{1.22}$$

Demostración. Obsérvese primero que la fórmula (1.18b) corresponde al producto de la fila i de A por la *columna* i de $(\operatorname{adj} A)$. Por otro lado, la fórmula (1.18c) representa el producto de la *fila* j de $(\operatorname{adj} A)$ por la columna j de A. Estas dos fórmulas juntas expresan que todo elemento *diagonal* de los productos $A\,(\operatorname{adj} A)$ y $(\operatorname{adj} A)\,A$ es igual a $\det A$.

Al multiplicar la fila k de A por la columna i de $(\operatorname{adj} A)$, con $k \neq i$, se obtiene

$$(-1)^{i+1}a_{k1}m_{i1} + (-1)^{i+2}a_{k2}m_{i2} + \cdots + (-1)^{i+n}a_{kn}m_{in} = \sum_{j=1}^{n}(-1)^{i+j}a_{kj}m_{ij}. \tag{1.23}$$

Esta es el determinante de la matriz A' obtenida al reemplazar la fila i de A por su fila k. Entonces las filas i y k de A' son iguales, así que $\det A' = 0$ por el Escolio 1.48. Luego la expresión (1.23) vale 0 cuando $k \neq i$. De igual modo, el producto punto de la fila l de A por la columna j de $(\operatorname{adj} A)$, con $l \neq j$, se anula. Luego, la entrada (i, j) de $A\,(\operatorname{adj} A)$ o de $(\operatorname{adj} A)\,A$ es $(\det A)\,[\![i = j]\!]$; lo cual comprueba las igualdades (1.22). □

La fórmula (1.22) proporciona *una fórmula para la matriz inversa* de una matriz no singular. Si $\det A \neq 0$, entonces

$$\boxed{A^{-1} = \frac{1}{\det A}\operatorname{adj} A.} \tag{1.24}$$

En el caso 2×2, esta relación toma la forma:

$$\begin{bmatrix} a & b \\ c & d \end{bmatrix}^{-1} = \frac{1}{ad - bc}\begin{bmatrix} d & -b \\ -c & a \end{bmatrix}.$$

[22] La matriz $\operatorname{adj} A$ se bautiza la *matriz adjunta* de A por algunos autores. Sin embargo, cuando $\mathbb{F} = \mathbb{C}$, ese término está reservado para el conjugado hermítico A^*, que se verá más adelante. Aquí se prefiere usar la palabra **adjugada**, una mezcla inelegante de "adjunta" y "conjugada", con disculpas a la Real Academia Española.

Proposición 1.53 (Regla de Cramer)**.** *Sean $A \in M_n(\mathbb{F})$, $\boldsymbol{b} \in \mathbb{F}^n$. Para cada $j = 1, \dots, n$, sea $B_j := [\boldsymbol{a}_1 \dots \boldsymbol{a}_{j-1}\, \boldsymbol{b}\, \boldsymbol{a}_{j+1} \dots \boldsymbol{a}_n]$ la matriz obtenida de A al reemplazar su columna $\boldsymbol{a}_j$ por $\boldsymbol{b}$. Entonces el sistema de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$ tiene solución única $\boldsymbol{x} \in \mathbb{F}^n$ si y solo si* $\det A \neq 0$, *en cuyo caso*

$$x_j = \frac{\det B_j}{\det A}, \quad \textit{para} \quad j = 1, \dots, n. \tag{1.25}$$

Demostración. Ya se sabe que $A\boldsymbol{x} = \boldsymbol{b}$ tiene solución única solo si $A\boldsymbol{x} = \boldsymbol{0}$ tiene solución única, solo si $\ker T_A = \{\boldsymbol{0}\}$, solo si $n(A) = 0$, solo si $r(A) = n$, solo si A es invertible, solo si $\det A \neq 0$. Por otro lado, si $\det A \neq 0$, entonces $\boldsymbol{x} = A^{-1}\boldsymbol{b}$ es la solución única.

Si $\det A \neq 0$, al premultiplicar ambos lados de la ecuación $A\boldsymbol{x} = \boldsymbol{b}$ por $(\operatorname{adj} A)$, se obtiene la ecuación

$$(\det A)\, \boldsymbol{x} = (\operatorname{adj} A)\, \boldsymbol{b}.$$

Para cada j, estos dos vectores de columna tienen las coordenadas j siguientes:

$$(\det A)x_j = (\text{fila } j \text{ de } \operatorname{adj} A) \cdot \boldsymbol{b} = \sum_{i=1}^{n} (-1)^{i+j} m_{ij} b_i = \det B_j,$$

al usar la expansión (1.18c) en la columna j para evaluar $\det B_j$. La fórmula (1.25) sigue, al dividir esta relación por $\det A$. □

1.6. Ejercicios sobre espacios vectoriales y matrices

Ejercicio 1.1. (a) Demostrar que los tres vectores $(1,1,0)$, $(1,1,1)$, $(0,1,-1)$ son linealmente independientes en $\mathbb{R}^3$. Expresar los vectores $\boldsymbol{e}_1$, $\boldsymbol{e}_2$, $\boldsymbol{e}_3$ de la base estándar como combinaciones lineales de ellos.

(b) Mostrar que los tres polinomios $\frac{1}{2}t(t-1)$, $1-t^2$, $\frac{1}{2}t(t+1)$ son linealmente independientes en $\mathbb{R}[t]$. Expresar los monomios 1, t, t^2 como combinaciones lineales de ellos.

Ejercicio 1.2. Si $p, q, r, s \in \mathbb{R}$ son distintos, demostrar que los cuatro vectores

$$(1,1,1,1), \quad (p,q,r,s), \quad (p^2,q^2,r^2,s^2), \quad (p^3,q^3,r^3,s^3)$$

son linealmente independientes en $\mathbb{R}^4$. ⟦ Indicación: si el polinomio $a_0 + a_1 t + a_2 t^2 + a_3 t^3$ no es el polinomio nulo, tiene a lo sumo tres raíces distintas. ⟧

Ejercicio 1.3. Demostrar que $\{1, (t-1), (t-1)^2, \dots, (t-1)^n\}$ es una base del espacio vectorial $\mathbb{F}_n[t]$ de polinomios de grado no mayor que n (Ejemplo 1.4).

Ejercicio 1.4. Demostrar el Escolio 1.9 (prolongación de una base parcial).

Ejercicio 1.5. Sean U y W dos subespacios de un espacio $\mathbb{F}$-vectorial V de dimensión finita. Comprobar que su intersección $U \cap W$ y su suma $U + W$ también son subespacios de V. Verificar la fórmula:

$$\dim(U + W) = \dim U + \dim W - \dim(U \cap W).$$

$[\![$ Indicación: una base dada de $U \cap W$ se puede prolongar de dos maneras para formar bases de U y de W. Mostrar que la unión de esas dos bases es una base de $U + W$. $]\!]$

Ejercicio 1.6. (a) Sean $c_0, c_1, \dots, c_n \in \mathbb{F}$ escalares distintos y sean $\{ \pi_k : k = 0, 1, \dots, n \}$ los polinomios en $\mathbb{F}_n[t]$ dados por

$$\pi_k(t) := \frac{(t - c_0) \cdots (t - c_{k-1})(t - c_{k+1}) \cdots (t - c_n)}{(c_k - c_0) \cdots (c_k - c_{k-1})(c_k - c_{k+1}) \cdots (c_k - c_n)}.$$

Ellos son los *polinomios interpolativos de Lagrange* para los "nudos" $c_0, c_1, \dots, c_n$. Verificar que $\pi_k(c_j) = [\![k = j]\!]$.

(b) Concluir que $\{\pi_0(t), \pi_1(t), \dots, \pi_n(t)\}$ es una base para $\mathbb{F}_n[t]$.

(c) Demostrar que la base dual del espacio dual $\mathbb{F}_n[t]^*$ es el juego de las *evaluaciones* $f_j \colon \mathbb{F}_n[t] \to \mathbb{F}$ definidas por $f_j(p(t)) := p(c_j)$, para $j = 0, 1, \dots, n$.

Ejercicio 1.7. Encontrar los subespacios $\underline{\ker T}$ y $\underline{T(\mathbb{R}^3)}$ y hallar las dimensiones $n(T)$ y $r(T)$, si la matriz de $T \in \mathcal{L}(\mathbb{R}^3, \mathbb{R}^3)$ con respecto de la base estándar $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3\}$ es

$$A = \begin{bmatrix} 1 & 2 & 3 \\ 2 & 4 & 6 \\ 3 & 6 & 9 \end{bmatrix}.$$

Ejercicio 1.8. Si $T \in \mathcal{L}(V, W)$ y $S \in \mathcal{L}(W, Z)$, demostrar las siguientes relaciones entre núcleos e imágenes:

$$\begin{aligned} \ker(ST) &\supseteq \ker T, & \ker((ST)^{\mathrm{t}}) &\supseteq \ker(S^{\mathrm{t}}), \\ ST(V) &\subseteq S(W), & (ST)^{\mathrm{t}}(Z^*) &\subseteq T^{\mathrm{t}}(W^*). \end{aligned}$$

Concluir que $\underline{r(ST) \leqslant r(S)}$ y que $\underline{r(ST) \leqslant r(T)}$.

Ejercicio 1.9. Las funciones continuas reales sobre un intervalo $[a, b] \subset \mathbb{R}$ forman el espacio $\mathbb{R}$-vectorial $\underline{C[a, b]} := \{ f \colon [a, b] \to \mathbb{R} \text{ continua} \}$. Defínase $T \colon C[a, b] \to C[a, b]$ por

$$\underline{Tf}(x) := \int_a^x f(y)\, dy.$$

Demostrar que T es lineal e inyectiva, pero no es sobreyectiva. Concluir que $C[a, b]$ es infinitodimensional.

Ejercicio 1.10. Si $A \in M_n(\mathbb{F})$ es una matriz invertible, demostrar que $(A^{\mathrm{t}})^{-1} = (A^{-1})^{\mathrm{t}}$. Concluir que A^{-1} es simétrica si y solo si A es simétrica.

Ejercicio 1.11. Calcular (por inducción sobre n) las potencias A^n, B^n, C^n de las siguientes matrices:

$$A := \begin{bmatrix} a & 1 \\ 0 & a \end{bmatrix}, \qquad B := \begin{bmatrix} a & 1 & 0 \\ 0 & a & 1 \\ 0 & 0 & a \end{bmatrix}, \qquad C := \begin{bmatrix} 1 & a \\ 0 & 1 \end{bmatrix}, \qquad \text{donde} \quad a \in \mathbb{F}.$$

Ejercicio 1.12. La base estándar para el espacio vectorial $M_2(\mathbb{F})$ es $\mathcal{E} := \{E_{11}, E_{12}, E_{21}, E_{22}\}$, donde

$$E_{11} = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad E_{12} = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}, \quad E_{21} = \begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}, \quad E_{22} = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}.$$

Si $A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, defínase

$$\underline{L_A}(B) := AB, \quad \underline{R_A}(B) := BA, \quad \text{y} \quad T(B) := B^{\mathrm{t}}.$$

Demostrar que L_A, R_A y T son aplicaciones lineales de $M_2(\mathbb{F})$ en sí mismo y calcular sus matrices 4×4 con respecto a la base estándar.

Ejercicio 1.13. Sea A una matriz triangular superior con ceros en la diagonal:

$$A := \begin{bmatrix} 0 & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & 0 & a_{23} & \dots & a_{2n} \\ 0 & 0 & 0 & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 0 \end{bmatrix},$$

esto es, $a_{ij} = 0$ para $i \geqslant j$. Demostrar que $A^n = 0$. Concluir que $1_n + A$ es invertible, con

$$(1_n + A)^{-1} = 1_n - A + A^2 - \cdots + (-1)^{n-1} A^{n-1}.$$

Usar esta relación para calcular el inverso de la matriz $B = \begin{bmatrix} 1 & a & b \\ 0 & 1 & c \\ 0 & 0 & 1 \end{bmatrix}$.

Ejercicio 1.14. Resolver el sistema de ecuaciones $A\boldsymbol{x} = \boldsymbol{b}$, con

$$A := \begin{bmatrix} 2 & 1 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ 0 & 1 & 2 & 1 \\ 0 & 0 & 1 & 2 \end{bmatrix}, \qquad \boldsymbol{b} := \begin{bmatrix} 2 \\ 1 \\ 4 \\ 8 \end{bmatrix},$$

por el método de eliminación gaussiana. Usar el resultado del cálculo para escribir explícitamente las matrices L, D, U de la factorización $A = LDU$.

Ejercicio 1.15. Para la matriz aumentada siguiente,

$$[A \mid \boldsymbol{b}] := \left[\begin{array}{ccccc|c} 1 & -1 & 2 & 0 & 3 & -1 \\ 0 & 2 & 1 & 3 & 1 & -4 \\ -1 & 1 & 5 & 1 & 0 & 3 \\ -1 & 0 & 1 & -1 & -2 & 4 \end{array}\right],$$

(a) calcular el rango $r(A)$;

(b) encontrar una base del espacio de soluciones de $A\boldsymbol{x} = \boldsymbol{0}$; y

(c) describir el conjunto de soluciones de $A\boldsymbol{x} = \boldsymbol{b}$.

Ejercicio 1.16. Demostrar que cada matriz de rango k es una suma de k matrices de rango 1. ⟦ Indicación: Usar la Proposición 1.41. ⟧

Ejercicio 1.17. Verificar el *determinante de Vandermonde*:

$$\begin{vmatrix} 1 & x & x^2 & x^3 \\ 1 & y & y^2 & y^3 \\ 1 & z & z^2 & z^3 \\ 1 & w & w^2 & w^3 \end{vmatrix} = (x-y)(x-z)(x-w)(y-z)(y-w)(z-w).$$

⟦ Indicación: Usar eliminación gaussiana. ⟧

Ejercicio 1.18. Verificar (por eliminación gaussiana) que

$$\begin{vmatrix} 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & -1 & 1 \\ 1 & -1 & 1 & -1 & 1 & 1 \\ 1 & -1 & -1 & 1 & 1 & -1 \end{vmatrix} = -160.$$

Ejercicio 1.19. Si $A \in M_m(\mathbb{F})$, $B \in \mathbb{F}^{m\times n}$, $C \in \mathbb{F}^{n\times m}$, $D \in M_n(\mathbb{F})$, demostrar que

$$\det \begin{bmatrix} A & B \\ 0 & D \end{bmatrix} = \det \begin{bmatrix} A & 0 \\ C & D \end{bmatrix} = (\det A)(\det D),$$

si en ambos casos 0 representa un rectángulo de ceros.

Ejercicio 1.20. Obtener el rango de la matriz

$$A := \begin{bmatrix} 3 & 5 & 1 & 4 \\ 2 & -1 & 1 & 1 \\ 5 & 4 & 2 & 5 \end{bmatrix}$$

(a): por cálculo de menores; y también (b): por un cambio a forma escalonada.

Ejercicio 1.21. Calcular la matriz adjugada $(\operatorname{adj} A)$ y la matriz inversa A^{-1} para

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

Ejercicio 1.22. (a) Si $A \in M_n(\mathbb{F})$ con $n \geqslant 2$, demostrar que $\det(\operatorname{adj} A) = (\det A)^{n-1}$.

(b) Concluir que $\operatorname{adj}(\operatorname{adj} A) = (\det A)^{n-2} A$ si $n \geqslant 3$.

⟦ Indicación: considerar el producto de tres matrices $A\,(\operatorname{adj} A)\,\operatorname{adj}(\operatorname{adj} A)$. ⟧

Ejercicio 1.23. Si $A \in M_n(\mathbb{F})$; $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{F}^n$; y $d \in \mathbb{F}$, sea $\begin{bmatrix} A & \boldsymbol{b} \\ \boldsymbol{c}^{\mathrm{t}} & d \end{bmatrix}$ la matriz $(n+1) \times (n+1)$ formada al bordear A por la columna $\boldsymbol{b}$, la fila $\boldsymbol{c}^{\mathrm{t}}$ y la entrada d. Calcular el producto de matrices

$$\begin{bmatrix} 1_n & \mathbf{0} \\ -\boldsymbol{c}^{\mathrm{t}} & \det A \end{bmatrix} \begin{bmatrix} \operatorname{adj} A & \mathbf{0} \\ \mathbf{0}^{\mathrm{t}} & 1 \end{bmatrix} \begin{bmatrix} A & \boldsymbol{b} \\ \boldsymbol{c}^{\mathrm{t}} & d \end{bmatrix}.$$

En seguida, demostrar que

$$\det \begin{bmatrix} A & \boldsymbol{b} \\ \boldsymbol{c}^{\mathrm{t}} & d \end{bmatrix} = d \det A - \boldsymbol{c}^{\mathrm{t}}(\operatorname{adj} A)\boldsymbol{b}.$$

⟦ Indicación: usar el resultado del Ejercicio 1.22(a). ⟧

Ejercicio 1.24. (a) Sean (x, y) las coordenadas de un punto del plano $\mathbb{R}^2$. Demostrar que la recta que pasa por dos puntos (x_1, y_1) y (x_2, y_2) tiene la ecuación

$$\begin{vmatrix} 1 & x & y \\ 1 & x_1 & y_1 \\ 1 & x_2 & y_2 \end{vmatrix} = 0.$$

(b) Demostrar que el círculo que pasa por tres puntos (x_1, y_1), (x_2, y_2) y (x_3, y_3) del plano tiene la ecuación

$$\begin{vmatrix} 1 & x & y & x^2 + y^2 \\ 1 & x_1 & y_1 & x_1^2 + y_1^2 \\ 1 & x_2 & y_2 & x_2^2 + y_2^2 \\ 1 & x_3 & y_3 & x_3^2 + y_3^2 \end{vmatrix} = 0.$$

⟦ Indicación: comprobar que esta ecuación representa un círculo; y que los los tres puntos dados la satisface. ⟧

2 Estructura de aplicaciones lineales

Alles sollte so einfach wie möglich gemacht sein,
aber nicht einfacher.

— Albert Einstein

La esencia del álgebra lineal es el estudio de las propiedades de aplicaciones lineales. En este capítulo se analizará la estructura de un operador lineal, esto es, una aplicación lineal de un espacio vectorial V en sí mismo. Por ahora, se considera la situación en donde V es finitodimensional sobre un cuerpo $\mathbb{F}$ cualquiera. (Se difiere al siguiente capítulo las consecuencias de dotar V de un producto escalar.) Cada aplicación lineal determina dos polinomios que revelan su estructura: su polinomio característico y su polinomio mínimo. También está acompañado de un juego de escalares (sus autovalores) que la distinguen.

2.1. Autovalores y autovectores

Definición 2.1. Sea V un espacio vectorial finitodimensional sobre un cuerpo $\mathbb{F}$. Un **operador lineal** sobre V es una aplicación lineal $T\colon V \to V$. El espacio vectorial $\mathcal{L}(V, V)$ de todos los operadores $\mathbb{F}$-lineales sobre V se denotará[1] por $\underline{\mathcal{L}(V)}$.

El espacio vectorial $\mathcal{L}(V)$ es también un *anillo*, cuya operación multiplicativa es la *composición de operadores*. En efecto, para $R, S, T \in \mathcal{L}(V)$, esta composición:

(a) es *asociativa*: $R(ST) = (RS)T$;

(b) tiene elemento identidad $1 \equiv 1_V : \boldsymbol{x} \mapsto \boldsymbol{x}$ que cumple $1\,T = T\,1 = T$;

(c) es *compatible* con la multiplicación escalar: $c(ST) = (cS)T = S(cT)$ para $c \in \mathbb{F}$;

(d) es *distributiva* sobre la suma: $T(R + S) = TR + TS$ y $(R + S)T = RT + ST$.

En otras palabras,[2] $\mathcal{L}(V)$ es un **álgebra** sobre $\mathbb{F}$. ◊

Definición 2.2. Sea V un espacio $\mathbb{F}$-vectorial y sea $T \in \mathcal{L}(V)$. Un **autovalor** de T es un escalar $\lambda \in \mathbb{F}$ tal que la ecuación

$$\boxed{T(\boldsymbol{x}) = \lambda\,\boldsymbol{x}} \tag{2.1}$$

tenga una solución $\boldsymbol{x} \neq \mathbf{0}$. Un vector *no nulo* $\boldsymbol{x} \in V$ que cumple (2.1) se llama un **autovector** asociado al autovalor λ. ⟦ La ecuación $T(\boldsymbol{x}) = \lambda\,\boldsymbol{x}$ tiene la solución trivial $\boldsymbol{x} = \mathbf{0}$ cualquiera que sea el coeficiente λ. Se descarta siempre la solución trivial: por regla, el vector $\mathbf{0}$ nunca puede ser autovector de un operador lineal. ⟧

[1] En las ocasiones cuando sea necesario indicar el cuerpo $\mathbb{F}$ de escalares, se escribe $\mathcal{L}_{\mathbb{F}}(V)$ en vez de $\mathcal{L}(V)$.

[2] Una $\mathbb{F}$-*álgebra* es un espacio $\mathbb{F}$-vectorial y un anillo a la vez, con operaciones algebraicas compatibles.

Algunos autores llaman a λ un *valor propio* y a $\boldsymbol{x}$ un *vector propio* de T.[3] $\lozenge$

Si $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base (ordenada) de V, la expansión $\boldsymbol{x} = c_1\boldsymbol{x}_1 + \cdots + c_n\boldsymbol{x}_n$ determina un isomorfismo lineal $V \to \mathbb{F}^n : \boldsymbol{x} \mapsto \boldsymbol{c} = [\boldsymbol{x}]^{\mathcal{B}}$ dado por (1.3). A su vez, la fórmula (1.9) es aplicable con la misma base en el dominio y el codominio de T, es decir,

$$T(\boldsymbol{x}_j) =: \sum_{i=1}^{n} a_{ij}\,\boldsymbol{x}_i, \qquad A = [T]_{\mathcal{B}}^{\mathcal{B}}. \tag{2.2}$$

De la fórmula (1.10) en la Sección 1.3, se sabe que $[T(\boldsymbol{x})]^{\mathcal{B}} = A\,[\boldsymbol{x}]^{\mathcal{B}}$, de modo que las correspondencias $T \leftrightarrow A \leftrightarrow T_A$ establecen isomorfismos lineales entre $\mathcal{L}(V)$, $M_n(\mathbb{F})$ y $\mathcal{L}(\mathbb{F}^n)$. Estas correspondencias convierten la composición de operadores en multiplicación de matrices y viceversa, de modo que estas tres $\mathbb{F}$-álgebras *son isomorfos como álgebras*.

Así las cosas, cada propiedad de aplicaciones lineales induce una propiedad paralela de matrices. Por ejemplo, las matrices también poseen autovalores y autovectores.

Definición 2.3. Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada. Un **autovalor de** A es un autovalor de T_A, es decir, un escalar $\lambda \in \mathbb{F}$ tal que la ecuación

$$\boxed{A\boldsymbol{x} = \lambda\,\boldsymbol{x}}$$

tenga solución $\boldsymbol{x} \neq \boldsymbol{0}$ en $\mathbb{F}^n$. Un vector no nulo $\boldsymbol{x} \in \mathbb{F}^n$ tal que $A\boldsymbol{x} = \lambda\,\boldsymbol{x}$ es un **autovector de** A asociado al autovalor λ. $\lozenge$

Lema 2.4. *Sea V un espacio $\mathbb{F}$-vectorial finitodimensional. Para un operador lineal $T \in \mathcal{L}(V)$ y $\lambda \in \mathbb{F}$, son equivalentes las siguientes condiciones:*

(a) *λ es un autovalor de T;*

(b) *el operador lineal $(T - \lambda\,1)$ no es invertible en $\mathcal{L}(V)$;*

(c) $\ker(T - \lambda\,1) \neq \{\boldsymbol{0}\}$.

Demostración. Ad (a) $\iff$ (b): Las siguientes afirmaciones dicen lo mismo:

$$\begin{aligned} \lambda \text{ es un autovalor de } T &\iff \text{hay } \boldsymbol{x} \in V \text{ con } \boldsymbol{x} \neq \boldsymbol{0} \text{ y } T(\boldsymbol{x}) = \lambda\boldsymbol{x} \\ &\iff \text{hay } \boldsymbol{x} \neq \boldsymbol{0} \text{ tal que } (T - \lambda\,1)(\boldsymbol{x}) = \boldsymbol{0} \\ &\iff (T - \lambda\,1) \text{ no es inyectivo} \\ &\iff (T - \lambda\,1) \text{ no es biyectivo.} \end{aligned}$$

[3] La terminología viene en primera instancia del alemán, cuando David Hilbert en 1904 empleó la palabra *Eigenwert*, en un artículo sobre ecuaciones integrales: *Eigen* = auto, *wert* = valor. (Huyan de las traducciones parciales tales como “eigenvalores” y “eigenvectores”.) La usanza moderna aparece en: John von Neumann, “Allgemeine Eigenwerttheorie Hermitescher Funktionaloperatoren”, *Mathematische Annalen* **102** (1929), 49–131. Von Neumann declaró: *ein Eigenwert ist eine Zahl λ, zu der es eine Funktion $f \neq 0$ mit $Rf = \lambda f$ gibt; f ist dann Eigenfunktion*.

La última equivalencia se debe a que $n(T - \lambda\,1) + r(T - \lambda\,1) = \dim V$, por el Teorema 1.23 (de rango y nulidad); por lo tanto, *un operador lineal es inyectivo si y solo si es sobreyectivo*.[4]

Ad (b) $\iff$ (c): Hay un vector $\boldsymbol{x} \neq \boldsymbol{0}$ tal que $(T - \lambda\,1)(\boldsymbol{x}) = \boldsymbol{0}$ si y solo si hay un vector $\boldsymbol{x} \in \ker(T - \lambda\,1)$ con $\boldsymbol{x} \neq \boldsymbol{0}$. □

Lema 2.5. *Para una matriz cuadrada $A \in M_n(\mathbb{F})$ y $\lambda \in \mathbb{F}$, estas condiciones son equivalentes:*

(a) *λ es un autovalor de A;*

(b) *la matriz $(A - \lambda\,1_n)$ no es invertible en $M_n(\mathbb{F})$;*

(c) $\det(A - \lambda\,1_n) = 0$.

Demostración. La equivalencia (a) $\iff$ (b) sigue del lema anterior, para el caso $T = T_A$. La equivalencia (b) $\iff$ (c) sigue de la Proposición 1.49. □

Corolario 2.6. *Los autovalores de una matriz* triangular $A \in M_n(\mathbb{F})$ *son sus elementos diagonales $a_{11}, a_{22}, \dots, a_{nn}$.*

Demostración. Supóngase que A es triangular superior, es decir, $a_{ij} = 0$ para $i > j$. Si λ es un autovalor de A, entonces $\det(A - \lambda\,1_n) = 0$ o bien, lo que es lo mismo, $\det(\lambda\,1_n - A) = 0$. Explícitamente,

$$\det(\lambda\,1_n - A) = \begin{vmatrix} \lambda - a_{11} & -a_{12} & \dots & -a_{1n} \\ 0 & \lambda - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda - a_{nn} \end{vmatrix} = (\lambda - a_{11})(\lambda - a_{22}) \cdots (\lambda - a_{nn}),$$

porque la matriz $\lambda\,1_n - A$ también es triangular superior. Entonces λ es un autovalor de A si y solo si $\lambda - a_{kk} = 0$ para algún k, si y solo si $\lambda \in \{a_{11}, a_{22}, \dots, a_{nn}\}$.

Para una matriz triangular inferior, la demostración es similar. □

Definición 2.7. Si $A \in M_n(\mathbb{F})$ es una matriz cuadrada, el **polinomio característico** de A es el polinomio $\underline{p_A(t)} \in \mathbb{F}[t]$ definido por

$$\boxed{p_A(t) := \det(t\,1_n - A).} \tag{2.3}$$

Nótese que $p_A(t) = (-1)^n \det(A - t\,1_n)$ también. ◊

[4]Esta conclusión solo es válida porque $\dim V$ es finita, así que $n(T - \lambda\,1) = 0$ y $r(T - \lambda\,1) = \dim V$ son equivalentes. Sobre espacios vectoriales infinitodimensionales, existen operadores lineales inyectivos pero no sobreyectivs. Véase el Ejercicio 1.9, por ejemplo.

Por ejemplo, si $n = 4$, el polinomio característico de A viene dado por

$$p_A(t) = \begin{vmatrix} t - a_{11} & -a_{12} & -a_{13} & -a_{14} \\ -a_{21} & t - a_{22} & -a_{23} & -a_{24} \\ -a_{31} & -a_{32} & t - a_{33} & -a_{34} \\ -a_{41} & -a_{42} & -a_{43} & t - a_{44} \end{vmatrix} = \begin{vmatrix} a_{11} - t & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} - t & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} - t & a_{34} \\ a_{41} & a_{42} & a_{43} & a_{44} - t \end{vmatrix}.$$

Los procedimientos de cálculo de determinantes muestran que $p_A(t)$ es un polinomio de grado n. Por ejemplo, la fórmula de Leibniz (1.19) muestra que

$$p_A(t) = (t - a_{11})(t - a_{22}) \cdots (t - a_{nn}) + \text{otros términos},$$

donde cada uno de los "otros términos" es (± 1) veces un producto de n entradas de la matriz $t\,1_n - A$, de las cuales a lo sumo $(n - 2)$ entradas serían diagonales: la suma de todos ellos forma un polinomio de grado no mayor que $(n - 2)$. Entonces se ve que

$$p_A(t) = t^n - (a_{11} + a_{22} + \cdots + a_{nn})\, t^{n-1} + O(t^{n-2}).$$

El polinomio $p_A(t)$ es un **polinomio mónico**,[5] es decir, su primer coeficiente no nulo es 1. ⟦ La Proposición 2.16, más adelante, ofrece fórmulas explícitas para todos los coeficientes de $p_A(t)$. ⟧

Lema 2.8. *Si $A, B \in M_n(\mathbb{F})$ son dos matrices semejantes, entonces* $\det B = \det A$.

Demostración. Las matrices A y B son semejantes si y solo si hay una matriz invertible P tal que $B = P^{-1}AP$. Como $(\det P^{-1})(\det P) = \det(P^{-1}P) = \det 1_n = 1$, se ve que $\det P^{-1} = 1/(\det P)$. Entonces

$$\det B = \det(P^{-1}AP) = (\det P^{-1})(\det A)(\det P) = \det A. \qquad \square$$

Corolario 2.9. *Si $A, B \in M_n(\mathbb{F})$ son matrices semejantes, entonces $p_B(t) = p_A(t)$.* ⊟

Definición 2.10. Si V es un espacio vectorial de dimensión finita sobre $\mathbb{F}$ y si $T \in \mathcal{L}(V)$, su **determinante** $\underline{\det T} \in \mathbb{F}$ se define por $\det T := \det [T]^{\mathcal{B}}_{\mathcal{B}}$, donde $\mathcal{B}$ es una base cualquiera de V.

El escalar $\det T$ está bien definida, porque si $A = [T]^{\mathcal{B}}_{\mathcal{B}}$ y $B = [T]^{\mathcal{C}}_{\mathcal{C}}$ son las matrices de T con respecto a dos bases distintas $\mathcal{B}$, $\mathcal{C}$ de V, entonces la matriz de cambio de base $P = [1_V]^{\mathcal{B}}_{\mathcal{C}}$ es invertible, con inverso $P^{-1} = [1_V]^{\mathcal{C}}_{\mathcal{B}}$; por tanto,

$$B = [T]^{\mathcal{C}}_{\mathcal{C}} = [1_V]^{\mathcal{C}}_{\mathcal{B}}\, [T]^{\mathcal{B}}_{\mathcal{B}}\, [1_V]^{\mathcal{B}}_{\mathcal{C}} = P^{-1}AP \tag{2.4}$$

y del Lema 2.8 se concluye que $\det B = \det A$. ◊

[5] Algunos libros definen $p_A(t) := \det(A - t\,1_n)$. Con ese convenio, el primer coeficiente no nulo sería $(-1)^n$. No hay gran diferencia entre los dos convenios; sin embargo, es más cómodo elegir el signo de manera que el polinomio característico sea mónico.

Definición 2.11. Sea $T \in \mathcal{L}(V)$ un operador lineal sobre un espacio $\mathbb{F}$-vectorial V de dimensión finita. El **polinomio característico de** T es el polinomio $p_T(t) \in \mathbb{F}[t]$ definido por:

$$\boxed{p_T(t) := \det(t\,1_V - T)} = (-1)^{\dim V} \det(T - t\,1_V). \qquad \Diamond$$

Proposición 2.12. *Los autovalores de una matriz cuadrada $A \in M_n(\mathbb{F})$ son las raíces de su polinomio característico $p_A(t)$. Por lo tanto, A posee a lo sumo n autovalores distintos.*

Demostración. El Lema 2.5 dice que λ es un autovalor de A si y solo si $p_A(\lambda) = 0$. □

Ejemplo 2.13. Considérese la siguiente matriz $J \in M_2(\mathbb{F})$:

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{2.5}$$

Su polinomio característico es $\begin{vmatrix} t & -1 \\ 1 & t \end{vmatrix} = t^2 + 1$. Ahora, si $\mathbb{F} = \mathbb{R}$ o $\mathbb{Q}$, el polinomio $t^2 + 1$ es irreducible[6] y no posee raíces en $\mathbb{F}$. Este es un ejemplo de *una matriz que no posee autovalor alguno* en $\mathbb{F}$.

Por otro lado, si $\mathbb{F} = \mathbb{C}$, el polinomio $t^2 + 1$ sí es reducible, al ser $t^2 + 1 = (t - i)(t + i)$. Esto muestra que $\{i, -i\}$ podrían ser autovalores de J. Es fácil obtener un par de autovectores en $\mathbb{C}^2$, para verificar que i y $-i$ son efectivamente autovalores de J; por ejemplo:

$$\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ i \end{bmatrix} = \begin{bmatrix} i \\ -1 \end{bmatrix} = i \begin{bmatrix} 1 \\ i \end{bmatrix}, \qquad \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -i \end{bmatrix} = \begin{bmatrix} -i \\ -1 \end{bmatrix} = -i \begin{bmatrix} 1 \\ -i \end{bmatrix}. \qquad \Diamond$$

► En general, la búsqueda de autovectores exige encontrar soluciones no triviales de sistemas de ecuaciones lineales homogéneas.

Ejemplo 2.14. El polinomio característico de la matriz

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

se obtiene del cálculo

$$\begin{aligned} p_A(t) = \det(t\,1_3 - A) = \begin{vmatrix} t-1 & 0 & -2 \\ 0 & t+1 & 2 \\ -2 & 2 & t \end{vmatrix} &= (t-1) \begin{vmatrix} t+1 & 2 \\ 2 & t \end{vmatrix} - 2 \begin{vmatrix} 0 & t+1 \\ -2 & 2 \end{vmatrix} \\ &= (t-1)(t^2 + t - 4) - 2(2t + 2) = t^3 - 9t \\ &= t(t-3)(t+3). \end{aligned}$$

[6] En el caso $\mathbb{F} = \mathbb{F}_p := \{\overline{0}, \overline{1}, \dots, \overline{p-1}\}$, el cuerpo finito de residuos módulo división por un entero primo p, la existencia de raíces de $t^2 + 1$ en $\mathbb{F}_p$ es un tema interesante de la teoría de números. Se sabe que -1 es un cuadrado módulo p si y solo si $p = 2$ o bien $p = 4m + 1$ para algún $m \in \mathbb{N}$.

Luego $p_A(t) = t^3 - 3t$, con raíces $\lambda = 0, 3, -3$; estos son los tres autovalores de A.

Ahora bien: para obtener los *autovectores* correspondientes, se debe resolver (por eliminación gaussiana) tres sistemas de ecuaciones de ecuaciones homogéneas $(\lambda\,1_3 - A)\boldsymbol{x} = \mathbf{0}$ para $\lambda = 0, 3, -3$ respectivamente. En cada caso, se cambia la matriz aumentada $[\lambda\,1_3 - A \mid \mathbf{0}]$ a la forma $[V \mid \mathbf{0}]$ con V triangular superior mediante operaciones de fila y se resuelve la ecuación $V\boldsymbol{x} = \mathbf{0}$ por “sustitución regresiva”. En cada caso, la última fila de $[V \mid \mathbf{0}]$ es nula, que corresponde a la ecuación trivial $0x_3 = 0$, con lo cual la variable x_3 queda libre: el autovector queda determinado hasta un múltiplo. En detalle:

$$\text{Caso } \underline{\lambda = 0}: \qquad \left[\begin{array}{ccc|c} -1 & 0 & -2 & 0 \\ 0 & 1 & 2 & 0 \\ -2 & 2 & 0 & 0 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} -1 & 0 & -2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 2 & 4 & 0 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} -1 & 0 & -2 & 0 \\ 0 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right];$$

en cuyo caso (leyendo las filas de abajo para arriba),

$$0x_3 = 0; \quad x_2 + 2x_3 = 0; \quad -x_1 - 2x_3 = 0 \implies \boldsymbol{x} = \begin{bmatrix} -2x_3 \\ -2x_3 \\ x_3 \end{bmatrix} = x_3 \begin{bmatrix} -2 \\ -2 \\ 1 \end{bmatrix}.$$

Además,

$$\text{Caso } \underline{\lambda = 3}: \qquad \left[\begin{array}{ccc|c} 2 & 0 & -2 & 0 \\ 0 & 4 & 2 & 0 \\ -2 & 2 & 3 & 0 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} 2 & 0 & -2 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 2 & 1 & 0 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} 2 & 0 & -2 & 0 \\ 0 & 4 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right];$$

cuya solución es

$$0x_3 = 0; \quad 4x_2 + 2x_3 = 0; \quad 2x_1 - 2x_3 = 0 \implies \boldsymbol{x} = \begin{bmatrix} x_3 \\ -\frac{1}{2}x_3 \\ x_3 \end{bmatrix} = \tfrac{1}{2}x_3 \begin{bmatrix} 2 \\ -1 \\ 2 \end{bmatrix}.$$

Seguidamente,

$$\text{Caso } \underline{\lambda = -3}: \qquad \left[\begin{array}{ccc|c} -4 & 0 & -2 & 0 \\ 0 & -2 & 2 & 0 \\ -2 & 2 & -3 & 0 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} -4 & 0 & -2 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 2 & -2 & 0 \end{array}\right] \longmapsto \left[\begin{array}{ccc|c} -4 & 0 & -2 & 0 \\ 0 & -2 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{array}\right];$$

cuya solución es

$$0x_3 = 0; \quad -2x_2 + 2x_3 = 0; \quad -4x_1 - 2x_3 = 0 \implies \boldsymbol{x} = \begin{bmatrix} -\frac{1}{2}x_3 \\ x_3 \\ x_3 \end{bmatrix} = \tfrac{1}{2}x_3 \begin{bmatrix} -1 \\ 2 \\ 2 \end{bmatrix}.$$

Estos tres autovectores *forman las columnas de una matriz*

$$P := \begin{bmatrix} -2 & 2 & -1 \\ -2 & -1 & 2 \\ 1 & 2 & 2 \end{bmatrix}.$$

Esta matriz cuadrada P es invertible: es fácil calcular que $\underline{\det P = 27}$ y que

$$\operatorname{adj} P = \begin{bmatrix} -6 & -6 & 3 \\ 6 & -3 & 6 \\ -3 & 6 & 6 \end{bmatrix}; \qquad P^{-1} = \frac{1}{\det P} \operatorname{adj} P = \frac{1}{9} \begin{bmatrix} -2 & -2 & 1 \\ 2 & -1 & 2 \\ -1 & 2 & 2 \end{bmatrix}.$$

Se ve por cálculo directo que $\underline{AP = PD}$, donde D es una matriz diagonal. En efecto,

$$AP = \begin{bmatrix} 1 & 0 & 2 \\ 0 & -1 & -2 \\ 2 & -2 & 0 \end{bmatrix} \begin{bmatrix} -2 & 2 & -1 \\ -2 & -1 & 2 \\ 1 & 2 & 2 \end{bmatrix} = \begin{bmatrix} 0 & 6 & 3 \\ 0 & -3 & -6 \\ 0 & 6 & -6 \end{bmatrix} = \begin{bmatrix} -2 & 2 & -1 \\ -2 & -1 & 2 \\ 1 & 2 & 2 \end{bmatrix} \begin{bmatrix} 0 & 0 & 0 \\ 0 & 3 & 0 \\ 0 & 0 & -3 \end{bmatrix} = PD.$$

Las entradas diagonales de la matriz D son precisamente los tres autovalores $0, 3, -3$ de la matriz A, en el orden que corresponde al orden de las columnas de P. La ecuación $AP = PD$ también puede escribirse en la forma

$$P^{-1}AP = D. \tag{2.6}$$

En otras palabras, *la matriz A es semejante a una matriz diagonal D*, mediante conjugación $A \mapsto P^{-1}AP$ por una matriz invertible P *cuyas columnas son los autovectores de A*. Dícese que la matriz A es **diagonalizable**. Más adelante se estudiará unas condiciones necesarias para que una determinada matriz cuadrada sea diagonalizable. ◊

► Se busca la prometida *fórmula* general para los coeficientes del polinomio característico. Esta se expresa cómodamente con la siguiente notación para submatrices.

Notación. Una matriz $A \in \mathbb{F}^{m\times n}$ tiene varias submatrices A_{IJ} formados así: tómese dos juegos de índices

$$I := \{i_1, \dots, i_k\} \subseteq \{1, \dots, m\} \quad \text{y} \quad J := \{j_1, \dots, j_l\} \subseteq \{1, \dots, n\},$$

numerados en orden creciente: $i_1 < i_2 < \cdots < i_k$ y también $j_1 < j_2 < \cdots < j_l$. Entonces las entradas de la **submatriz** $\underline{A_{IJ}} \in \mathbb{F}^{k\times l}$ tiene las entradas a_{ij} con $i \in I$, $j \in J$.

Sean $\underline{I'} := \{1, \dots, m\} \setminus I$ y también $\underline{J'} := \{1, \dots, n\} \setminus J$. Dícese que la submatriz $A_{I'J'}$ es *complementaria* a A_{IJ}.

Definición 2.15. Si $A \in M_n(\mathbb{F})$ es una matriz cuadrada y si $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$, la submatriz diagonal $\underline{A_{II}} \in M_k(\mathbb{F})$ se llama una **submatriz principal** de A. Su determinante $\underline{m_{II}} := \det A_{II}$ es un **menor principal** de A. Para cada $k = 1, \dots, n$, hay $\binom{n}{k}$ menores principales de A obtenidos de submatrices $k \times k$ principales. ◊

Proposición 2.16. *Si $A \in M_N(\mathbb{F})$ es una matriz cuadrada, su polinomio característico tiene la forma*

$$p_A(t) = t^n - \tau_1(A)\, t^{n-1} + \tau_2(A)\, t^{n-2} - \cdots + (-1)^{n-1}\tau_{n-1}(A)\, t + (-1)^n \tau_n(A), \tag{2.7a}$$

donde

$$\boxed{\tau_1(A) = \underline{\operatorname{tr} A} := a_{11} + \cdots + a_{nn}} \quad \textit{es la } \textbf{\textit{traza}} \textit{ de } A; \quad \boxed{\tau_n(A) = \det A;}$$

y en general

$$\tau_k(A) = \sum_{|I|=k} \det A_{II} \quad \textit{para} \quad k = 1, \dots, n \tag{2.7b}$$

es la suma de todos los menores principales $k \times k$ de la matriz A.

Demostración. En el desarrollo de Leibniz del determinante

$$p_A(t) = \det(t\, 1_n - A) = \begin{vmatrix} t - a_{11} & -a_{12} & \dots & -a_{1n} \\ -a_{21} & t - a_{22} & \dots & -a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ -a_{n1} & -a_{n2} & \dots & t - a_{nn} \end{vmatrix},$$

el coeficiente de t^{n-k} es la suma de todos los términos obtenidos de la siguiente forma: elíjanse k índices $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$; tómese el término t del binomio $(t - a_{ll})$ para $l \notin I$; fórmese el producto de estos t con términos $(-a_{ij})$ de las filas I y las columnas I sin repetir filas ni columnas; multiplíquese por el signo de la permutación de filas contra columnas. De este modo, el coeficiente de t^{n-k} es

$$\sum_{|I|=k} \sum_{\sigma} (-1)^{\sigma} (-a_{i_1 j_1}) \cdots (-a_{i_k j_k}), \tag{2.8}$$

donde la suma recorre las permutaciones $\sigma \in S_n$ tales que $\sigma(i_r) = j_r$ para $r = 1, \dots, k$, mientras $\sigma(l) = l$ para $l \notin I$. Sea $\rho \in S_n$ tal que $\rho(r) = i_r \in I$ para $r = 1, \dots, k$ y $\rho(s) \in I'$ para $s = k + 1, \dots, n$. Sea π la permutación de $\{1, \dots, k\}$ determinado por $j_r =: i_{\pi(r)}$. ⟦ Es natural considerar π como un elemento de S_n que deja fija la parte $\{k + 1, \dots, n\}$. ⟧ Para $r \in \{1, \dots, k\}$, se ve que

$$r \overset{\rho}{\longmapsto} i_r \overset{\sigma}{\longmapsto} j_r = i_{\pi(r)} \overset{\rho^{-1}}{\longmapsto} \pi(r) \in \{1, \dots, k\},$$

lo cual dice que $\pi := \rho^{-1}\sigma\rho$, una permutación compuesta. ⟦ Como $\sigma(\rho(s)) = \rho(s) \in I'$ para $s > k$, se ve que $\rho^{-1}\sigma\rho$ también deja fijo $\{k+1,\dots,n\}$. ⟧ Esto implica que $(-1)^\pi = (-1)^\rho(-1)^\sigma(-1)^\rho = (-1)^\sigma$.

Ahora el coeficiente (2.8) de t^{n-k} tiene el formato:

$$\begin{aligned}\sum_{|I|=k}\sum_{\pi}(-1)^\pi(-a_{i_1 i_{\pi(1)}})\cdots(-a_{i_k i_{\pi(k)}}) &= (-1)^k\sum_{|I|=k}\sum_{\pi}(-1)^\pi a_{i_1 i_{\pi(1)}}\cdots a_{i_k i_{\pi(k)}} \\ &= (-1)^k\sum_{|I|=k}\det A_{II} = (-1)^k\tau_k(A).\end{aligned}$$

Por lo tanto, $p_A(t) = \sum_{k=0}^n (-1)^k\,\tau_k(A)\,t^{n-k}$. □

Corolario 2.17. *Las sumas de menores principales son invariantes bajo semejanza: toda vez que $A \in M_n(\mathbb{F})$ y $P \in M_n(\mathbb{F})$ es invertible, vale*

$$\tau_k(A) = \tau_k(P^{-1}AP) \quad \textit{para} \quad k = 1,\dots,n. \qquad ⊟$$

Si A es una matriz *triangular*, con autovalores $\lambda_1,\dots,\lambda_n$ en la diagonal, las submatrices principales A_{II} son también triangulares; en este caso, los menores principales $k \times k$ son productos de k elementos diagonales. Del Corolario 2.6, se ve que

$$\begin{aligned}\operatorname{tr} A &= \lambda_1 + \lambda_2 + \cdots + \lambda_n, \\ \tau_2(A) &= \lambda_1\lambda_2 + \lambda_1\lambda_3 + \cdots + \lambda_{n-1}\lambda_n, \\ \tau_3(A) &= \lambda_1\lambda_2\lambda_3 + \lambda_1\lambda_2\lambda_4 + \cdots + \lambda_{n-2}\lambda_{n-1}\lambda_n, \\ \vdots\quad & \quad\ \vdots \\ \det A &= \lambda_1\lambda_2\cdots\lambda_n. \end{aligned} \tag{2.9}$$

De hecho, estas fórmulas valen para cualquier matriz A cuyos autovalores son $\lambda_1,\dots,\lambda_n$, como se verá más adelante.

► El argumento de la demostración anterior es aplicable al cálculo de determinantes. Hay una generalización importante del desarrollo según una fila (o columna), que consiste en expandir en varias filas (o columnas) a la vez. La fórmula siguiente se conoce como el **desarrollo de Laplace** de un determinante.[7]

Escolio 2.18. *Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada y sea $I = \{i_1,\dots,i_k\}$ un juego de índices de filas de A. Si $J = \{j_1,\dots,j_k\}$ es un juego de índices de k columnas cualesquiera, sea*

[7] *Pierre-Simon de Laplace*, matemático y astrónomo francés, dio la regla de expansión en 1772, en uno de sus primeros trabajos sobre las órbitas planetarias, en el cual tuvo que resolver algunos sistemas de ecuaciones lineales.

$\underline{s(I,J)} := i_1 + \cdots + i_k + j_1 + \cdots + j_k$. *Entonces*

$$\det A = \sum_{|J|=k} (-1)^{s(I,J)} (\det A_{IJ})\,(\det A_{I'J'}), \tag{2.10}$$

donde la sumatoria recorre las $\binom{n}{k}$ *posibilidades para* J. ⊟

2.2. El teorema de Cayley y Hamilton

A esta altura, conviene recordar ciertos detalles acerca de la división de polinomios. Se sabe que los polinomios (con una sola incógnita t) sobre un cuerpo $\mathbb{F}$ forman una $\mathbb{F}$-álgebra conmutativa $\mathbb{F}[t]$. Esta álgebra no posee "divisores de cero": si $f(t) \neq 0$ y $g(t) \neq 0$, entonces $f(t)g(t) \neq 0$ también. (Dícese que $\mathbb{F}[t]$ es un álgebra *entera*.)[8] Esto es evidente al recordar la ley de producto:

$$\biggl(\sum_{j=0}^{n} a_j\, t^j\biggr)\biggl(\sum_{k=0}^{m} b_k\, t^k\biggr) = \sum_{j=0}^{n}\sum_{k=0}^{m} a_j b_k\, t^{j+k} = \sum_{r=0}^{n+m}\biggl(\sum_{j+k=r} a_j b_k\biggr) t^r,$$

porque $a_n \neq 0$, $b_m \neq 0$ implican $a_n b_m \neq 0$. Los *grados* se suman: si $\operatorname{gr} f(t) = n$, $\operatorname{gr} g(t) = m$, entonces $\operatorname{gr}(f(t)g(t)) = n + m$.

Un polinomio $g(t)$ es un **factor** de otro polinomio $f(t)$ si $f(t) = q(t)\,g(t)$ para algún polinomio $q(t)$. En este caso, se dice que $g(t)$ **divide** $f(t)$ y se escribe $\underline{g(t) \setminus f(t)}$. De lo contrario, si $g(t)$ no divide $f(t)$, se puede ejecutar una *división con residuo*, según la proposición que sigue.[9]

Proposición 2.19. *Si* $f(t), g(t) \in \mathbb{F}[t]$ *son dos polinomios con* $g(t) \neq 0$*, hay un único par de polinomios* $q(t)$*,* $r(t)$ *tales que*

$$\boxed{f(t) = q(t)g(t) + r(t),} \quad \text{con} \quad \begin{cases} \operatorname{gr} r(t) < \operatorname{gr} g(t), \\ \textit{o bien} \quad r(t) = 0. \end{cases} \tag{2.11}$$

Demostración. Escríbase $f(t) \equiv a_n t^n + \cdots + a_1 t + a_0$ y $g(t) \equiv b_m t^m + \cdots + b_1 t + b_0$. En el caso de que $m > n$, tómese $q(t) := 0$, $r(t) := f(t)$.

[8] Un anillo conmutativo A (estructura con suma y producto compatibles) es un **anillo entero** si para $a, b \in A$, la relación $ab = 0$ implica $a = 0$ o bien $b = 0$. Esta es una propiedad clave de los números enteros $\mathbb{Z}$. Algunos libros lo llaman *dominio entero* o, menos correctamente, *dominio de integridad*: Kronecker empleó este término para distinguirlo de un cuerpo, que él llamó *dominio de racionalidad*.

[9] El uso de la raya inclinada para denotar división se prefiere sobre la raya vertical $g(t) \mid f(t)$, por recomendación del libro: Ronald Graham, Donald Knuth y Oren Patashnik, *Concrete Mathematics*, Addison-Wesley, Reading, MA, 1989.

En cambio, si $m \leqslant n$, entonces

$$f_1(t) := f(t) - \frac{a_n}{b_m}\, t^{n-m}\, g(t)$$

es un polinomio con $\operatorname{gr} f_1(t) < n$. Al invocar inducción sobre n, se puede suponer que $f_1(t) = q_1(t)g(t) + r(t)$, con $\operatorname{gr} r(t) < m$ o bien $r(t) = 0$. Entonces

$$f(t) = \biggl(\frac{a_n}{b_m}\, t^{n-m} + q_1(t)\biggr) g(t) + r(t),$$

y la existencia de un par $q(t)$, $r(t)$ que cumple (2.11) sigue por la inducción sobre n.

Para la *unicidad* de $q(t)$ y $r(t)$, nótese que si $q(t)g(t) + r(t) = \tilde{q}(t)g(t) + \tilde{r}(t)$, entonces

$$\bigl(q(t) - \tilde{q}(t)\bigr) g(t) = \tilde{r}(t) - r(t).$$

Si esta ecuación no es $\underline{0 = 0}$, entonces su lado izquierdo tendría grado $\geqslant m$, mientras su lado derecho tendría $< m$, lo cual es imposible. Se deduce que $\tilde{r}(t) = r(t)$ y $\bigl(q(t) - \tilde{q}(t)\bigr) g(t) = 0$. Como $\mathbb{F}[t]$ es entero y $g(t) \neq 0$, se obtiene $q(t) - \tilde{q}(t) = 0$. □

Lema 2.20 (“Teorema del residuo”). *Si $a \in \mathbb{F}$, el residuo de la división de un polinomio $f(t) \in \mathbb{F}[t]$ por $(t - a)$ es igual a $f(a)$.*

Demostración. Escríbase $f(t) = (t-a)q(t)+r(t)$, según (2.11). Entonces $r(t)$ es un polinomio constante r_0, porque si no es nulo su grado es menor que $\operatorname{gr}(t-a) = 1$. Al evaluar esta ecuación polinomial en $a \in \mathbb{F}$, se obtiene $f(a) = (a - a)q(a) + r(a) = r(a) = r_0$. □

Corolario 2.21 (“Teorema del factor”). *Un polinomio $f(t)$ tiene $(t-a)$ como factor de primer grado si y solo si $f(a) = 0$.* ⊟

Definición 2.22. Si $f(t), g(t)$ son dos polinomios en $\mathbb{F}[t]$, su **máximo común divisor** $k(t) = \underline{\operatorname{mcd}(f(t), g(t))}$ es el (único) polinomio tal que

(i) $k(t) \setminus f(t), \quad k(t) \setminus g(t)$;

(ii) si $h(t) \setminus f(t)$ y $h(t) \setminus g(t)$, entonces $h(t) \setminus k(t)$;

(iii) $k(t)$ es *mónico*, es decir, de la forma $k(t) = t^m + c_{m-1}\, t^{m-1} + \cdots + c_1\, t + c_0$. ◊

Su existencia puede comprobarse con el *algoritmo euclidiano*, el mismo procedimiento que se usa para encontrar el máximo común divisor de dos números enteros en $\mathbb{Z}$. La Proposición 2.19 produce una sucesión finita de divisiones con residuo:

$$\begin{aligned} f(t) &= q_1(t)g(t) + r_1(t), & g(t) &= q_2(t)r_1(t) + r_2(t), \\ r_1(t) &= q_3(t)r_2(t) + r_3(t), & r_2(t) &= q_4(t)r_3(t) + r_4(t), \ \ldots \\ r_{j-2}(t) &= q_j(t)r_{j-1}(t) + r_j(t), & r_{j-1}(t) &= q_{j+1}(t)r_j(t) + 0, \end{aligned} \tag{2.12}$$

donde los grados de los residuos decrecen hasta que algún residuo se anule. Si es el último residuo no nulo es $r_j(t) = d_m t^m + \cdots + d_0$, no es difícil comprobar que $k(t) := d_m^{-1} r_j(t)$ cumple las tres propiedades de la Definición anterior.

Si $\tilde{k}(t)$ es cualquier máximo común divisor de $f(t)$ y $g(t)$, es obvio que $k(t) \setminus \tilde{k}(t)$ y $\tilde{k}(t) \setminus k(t)$, así que $\tilde{k}(t) = c\,k(t)$ para algún $c \in \mathbb{F}$. Como $\tilde{k}(t)$ también es mónico, se ve que $c = 1$. En resumen: $\operatorname{mcd}(f(t), g(t))$ existe y es único.

Lema 2.23. *Dados $f(t), g(t) \in \mathbb{F}[t]$, existen otros dos polinomios $a(t)$, $b(t)$ tales que*

$$\boxed{a(t)\,f(t) + b(t)\,g(t) = \operatorname{mcd}(f(t), g(t)).} \tag{2.13}$$

Demostración. Fíjese que $r_1(t) = f(t) - q_1(t)g(t)$, a partir de (2.12). Además,

$$\begin{aligned} r_2(t) &= g(t) - q_2(t)r_1(t) \\ &= g(t) - q_2(t)\bigl(f(t) - q_1(t)g(t)\bigr) \\ &= -q_2(t)f(t) + \bigl(q_1(t)q_2(t) + 1\bigr)\,g(t). \end{aligned}$$

Al repetir este argumento para $r_2(t)$, $r_3(t)$, etc., se puede hallar $a_i(t), b_i(t) \in \mathbb{F}[t]$ tales que $r_i(t) = a_i(t)f(t) + b_i(t)g(t)$, para $i = 1, 2, \dots, j$. Al dividir la última de estas ecuaciones, $a_j(t)f(t) + b_j(t)g(t) = r_j(t)$ por el primer coeficiente d_m de $r_j(t)$, se obtiene la fórmula (2.13). □

Corolario 2.24 (Identidad de Bézout)**.** *Dos polinomios $f(t)$, $g(t)$ son* ***relativamente primos****, esto es, $\underline{\operatorname{mcd}(f(t), g(t)) = 1}$, si y solo si hay polinomios $a(t), b(t) \in \mathbb{F}[t]$ tales que*

$$\boxed{a(t)\,f(t) + b(t)\,g(t) = 1.}$$ ⊟

► Un procedimiento útil, llamado *cálculo funcional*, consiste en reemplazar las potencias t^n del indeterminado t por las potencias de un elemento de alguna $\mathbb{F}$-álgebra. En particular, podemos sustituir t por una matriz en $M_n(\mathbb{F})$, o bien por una aplicación lineal en $\mathcal{L}(V)$.

Definición 2.25. Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada. Si $f(t) = c_n t^n + \cdots + c_1 t + c_0$ es un polinomio en $\mathbb{F}[t]$, se define

$$\underline{f(A)} := c_n A^n + \cdots + c_1 A + c_0\, 1_n \in M_n(\mathbb{F}). \tag{2.14a}$$

La aplicación $f(t) \mapsto f(A) : \mathbb{F}[t] \to M_n(\mathbb{F})$ es $\mathbb{F}$-lineal y lleva productos de polinomios en productos de matrices, es decir, es un *homomorfismo de álgebras*.

De igual manera, sea $T \in \mathcal{L}(V)$, donde V es un espacio $\mathbb{F}$-vectorial. Defínase

$$\underline{f(T)} := c_n T^n + \cdots + c_1 T + c_0\, 1_V \in \mathcal{L}(V). \tag{2.14b}$$

La aplicación $f(t) \mapsto f(T) : \mathbb{F}[t] \to \mathcal{L}(V)$ es $\mathbb{F}$-lineal y lleva productos de polinomios en composiciones de operadores: este es otro homomorfismo de álgebras. ◊

Si $A = [T]^{\mathcal{B}}_{\mathcal{B}}$ es la matriz de T respecto de una base $\mathcal{B}$ de V, entonces $p(A) = [p(T)]^{\mathcal{B}}_{\mathcal{B}}$.

Los homomorfismos de la Definición 2.25 no son sobreyectivos, porque las álgebras $M_n(\mathbb{F})$ y $\mathcal{L}(V)$ no son conmutativos (si $n > 1$ o $\dim V > 1$). Tampoco son inyectivos, porque $M_n(\mathbb{F})$ y $\mathcal{L}(V)$ son finitodimensionales y $\mathbb{F}[t]$ es tiene dimensión infinita. Entonces, para cada matriz A, debe haber polinomios no nulos $f(t)$ tales que $f(A) = 0$. El siguiente teorema, debido a Hamilton[10] y a Cayley,[11] proporciona un polinomio específico con esta propiedad, el cual es justamente el polinomio característico de A.

Teorema 2.26 (Cayley y Hamilton)**.** *Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada y sea $p_A(t) \in \mathbb{F}[t]$ su polinomio característico. Entonces $\underline{p_A(A) = 0}$ en $M_n(\mathbb{F})$.*

Demostración. La Proposición 1.52, antecedente de la regla de Cramer, demuestra que[12]

$$(t\,1_n - A)\ \operatorname{adj}(t\,1_n - A) = \det(t\,1_n - A)\,1_n = p_A(t)\,1_n\,. \tag{2.15}$$

Las entradas de la matriz $\operatorname{adj}(t\,1_n - A)$ son, salvo signo, menores $(n-1)\times(n-1)$ de la matriz $t\,1_n - A$. Como tal, son polinomios de grado no mayor que $(n-1)$. Al combinar términos según las potencias de t, se obtienen matrices $B_0, B_1, \dots, B_{n-1} \in M_n(\mathbb{F})$ tales que

$$\operatorname{adj}(t\,1_n - A) = B_{n-1}\,t^{n-1} + \cdots + B_1\,t + B_0.$$

Al escribir $p_A(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0$, la ecuación (2.15) tiene el formato:[13]

$$(t\,1_n - A)\,(B_{n-1}\,t^{n-1} + \cdots + B_1\,t + B_0) = (t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0)\,1_n.$$

Al igualar potencias de t en ambos lados de esta ecuación, resultan las siguientes igualdades:

$$-AB_0 = c_0 1_n, \quad B_0 - AB_1 = c_1 1_n,\ \dots,\quad B_{n-2} - AB_{n-1} = c_{n-1} 1_n, \quad B_{n-1} = 1_n\,.$$

[10] *William Rowan Hamilton* desarrolló la teoría de *cuaterniones*, que combinan escalares y 3-vectores reales en un espacio vectorial $\mathbb{H} := \mathbb{R} \oplus \mathbb{R}^3$, dotado de un producto no conmutativo. Las aplicaciones lineales en $\mathcal{L}_{\mathbb{R}}(\mathbb{H})$ se representan por matrices en $M_4(\mathbb{R})$. Hamilton mostró que cada aplicación satisface su polinomio característico, en su libro *Lectures on Quaternions*, Dublin, 1852.

[11] *Arthur Cayley* introdujo la definición moderna de *matriz* en su artículo: “Memoir on the theory of matrices”, *Philosophical Transactions of the Royal Society of London* **148** (1858), 17–37. Allí enunció el teorema para matrices cuadradas en general, aunque sólo demostró los casos 2×2 y 3×3.

[12] La regla de Cramer es válido para matrices con entradas escalares. Para justificar (2.15), se puede reemplazar t por λ ya que se verifica la ecuación correspondiente para todo $\lambda \in \mathbb{F}$. Mejor aun, se ve que la fórmula $B \operatorname{adj} B = (\det B)\,1_n$ de (1.22) es una abreviatura para n^2 identidades polinomiales en las entradas de $B \in M_n(\mathbb{F})$, que sigue válido (por cálculo funcional) cuando se reemplaza el cuerpo $\mathbb{F}$ por el álgebra $\mathbb{F}[t]$.

[13] Ya se sabe por (2.7) que $c_k = (-1)^{n-k}\tau_{n-k}(A)$, pero esta demostración no requiere la fórmula explícita.

Al multiplicarlas por potencias sucesivas de A, se obtiene un sistema de ecuaciones:

$$\begin{aligned} -AB_0 &= c_0 1_n, \\ AB_0 - A^2B_1 &= c_1 A, \\ A^2B_1 - A^3B_2 &= c_2 A^2, \\ \vdots \quad &= \ \vdots \\ A^{n-1}B_{n-2} - A^n B_{n-1} &= c_{n-1}A^{n-1}, \\ A^n B_{n-1} &= A^n. \end{aligned}$$

Una *suma telescópica* de estas relaciones produce el resultado deseado:

$$0 = A^n + c_{n-1}A^{n-1} + \cdots + c_1 A + c_0 \, 1_n = p_A(A). \qquad \square$$

Corolario 2.27. *Sea $T \in \mathcal{L}(V)$ un operador lineal sobre el espacio vectorial V, con polinomio característico $p_T(t) \in \mathbb{F}[t]$. Entonces $p_T(T) = 0$ en $\mathcal{L}(V)$.* ⊟

► Amén del polinomio característico $p_A(t)$, hay otros polinomios cuya evaluación en A es nula. (Cualquier polinomio divisible por el propio $p_A(t)$ tiene esa propiedad.) Interesa obtener un polinomio de mínimo grado en donde A se anula. A veces este polinomio es $p_A(t)$, pero otras veces tiene grado menor.

Proposición 2.28. *Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada. Entre todos los polinomios $f(t)$ en $\mathbb{F}[t]$ tales que $f(A) = 0$, hay un único polinomio* mónico $\underline{q_A(t)}$ de mínimo grado. *Este $q_A(t)$ divide cualquier $f(t)$ que cumple $f(A) = 0$.*

Demostración. Sea $f(t)$ cualquier polinomio que satisface $f(A) = 0$. Como n es el grado del polinomio característico $p_A(t)$, el menor grado posible[14] para $f(t)$ es un número $m \in \{1, 2, \dots, n\}$. Sea $q_A(t)$ un polinomio mónico de grado m tal que $q_A(A) = 0$ en $M_n(\mathbb{F})$.

Por la divisibilidad con residuo (2.11), se puede escribir

$$f(t) = s(t)\, q_A(t) + r(t),$$

para un único par de polinomios $s(t)$, $r(t)$, donde $\operatorname{gr} r(t) < \operatorname{gr} q_A(t) = m$ o bien $r(t) = 0$.

Además, $r(A) = f(A) - s(A)\, q_A(A) = 0$. Como esto contradiría la minimalidad de grado m si $r(t)$ fuera un polinomio no nulo, se concluye que $r(t) = 0$. Por lo tanto, $f(t) = s(t)\, q_A(t)$, de modo que $q_A(t) \setminus f(t)$.

Si $\tilde{q}(t)$ es un polinomio mónico cualquiera de grado m con $\tilde{q}(A) = 0$, el mismo argumento muestra que $\tilde{q}(t) = u(t)\, q_A(t)$ para algún polinomio $u(t)$. Por conteo de grados, se ve que $u(t)$ es un polinomio constante. Por ser $q_A(t)$ y $\tilde{q}(t)$ mónicos, se deduce que $u(t) = 1$; luego $\tilde{q}(t) = q_A(t)$. Esto establece la unicidad de $q_A(t)$. □

[14] Un polinomio de grado 0 es *una constante no nula* $g(t) = c_0 \neq 0$. Por la fórmula (2.14a), $g(A) = c_0 \, 1_n \neq 0$. Por convenio, el grado del polinomio nulo $g(t) = 0$ no se define. Entonces $f(t)$ debe tener grado al menos 1.

Corolario 2.29. *Sea $T \in \mathcal{L}(V)$ un operador lineal. Entre todos los polinomios mónicos $f(t) \in \mathbb{F}[t]$ tales que $f(T) = 0$, hay un único polinomio $\underline{q_T(t)}$ de mínimo grado. Este $q(t)$ divide cualquier $f(t)$ tal que $f(T) = 0$.* □

Definición 2.30. Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada. El polinomio mónico $\underline{q_A(t)}$ de mínimo grado tal que $q_A(A) = 0$ es el **polinomio mínimo** de A.

El teorema de Cayley y Hamilton muestra que $p_A(A) = 0$. Por lo tanto, $\operatorname{gr} q_A(t) \leqslant n$. La Proposición 2.28 muestra que $q_A(t) \setminus p_A(t)$. En particular, todas las raíces de $q_A(t)$ son autovalores de A (por la Proposición 2.12). (El resultado inverso también es válido: el Corolario 2.41, abajo, muestra que todo autovalor de A es una raíz de su polinomio mínimo.)

De manera similar, si $T \in \mathcal{L}(V)$, el polinomio mónico $q_T(t)$ de mínimo grado tal que $q_T(T) = 0$ se llama el **polinomio mínimo** de T. Además, $q_T(t) \setminus p_T(t)$. ◊

Ejemplo 2.31. Considérese la matriz triangular

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

Su polinomio característico $p_A(t)$ es entonces

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

Son candidatos *a priori* para el polinomio mínimo los factores: $(t-3)$; $(t-2)$; $(t-3)^2$; $(t-2)^2$; $(t-3)(t-2)$; $(t-3)^2(t-2)$; $(t-3)(t-2)^2$; y $(t-3)^2(t-2)^2$. Obsérvese que

$$(A - 3\,1_4)(A - 2\,1_4) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \neq 0$$

pero que $(A - 3\,1_4)^2(A - 2\,1_4) = 0$, por cálculo directo. Se concluye que $q_A(t) = (t-3)^2(t-2)$. ⟦ *Moraleja*: el polinomio mínimo puede tener factores repetidos. ⟧ ◊

2.3. Matrices diagonalizables

Entre todas las matrices cuadradas que representan un operador lineal T, se busca una que sea lo más sencilla posible. Hay varias posibilidades para $[T]^{\mathcal{C}}_{\mathcal{C}}$ porque hay varias maneras de elegir la base $\mathcal{C}$ del espacio vectorial subyacente. En algunos casos, pero no siempre, se puede elegir esta base de tal manera que la matriz $[T]^{\mathcal{C}}_{\mathcal{C}}$ sea diagonal.

La búsqueda del representante diagonal se reduce a un problema de clasificar las matrices cuadradas. Si $A = [T]^{\mathcal{B}}_{\mathcal{B}}$ es una matriz que representa $T \in \mathcal{L}(V)$ en una base dada $\mathcal{B}$ de V, se obtiene cualquier otro representante $[T]^{\mathcal{C}}_{\mathcal{C}}$ por un cambio de base de $\mathcal{B}$ a $\mathcal{C}$. Si $P = [1_V]^{\mathcal{B}}_{\mathcal{C}}$ es la matriz de cambio de base, entonces se pasa de $A = [T]^{\mathcal{B}}_{\mathcal{B}}$ a $P^{-1}AP = [T]^{\mathcal{C}}_{\mathcal{C}}$, según la fórmula (2.4). El procedimiento se reduce al siguiente problema matricial: *dada una matriz $A \in M_n(\mathbb{F})$, encontrar una matriz invertible P tal que $P^{-1}AP$ sea diagonal.*

Este problema admite una solucíónen el caso de que la matriz A posee *autovalores distintos*, en vista de los siguientes resultados.

Lema 2.32. *Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada. Si A posee $\underline{k}$* autovalores distintos *$\{\lambda_1, \dots, \lambda_k\}$ con unos autovectores correspondientes $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$, entonces esos autovectores son linealmente independientes.*

Demostración. Por inducción sobre k, se puede asumir que cualquier colección de $(k-1)$ autovectores para autovalores distintos son linealmente independientes. (Si $k = 1$, esto es evidente, porque $\{\boldsymbol{x}_1\}$ es linealmente independiente ya que $\boldsymbol{x}_1 \neq \boldsymbol{0}$, por ser $\boldsymbol{x}_1$ un autovector.)

Si $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ no fueran linealmente independientes, habría una relación de dependencia

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_k\boldsymbol{x}_k = \boldsymbol{0}, \tag{2.16}$$

con $c_1, \dots, c_k$ no todos cero. Renumerando la lista si fuera necesario, se puede suponer que $c_1 \neq 0$. Luego, $c_2, \dots, c_k$ no son todos cero porque $c_1\boldsymbol{x}_1 \neq \boldsymbol{0}$. Al aplicar la matriz A a los dos lados de esta ecuación, resulta

$$c_1\lambda_1\boldsymbol{x}_1 + c_2\lambda_2\boldsymbol{x}_2 + \cdots + c_k\lambda_k\boldsymbol{x}_k = \boldsymbol{0}.$$

Al restar λ_1 veces (2.16) de esta relación, se obtiene

$$c_2(\lambda_2 - \lambda_1)\boldsymbol{x}_2 + c_3(\lambda_3 - \lambda_1)\boldsymbol{x}_3 + \cdots + c_k(\lambda_k - \lambda_1)\boldsymbol{x}_k = \boldsymbol{0}. \tag{2.17}$$

Los coeficientes en la ecuación (2.17) no son todos cero porque los λ_j son distintos y $c_2, \dots, c_k$ no son todos cero. Pero entonces $\boldsymbol{x}_2, \dots, \boldsymbol{x}_k$ serían linealmente dependientes, contrario a la hipótesis inductiva. Se concluye que $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ deben ser linealmente independientes. □

Proposición 2.33. *Si $A \in M_n(\mathbb{F})$ posee $\underline{n}$ autovalores distintos, entonces A es diagonalizable.*

Demostración. Sean $\lambda_1, \dots, \lambda_n$ los n autovalores distintos de A y sea $\{\boldsymbol{p}_1, \dots, \boldsymbol{p}_n\}$ un juego de n autovectores correspondientes. Por el Lema 2.32 anterior, $\mathcal{B} = \{\boldsymbol{p}_1, \dots, \boldsymbol{p}_n\}$ es una base de $\mathbb{F}^n$. Cada $\boldsymbol{p}_s$ es una combinación lineal de los vectores de la base estándar $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$:

$$\boldsymbol{p}_s = \sum_{j=1}^{n} p_{js}\boldsymbol{e}_j = (\text{la columna número } s \text{ de una matriz } P).$$

Aquí $P = [p_{js}]$ es la matriz $[1]^{\mathcal{E}}_{\mathcal{B}}$ de cambio de base (de $\mathcal{E}$ a $\mathcal{B}$). El producto de matrices AP es entonces

$$\begin{aligned} AP &= A\,[\boldsymbol{p}_1\ \boldsymbol{p}_2\ \dots \boldsymbol{p}_n] = [A\boldsymbol{p}_1\ A\boldsymbol{p}_2\ \dots A\boldsymbol{p}_n] \\ &= [\lambda_1\boldsymbol{p}_1\ \lambda_2\boldsymbol{p}_2\ \dots \lambda_n\boldsymbol{p}_n] = [\boldsymbol{p}_1\ \boldsymbol{p}_2\ \dots \boldsymbol{p}_n] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \equiv PD, \end{aligned} \tag{2.18}$$

donde D es la matriz diagonal con entradas diagonales $\lambda_1, \dots, \lambda_n$. (Es útil recordar que la multiplicación a la derecha $P \mapsto PD$ efectúa un juego de operaciones de columna.)

La matriz P es invertible porque su rango es n, ya que tiene n columnas linealmente independientes. Entonces $\underline{AP = PD}$ es equivalente a $\underline{P^{-1}AP = D}$, con D diagonal. ◻

Notación. Se denota la matriz diagonal con entradas diagonales $\lambda_1, \dots, \lambda_n$ (en ese orden) con la notación compacta

$$\underline{\operatorname{diag}[\lambda_1, \lambda_2, \dots, \lambda_n]} := \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{2.19}$$

Si $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V tal que $[T]^{\mathcal{B}}_{\mathcal{B}} = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$, entonces $T(\boldsymbol{x}_j) = \sum_{i=1}^n \lambda_i [\![i = j]\!]\, \boldsymbol{x}_i = \lambda_j \boldsymbol{x}_j$ en vista de (1.9). En otras palabras, cada $\boldsymbol{x}_j$ es un autovector de T.

Proposición 2.34. *Una matriz $A \in M_n(\mathbb{F})$ es diagonalizable si y solo si A tiene n autovectores linealmente independientes.*

Demostración. Si $\boldsymbol{p}_1, \dots, \boldsymbol{p}_n \in \mathbb{F}^n$ son n autovectores de A que son linealmente independientes, entonces la matriz $P = [\boldsymbol{p}_1\ \boldsymbol{p}_2\ \dots \boldsymbol{p}_n]$ tiene n columnas linealmente independientes, con lo cual su rango es n y la matriz P es *invertible*. Fíjese que $A\boldsymbol{p}_k = \lambda_k\, \boldsymbol{p}_k$ para $k = 1, \dots, n$, donde λ_k es el autovalor que corresponde al autovector $\boldsymbol{p}_k$. Si $D := \operatorname{diag}[\lambda_1, \dots, \lambda_n]$ es la matriz diagonal cuyas entradas diagonales son estos autovalores, tomados en orden, entonces el cálculo (2.18) muestra de nuevo que $AP = PD$. Se concluye que $P^{-1}AP = D$; esto dice que A es diagonalizable con forma diagonal D.

Por otro lado, si A es diagonalizable, hay una matriz invertible P y una matriz diagonal $D := \operatorname{diag}[\lambda_1, \dots, \lambda_n]$ tal que $P^{-1}AP = D$. Por ende, vale $AP = PD$; luego, al comparar la k-ésima columna de ambos lados de esta igualdad matricial, se ve de (2.18) que $A\boldsymbol{p}_k = \lambda_k\, \boldsymbol{p}_k$ para $k = 1, \dots, n$. En consecuencia, cada λ_k es un autovalor de A y cada columna $\boldsymbol{p}_k$ es un autovector. La matriz invertible P tiene rango n, así que sus columnas son linealmente independientes. ◻

Corolario 2.35. *Una matriz $A \in M_n(\mathbb{F})$ es diagonalizable si y solo si $\mathbb{F}^n$ posee una base que consiste de autovectores de A.* ⊟

La Proposición 2.34 no exige que los autovalores de una matriz diagonalizable sean distintos. De hecho, cualquier matriz diagonal D es *ipso facto* diagonalizable: sus autovalores son sus entradas diagonales y su base de autovectores es la base estándar $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$.

Denótese por $\{\alpha_1, \dots, \alpha_r\}$ los *elementos distintos* del juego de autovalores $(\lambda_1, \dots, \lambda_n)$. Si α_1 ocurre k_1 veces, α_2 ocurre k_2 veces,..., α_r ocurre k_r veces, con $k_1 + k_2 + \cdots + k_r = n$, se puede permutar los λ_i para obtener

$$(\lambda_1, \dots, \lambda_n) \longmapsto (\underbrace{\alpha_1, \dots, \alpha_1}_{k_1 \text{ veces}}, \underbrace{\alpha_2, \dots, \alpha_2}_{k_2 \text{ veces}}, \dots, \underbrace{\alpha_r, \dots, \alpha_r}_{k_r \text{ veces}}).$$

Se dice que k_i es la **multiplicidad** del autovalor α_i. El polinomio característico de la matriz $D = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$ es entonces

$$\boxed{p_D(t) = (t - \alpha_1)^{k_1} (t - \alpha_2)^{k_2} \cdots (t - \alpha_r)^{k_r}.}$$

En el caso diagonal, el teorema de Cayley y Hamilton tiene una prueba directa: en el producto de matrices diagonales $(D - \alpha_1)^{k_1} (D - \alpha_2)^{k_2} \cdots (D - \alpha_r)^{k_r}$ al menos uno de los factores tiene una entrada diagonal 0 en cada fila; por lo tanto, el producto es la matriz nula.

El **polinomio mínimo** de esta matriz D es

$$\boxed{q_D(t) = (t - \alpha_1)\,(t - \alpha_2) \cdots (t - \alpha_r),} \tag{2.20}$$

con r factores distintos de primer grado. El producto $(D - \alpha_1)\,(D - \alpha_2) \cdots (D - \alpha_r)$ es la matriz 0 porque cada entrada diagonal es un producto de r escalares que incluyen un cero. Si se suprimiera uno de los factores $(t - \alpha_i)$, el producto de matrices con $(D - \alpha_i)$ omitido tendría al menos una entrada diagonal no nula.

Resulta que el resultado inverso también es válido: si el polinomio mínimo de una matriz A se descompone en factores lineales distintos, entonces A es diagonalizable. Antes de comprobar eso, conviene examinar otros aspectos estructurales de los operadores lineales.

2.4. Descomposición primaria de un operador lineal

En esta sección se verá que una matriz no diagonalizable es similar, por un cambio de base apropiado, a una suma directa de *bloques* centrados en la diagonal. Para apreciar mejor esa estructura, es preferible examinar la aplicación lineal subyacente con un tratamiento "libre de coordenadas".

Definición 2.36. Sea V un espacio vectorial sobre $\mathbb{F}$ y sea $T \in \mathcal{L}(V)$ un operador lineal. se dice que un subespacio $U \leqslant V$ es un **subespacio invariante** para T si $\underline{T(U) \subseteq U}$. ◊

Si $\dim V = n$ y si U es un subespacio invariante para $T \in \mathcal{L}(V)$ con $\dim U = m \leqslant n$, sea $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base de V cuya porción inicial $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es una base de U. (Es cuestión de elegir una base de U y luego completarla en una base de V, según el Escolio 1.9.) En el desarrollo (2.2) del operador T en esta base, la condición $T(U) \subseteq U$ implica que $a_{ij} = 0$ cuando $i > m$ y $j \leqslant m$; en otras palabras, la matriz $[T]_{\mathcal{B}}^{\mathcal{B}}$ tiene el formato

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} A & X \\ 0 & B \end{bmatrix}, \tag{2.21}$$

donde $A \in M_m(\mathbb{F})$, $B \in M_{n-m}(\mathbb{F})$, $X \in \mathbb{F}^{m\times(n-m)}$; y $0 \in \mathbb{F}^{(n-m)\times m}$ es un bloque de ceros.

Definición 2.37. Dícese que un subespacio invariante $U \leqslant V$ **reduce** el operador lineal $T \in \mathcal{L}(V)$ si hay otro subespacio invariante $W \leqslant V$ tal que $\underline{V = U \oplus W}$. En tal caso, existe una base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V tal que $U = \operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_m\rangle$ y $W = \operatorname{lin}\langle \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n\rangle$, y con respecto a esta base la esquina X de la matriz $[T]_{\mathcal{B}}^{\mathcal{B}}$ en (2.21) es otro bloque de ceros:

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}. \tag{2.22}$$

La matriz al lado derecho de (2.22) se llama la **suma directa** de las matrices cuadradas $A \in M_m(\mathbb{F})$ y $B \in M_{n-m}(\mathbb{F})$. ◊

Proposición 2.38. *Si el polinomio característico de un operador lineal $T \in \mathcal{L}(V)$ se factoriza en $p_T(t) = h(t)\,k(t)$ con $\operatorname{mcd}(h(t), k(t)) = 1$, entonces $\underline{V = U \oplus W}$, donde $U = \ker h(T)$ y $W = \ker k(T)$ son subespacios invariantes para T.*

Demostración. Por el Corolario 2.24, la condición $\operatorname{mcd}(h(t), k(t)) = 1$ implica que existen dos polinomios $a(t), b(t) \in \mathbb{F}[t]$ que cumplen la identidad de Bézout:

$$h(t)\,a(t) + k(t)\,b(t) = 1.$$

Luego $h(T)\,a(T) + k(T)\,b(T) = 1_V$ en $\mathcal{L}(V)$. Ahora defínase

$$\underline{U} := \operatorname{im} k(T) \quad \text{y} \quad \underline{W} := \operatorname{im} h(T).$$

Ellos son subespacios invariantes para T, porque $T(h(T)\boldsymbol{y}) = h(T)(T\boldsymbol{y})$ y de igual modo $T(k(T)\boldsymbol{x}) = k(T)(T\boldsymbol{x})$. Además, cada $\boldsymbol{x} \in V$ cumple

$$\boldsymbol{x} = k(T)(b(T)\boldsymbol{x}) + h(T)(a(T)\boldsymbol{x}) \in U + W,$$

así que $V = U + W$ como suma simple de subespacios vectoriales. Para ver que esta suma es *directa*, tómese $\boldsymbol{x} \in U \cap W$. Entonces existen $\boldsymbol{y}, \boldsymbol{z} \in V$ tales que $\boldsymbol{x} = k(T)\boldsymbol{y} = h(T)\boldsymbol{z}$.

Del teorema de Cayley y Hamilton sigue $h(T)\boldsymbol{x} = p_T(T)\boldsymbol{y} = \boldsymbol{0}$ y $k(T)\boldsymbol{x} = p_T(T)\boldsymbol{z} = \boldsymbol{0}$. La identidad de Bézout entonces muestra que

$$\boldsymbol{x} = a(T)(h(T)\boldsymbol{x}) + b(T)(k(T)\boldsymbol{x}) = a(T)(\boldsymbol{0}) + b(T)(\boldsymbol{0}) = \boldsymbol{0}.$$

Se concluye que $U \cap W = \{\boldsymbol{0}\}$ y en consecuencia $V = U \oplus W$.

Si $\boldsymbol{x} \in U$, existe $\boldsymbol{y} \in V$ tal que $\boldsymbol{x} = k(T)\boldsymbol{y}$. Entonces $h(T)\boldsymbol{x} = h(T)(k(T)\boldsymbol{y}) = p_T(T)\boldsymbol{y} = \boldsymbol{0}$. Esto muestra que $U \subseteq \ker h(T)$. Al contar dimensiones, el Teorema 1.23 (de rango y nulidad) implica que

$$\dim U = n - \dim W = n - r(h(T)) = n(h(T)) = \dim(\ker h(T)),$$

por tanto $U = \ker h(T)$. De igual modo, se ve que $W = \ker k(T)$. □

Proposición 2.39. *Sea $T \in \mathcal{L}(V)$ un operador lineal cuyo polinomio característico escinde*[15] *en $\mathbb{F}[t]$. Supóngase que $p_T(t) = h(t)\,k(t)$ donde $h(t)$ y $k(t)$ son polinomios mónicos con* $\operatorname{mcd}(h(t), k(t)) = 1$. *Entonces las restricciones de T a los subespacios $U = \ker h(T)$ y $W = \ker k(T)$ tienen polinomios característicos $h(t)$ y $k(t)$, respectivamente.*

Demostración. Obsérvese, por la demostración de la Proposición 2.38, que los subespacios U y W reducen T, esto es: $T(U) \subseteq U$ y $T(W) \subseteq W$. Sean $T' \in \mathcal{L}(U)$ y $T'' \in \mathcal{L}(W)$ las restricciones respectivas de T a U y W. Sea $\mathcal{B}$ una base de U y $\mathcal{C}$ una base de W, y por ende $\mathcal{B} \uplus \mathcal{C}$ (unión disjunta) es una base de $U \oplus W = V$. Si $A := [T']^{\mathcal{B}}_{\mathcal{B}}$ y $B := [T'']^{\mathcal{C}}_{\mathcal{C}}$, la matriz de T para la base $\mathcal{B} \uplus \mathcal{C}$ es

$$[T]^{\mathcal{B}\uplus\mathcal{C}}_{\mathcal{B}\uplus\mathcal{C}} = \begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}. \tag{2.23}$$

Entonces, si $r = \dim U$, $s = \dim W$, el polinomio característico de T se factoriza así:

$$p_T(t) = \det \begin{bmatrix} t\,1_r - A & 0 \\ 0 & t\,1_s - B \end{bmatrix} = \det(t\,1_r - A)\det(t\,1_s - B) = p_A(t)\,p_B(t).$$

Si λ es una raíz de $p_A(t)$, entonces hay $\boldsymbol{y} \in U \setminus \{\boldsymbol{0}\}$ tal que $T(\boldsymbol{y}) = T'(\boldsymbol{y}) = \lambda\boldsymbol{y}$. Luego $T^2(\boldsymbol{y}) = \lambda^2\boldsymbol{y}$, $T^3(\boldsymbol{y}) = \lambda^3\boldsymbol{y}$, etc., de modo que $\boldsymbol{0} = h(T)(\boldsymbol{y}) = h(\lambda)\boldsymbol{y}$ y por ende $h(\lambda) = 0$. Además, $k(\lambda) \neq 0$ porque $h(t)$ y $k(t)$ no tienen una raíz común.

[15] Se dice que un polinomio $f(t) \in \mathbb{F}[t]$ **escinde** si $f(t) = a_n(t - \alpha_1)(t - \alpha_2)\cdots(t - \alpha_n)$, con *raíces* $\alpha_1, \dots, \alpha_n \in \mathbb{F}$, no necesariamente distintos. En el caso $\mathbb{F} = \mathbb{C}$, *todo polinomio en* $\mathbb{C}[t]$ *escinde*: esto es el llamado Teorema Fundamental del Álgebra.

Entonces, cada raíz de $p_A(t)$ es una raíz de $h(t)$ pero no de $k(t)$. De igual manera, cada raíz de $p_B(t)$ es una raíz de $k(t)$ pero no de $h(t)$. Por lo tanto,

$$p_T(t) = k(t)\,h(t) = p_A(t)\,p_B(t) =: (t - \lambda_1) \cdots (t - \lambda_n)$$

donde la repartición de los monomios $(t - \lambda_i)$ entre los dos factorizaciones de $p_T(t)$ obliga las igualdades de polinomios *mónicos* $p_A(t) = h(t)$ y $p_B(t) = k(t)$. $\square$

En la Proposición 2.39, la hipótesis de que $p_T(t)$ escinde en $\mathbb{F}[t]$ (usada en el último paso de la demostración) no es indispensable. Es posible apelar a un teorema de Kronecker, que dice que cada polinomio $p(t) \in \mathbb{F}[t]$ tiene una raíz en algún cuerpo que *extiende* $\mathbb{F}$ (es decir, que que incluye $\mathbb{F}$ como subcuerpo).

Es posible, entonces, extender el cuerpo original $\mathbb{F}$ a otro cuerpo $\mathbb{K}$ tal que $p_T(t)$ escinde en $\mathbb{K}[t]$; las igualdades $p_A(t) = h(t)$ y $p_B(t) = k(t)$ entonces se verifican en $\mathbb{K}[t]$ y de rebote también en $\mathbb{F}[t]$. Este artificio es particularmente útil en el caso en donde $\mathbb{F} = \mathbb{R}$, porque hay polinomios reales cuadráticas que no escinden en $\mathbb{R}[t]$ pero sí en $\mathbb{C}[t]$ (en vista del Teorema Fundamental del Álgebra).

► En seguida, se verá que la factorización anterior del polinomio caraterístico (si existe) conlleva una factorización correspondiente del polinomio mínimo.

Lema 2.40. *Sea $V = U \oplus W$, con $U = \ker h(T)$ y $W = \ker k(T)$, la descomposición de V obtenida de una factorización del polinomio característico $p_T(t) = h(t)\,k(t)$ en factores relativamente primos. Entonces hay una factorización correspondiente del polinomio mínimo $q_T(t) = r(t)\,s(t)$ en factores relativamente primos, donde $r(t) \setminus h(t)$; $s(t) \setminus k(t)$; $r(T)$ se anula en U y $s(T)$ se anula en W.*

Demostración. Elíjase una base de $V = U \oplus W$ tal que T tenga una matriz en bloques de la forma (2.23). Entonces la relación $q_T(T) = 0$ conlleva la relación

$$\begin{bmatrix} q_T(A) & 0 \\ 0 & q_T(B) \end{bmatrix} = q_T\left(\begin{bmatrix} A & 0 \\ 0 & B \end{bmatrix}\right) = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix},$$

así que $q_T(A) = 0$ en $\mathcal{L}(U)$ y $q_T(B) = 0$ en $\mathcal{L}(W)$. Defínase

$$\underline{r(t)} := \operatorname{mcd}(q_T(t), h(t)) \quad \text{y} \quad \underline{s(t)} := \operatorname{mcd}(q_T(t), k(t)).$$

Entonces $r(t)$ y $s(t)$ no tienen factor común; y además

$$r(t)s(t) = \operatorname{mcd}((q_T(t), h(t)k(t)) = q_T(t),$$

porque $q_T(t)$ divide $h(t)k(t) = p_T(t)$.

Como $h(A) = p_A(A) = 0$ por la Proposición 2.39 y el Lema 2.23 dice que $r(t) = a(t)q_T(t) + b(t)h(t)$ para algunos polinomios $a(t)$ y $b(t)$, se deduce que $r(A) = 0$.

Entonces, si T' es la restricción de T a U, sigue que $r(T') = 0$ en $\mathcal{L}(U)$. Del mismo modo, si T'' es la restricción de T a W, se obtiene $s(B) = 0$ y $s(T'') = 0$ en $\mathcal{L}(W)$. □

Corolario 2.41. *Cada raíz del polinomio característico de un operador lineal T es también una raíz de su polinomio mínimo.*

Demostración. Si λ es un autovalor de T, entonces $p_T(t) = (t-\lambda)^m\, k(t)$ para algún $m \geqslant 1$, con $k(\lambda) \neq 0$. El Lema 2.40 anterior muestra que $q_T(t) = (t-\lambda)^l\, s(t)$, con $l \leqslant m$, donde $(T - \lambda\, 1)^l$ anula $\ker((T-\lambda\, 1)^m)$, lo cual sería imposible si fuera $l = 0$. ⟦ El cálculo funcional de la Definición 2.25 dice que $\underline{T^0 := 1}$ cuando T es un operador no nulo. ⟧ Luego $l \in \{1, \dots, m\}$ y por ende λ es una raíz de $q_T(t)$. □

Proposición 2.42. *Si el polinomio* mínimo *de un operador lineal $T \in \mathcal{L}(V)$ se factoriza en $q_T(t) = r(t)\, s(t)$ con $r(t)$, $s(t)$ mónicos y relativamente primos, entonces $V = U' \oplus W'$, donde $U' := \ker r(T)$ y $W' := \ker s(T)$ son subespacios invariantes para T. Los polinomios mínimos de las restricciones de T a U' y W' son $r(t)$ y $s(t)$, respectivamente.*

Demostración. La demostración de la Proposición 2.38 se repite en forma casi idéntica, con $q_T(t)$, $r(t)$, $s(t)$, U', W' en los lugares respectivos de $p_T(t)$, $h(t)$, $k(t)$, U y W. En vez de usar $p_T(T) = 0$ por el teorema de Cayley y Hamilton, se usa $q_T(T) = 0$ por la definición del polinomio mínimo. Se obtiene $V = U' \oplus W'$ con $U' := \ker r(T)$ y $W' := \ker s(T)$. Además, se establecen las igualdades $U' = \operatorname{im} s(T)$ y $W' = \operatorname{im} r(T)$.

Sea $T' \in \mathcal{L}(U')$ la restricción del operador T al subespacio invariante U'. Se ve que $r(T') = 0$ en $\mathcal{L}(U')$ porque $r(T')\boldsymbol{y} = r(T)\boldsymbol{y} = \mathbf{0}$ para todo $\boldsymbol{y} \in U'$, porque $U' = \ker r(T)$.

Si $f(t) \in \mathbb{F}[t]$ es un polinomio tal que $f(T') = 0$ en $\mathcal{L}(U')$, entonces $f(T)\boldsymbol{y} = \mathbf{0}$ para todo $\boldsymbol{y} = s(T)\boldsymbol{x} \in U'$. Por tanto, $f(T)(s(T)\boldsymbol{x}) = \mathbf{0}$ para todo $\boldsymbol{x} \in V$; esto es, $f(T)\, s(T) = 0$ en $\mathcal{L}(V)$. Luego $r(t)s(t) = q_T(t) \setminus f(t)s(t)$; en otras palabras, hay un polinomio $g(t)$ tal que $f(t)s(t) = g(t)r(t)s(t)$. De ahí sigue[16] la relación $f(t) = g(t)r(t)$. En resumen: si $f(T') = 0$, entonces $r(t) \setminus f(t)$. Se ha mostrado que el polinomio mónico $r(t)$ es el polinomio mínimo de $T' \in \mathcal{L}(U')$.

De igual modo, $s(t)$ es el polinomio mínimo de la restricción de T al subespacio W'. □

Ahora, se puede abordar el caso general de la factorización completa del polinomio mínimo de un operador en factores relativamente primos. Sin perder generalidad, se puede suponer que todos esos factores son polinomios mónicos.

[16] La cancelación del factor común $s(t)$ es válida porque el álgebra $\mathbb{F}[t]$ no posee "divisores de cero".

Teorema 2.43 (Descomposición primaria)**.** *Si el polinomio mínimo de un operador lineal $T \in \mathcal{L}(V)$ se factoriza como $q_T(t) = h_1(t) \cdots h_r(t)$ en factores mónicos relativamente primos $h_1(t), \dots, h_r(t)$, entonces $\underline{V = U_1 \oplus \cdots \oplus U_r}$, donde $\underline{U_i = \ker h_i(T)}$, $i = 1, \dots, r$. Además, cada $h_i(t)$ es el polinomio mínimo de la restricción de T al subespacio invariante U_i.*

Demostración. Por inducción sobre r. El caso $r = 1$ es trivial (nada que mostrar), y el caso $r = 2$ es la Proposición 2.42 anterior.

Sea $k_1(t) := h_2(t) \cdots h_r(t)$, así que $q_T(t) = h_1(t)\, k_1(t)$ con $\operatorname{mcd}(h_1(t), k_1(t)) = 1$. Ahora la Proposición 2.42 muestra que $V = U_1 \oplus W_1$, donde $U_1 := \ker h_1(T)$ y $W_1 := \ker k_1(T) = \operatorname{im} h_1(T)$. También muestra que las restricciones de T a U_1 y W_1 tienen polinomios mínimos respectivos $h_1(t)$ y $k_1(t)$.

Por inducción sobre r, se puede suponer que el resultado es válido para la restricción de T al subespacio W_1, con polinomio mínimo $k_1(t) = h_2(t) \cdots h_r(t)$. Se obtiene $W_1 = U_2 \oplus \cdots \oplus U_r$, donde $U_i = \ker h_i(T)$, con polinomio mínimo $h_i(T)$ en cada subespacio U_i, para $i = 2, \dots, r$. El resultado ahora es evidente. □

Corolario 2.44. *Sea $T \in \mathcal{L}(V)$ un operador lineal cuyo polinomio característico $p_T(t)$ escinde en $\mathbb{F}[t]$. Su factorización completa tiene la forma*

$$\boxed{p_T(t) = (t - \alpha_1)^{k_1} (t - \alpha_2)^{k_2} \cdots (t - \alpha_r)^{k_r}} \tag{2.24}$$

con $\alpha_1, \dots, \alpha_r \in \mathbb{F}$ distintos. Entonces su polinomio mínimo tiene la forma

$$\boxed{q_T(t) = (t - \alpha_1)^{l_1} (t - \alpha_2)^{l_2} \cdots (t - \alpha_r)^{l_r}} \tag{2.25}$$

con $l_1, \dots, l_r \in \mathbb{N}$ y $1 \leqslant l_i \leqslant k_i$ para $i = 1, \dots, r$. Sea $V = U_1 \oplus \cdots \oplus U_r$ la descomposición primaria correspondiente, donde $U_i := \ker\big((T - \alpha_i\, 1)^{l_i}\big)$. Entonces $(t - \alpha_i)^{k_i}$ es el polinomio característico de la restricción de T al subespacio invariante U_i.

Demostración. Por inducción sobre r; el resultado es obvio si $r = 1$. Sea $h(t) := (t - \alpha_1)^{k_1}$, $k(t) := (t - \alpha_2)^{k_2} \cdots (t - \alpha_r)^{k_r}$, de modo que $V = U \oplus W$ con $U = \ker h(T)$ y $W = \ker k(T)$, por la Proposición 2.38. El Lema 2.40 muestra que $q_T(t) = r(t)s(t)$, donde $r(t) = (t - \alpha_1)^{l_1}$ y $s(t) = (t - \alpha_2)^{l_2} \cdots (t - \alpha_r)^{l_r}$ con $l_i \leqslant k_i$ para cada i; además, $r(T)$ anula U y $s(T)$ anula W. El Corolario 2.41 muestra que $l_1 \geqslant 1$.

Como $r(t)$ divide $h(t)$, es inmediato que $U = \ker h(T) \subseteq \ker r(T) = U_1$. Por otro lado, $r(t)$ y $k(t)$ son relativamente primos, así que $a(t)r(t) + b(t)k(t) = 1$ para ciertos polinomios $a(t), b(t)$; luego, si $\boldsymbol{z} \in U_1 \cap W$, entonces

$$\boldsymbol{z} = a(T)(r(T)\boldsymbol{z}) + b(T)(k(T)\boldsymbol{z}) = a(T)(\mathbf{0}) + b(T)(\mathbf{0}) = \mathbf{0}.$$

Cualquier $\boldsymbol{x} \in V$ es de la forma $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$ con $\boldsymbol{y} \in U$, $\boldsymbol{z} \in W$. Si $\boldsymbol{x} \in U_1$, entonces $\boldsymbol{y} \in U_1$ y $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$ queda también en U_1 y por tanto $\boldsymbol{z} = \mathbf{0}$ y $\boldsymbol{x} = \boldsymbol{y}$. Se ha mostrado que $U_1 = U$.

Se concluye que $U_1 \oplus (U_2 \oplus \cdots \oplus U_r)$ es la descomposición de V que corresponde a la factorización $p_T(t) = h(t)\,k(t)$ por la Proposición 2.38. La Proposición 2.39 ahora muestra que el polinomio característico de T restringido a U_1 es $h(t) = (t-\alpha_1)^{k_1}$. Además, el polinomio característico de T restringido a $U_2 \oplus \cdots \oplus U_r$ es $k(t) = (t-\alpha_2)^{k_2} \cdots (t-\alpha_r)^{k_r}$, que es lo que se requiere para aplicar la hipótesis inductiva. $\square$

► Con la descomposición primaria ya en mano, ahora es posible ofrecer otro criterio de diagonalizabilidad.

Proposición 2.45. *Una matriz $A \in M_n(\mathbb{F})$ es diagonalizable si y solo si su polinomio mínimo $q_A(t)$ se descompone en* factores distintos de primer grado.

Demostración. Si $A = PDP^{-1}$ con D diagonal, entonces $q_A(t) = q_D(t)$ es de la forma (2.20), con factores distintos de primer grado.

Inversamente, si $q_A(t) = (t-\alpha_1)(t-\alpha_2)\cdots(t-\alpha_r)$ con $\alpha_1, \dots, \alpha_r$ distintos, el Teorema 2.43 muestra que $\mathbb{F}^n$ posee una descomposición primaria de la forma

$$\mathbb{F}^n = U_1 \oplus U_2 \oplus \cdots \oplus U_r, \quad \text{donde cada} \quad U_k := \ker(T_A - \alpha_k\,1).$$

Sea $\mathcal{B}_k$ una base del subespacio U_k, de modo que su unión disjunta $\mathcal{B} := \mathcal{B}_1 \uplus \cdots \uplus \mathcal{B}_r$ es una base de $\mathbb{F}^n$. Si $\boldsymbol{x} \in \mathcal{B}_k$, entonces $A\boldsymbol{x} - \alpha_k\boldsymbol{x} = (T_A - \alpha_k\,1)(\boldsymbol{x}) = \boldsymbol{0}$, así que $A\boldsymbol{x} = \alpha_k\boldsymbol{x}$: cada elemento de la base $\mathcal{B}$ es un *autovector* de A. El Corolario 2.35 ahora dice que A es diagonalizable. Concretamente, el cambio de la base estándar $\mathcal{E}$ a la base $\mathcal{B}$ diagonaliza la matriz A:

$$P^{-1}AP = [T_A]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} \alpha_1 & \dots & 0 & & & & & & \\ \vdots & \ddots & \vdots & & & & & & \\ 0 & \dots & \alpha_1 & & & & & & \\ & & & \alpha_2 & \dots & 0 & & & \\ & & & \vdots & \ddots & \vdots & & & \\ & & & 0 & \dots & \alpha_2 & & & \\ & & & & & & \ddots & & \\ & & & & & & & \alpha_r & \dots & 0 \\ & & & & & & & \vdots & \ddots & \vdots \\ & & & & & & & 0 & \dots & \alpha_r \end{bmatrix}$$ $\square$

Ejemplo 2.46. La matriz $J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ no es diagonalizable sobre $\mathbb{R}$, pero sí es diagonalizable sobre $\mathbb{C}$. En efecto, su polinomio característico $p_J(t) = t^2 + 1$ es *irreducible* sobre $\mathbb{R}$. Como $q_J(t) \setminus p_J(t)$, también vale $q_J(t) = t^2 + 1$, el cual no posee factores reales de primer grado.

En cambio, al tomar $\mathbb{F} = \mathbb{C}$, se ve que $p_J(t) = (t-i)(t+i)$ y también $q_J(t) = (t-i)(t+i)$. Luego J es diagonalizable sobre $\mathbb{C}$, con forma diagonal $\operatorname{diag}[i, -i]$, como ya se ha visto en el Ejemplo 2.13. ◊

Ejemplo 2.47. Sea $\mathbb{F}$ un cuerpo cualquiera, y considérese la matriz triangular

$$A = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \quad \text{con} \quad \lambda \in \mathbb{F}.$$

Esta matriz *no es diagonalizable*. En efecto, vale $p_A(t) = (t-\lambda)^2$, así que $q_A(t) = t - \lambda$ o bien $q_A(t) = (t-\lambda)^2$. La posibilidad $q_A(t) = t - \lambda$ queda excluida porque $A - \lambda\,1_2 \neq 0$; por otro lado, se ve que $(A - \lambda\,1_2)^2 = 0$ por el teorema de Cayley y Hamilton o bien por cálculo directo.

El polinomio mínimo $q_A(t) = (t-\lambda)^2$ no es un producto de factores *distintos* de primer grado. Fíjese, *en passant*, que se ha comprobado lo siguiente:

$$\text{Las dos matrices} \quad \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix} \quad \text{y} \quad \begin{bmatrix} \lambda & 0 \\ 0 & \lambda \end{bmatrix} \quad \textit{no son semejantes}.$$ ◊

2.5. La forma de Jordan de una matriz compleja

El Ejemplo 2.46 anterior hace ver que la diagonalizabilidad de una matriz depende del cuerpo de los escalares. El cuerpo $\mathbb{C}$ (los números complejos) posee una propiedad fundamental, el "Teorema Fundamental del Álgebra":[17] *cualquier polinomio de grado n en* $\mathbb{C}[t]$ *posee n raíces complejas (no necesariamente distintas)*, o lo que es lo mismo, cualquier $p(t) \in \mathbb{C}[t]$ posee una factorización completa

$$p(t) = a_n(t-\alpha_1)\cdots(t-\alpha_n),$$

con $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ no necesariamente distintos.

Para simplificar la discusión, en esta sección se supondrá que $\mathbb{F} = \mathbb{C}$, de modo que los polinomios $p_A(t)$ y $q_A(t)$ tengan factores irreducibles de primer grado solamente.

El Teorema 2.43 de la descomposición primaria y su Corolario 2.44 muestran que cualquier operador lineal posee una matriz en $M_n(\mathbb{C})$ que es una suma directa de bloques diagonales:

$$[T]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} A_1 & 0 & \dots & 0 \\ 0 & A_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & A_r \end{bmatrix}. \tag{2.26}$$

[17]La primera demostración rigurosa de este teorema fue dada por Jean-Robert Argand en 1806. (Una prueba corta emplea el teorema de Liouville, que dice que una función holomorfa acotada definida en toda $z \in \mathbb{C}$ es necesariamente constante. Si $p(t) \in \mathbb{C}[t]$ no tuviera raíz alguna, entonces $z \mapsto 1/p(z)$ sería una función holomorfa acotada y por tanto constante.) Nótese que si $\operatorname{gr} p(t) \geqslant 1$, solo hace falta que $p(t)$ tenga *una* raíz α, porque se puede considerar el cociente $p(t)/(t-\alpha)$ para obtener otra raíz, y así sucesivamente.

Esto se ve al elegir bases $\mathcal{B}_1, \dots, \mathcal{B}_r$ para los subespacios $U_1, \dots, U_r$ de la descomposición primaria y tomar la base $\mathcal{B}$ de V como su unión disjunta: $\mathcal{B} = \mathcal{B}_1 \uplus \cdots \uplus \mathcal{B}_r$. Como cada subespacio U_i reduce T, los bloques no diagonales son rectángulos de ceros. Cada A_i es una matriz cuadrada $k_i \times k_i$, donde $k_i = \dim U_i$; su polinomio característico es $p_{A_i}(t) = (t - \alpha_i)^{k_i}$. (Fíjese que $k_1 + \cdots + k_r = n$.)

Al restar $\alpha_i \, 1_{k_i}$ de cada bloque, se obtiene una matriz $\underline{N_i} := A_i - \alpha_i \, 1_{k_i}$. El teorema de Cayley y Hamilton para la matriz A_i muestra que

$$N_i^{k_i} = (A_i - \alpha_i \, 1_{k_i})^{k_i} = 0.$$

Definición 2.48. Un operador $T \in \mathcal{L}(V)$ es **nilpotente** si $\underline{T^k = 0}$ para algún $k \in \{1, 2, 3, \dots\}$. Una matriz $A \in M_n(\mathbb{C})$ es una **matriz nilpotente** si $\underline{A^k = 0}$ para algún $k \in \{1, 2, 3, \dots\}$. $\Diamond$

► Si λ es un autovalor de un operador lineal nilpotente T, con un autovector $\boldsymbol{x} \neq \mathbf{0}$, entonces $\lambda^k \boldsymbol{x} = T^k(\boldsymbol{x}) = \mathbf{0}$ y por tanto $\lambda^k = 0$, luego $\lambda = 0$. Por lo tanto, 0 *es el único autovalor de un operador nilpotente*. Los vectores no nulos en $\ker T$ son los autovectores correspondientes. El polinomio característico de T es $p_T(t) = (t - 0)^n = t^n$. Si $k \in \mathbb{N}$ es el menor entero positivo tal que $T^k = 0$, el polinomio mínimo de T es $q_T(t) = t^k$.

Proposición 2.49. *Cualquier matriz $A \in M_n(\mathbb{C})$ es de la forma $\underline{A = B + N}$, donde B es diagonalizable, N es nilpotente y además $BN = NB$.*

Demostración. El polinomio característico de A se descompone en factores de primer grado:

$$p_A(t) = (t - \alpha_1)^{k_1}(t - \alpha_2)^{k_2} \cdots (t - \alpha_r)^{k_r},$$

donde $\alpha_1, \dots, \alpha_r \in \mathbb{C}$ son *distintos* y $k_1, \dots, k_r$ son enteros positivos.

Después de una cambio de base $A \mapsto P^{-1}AP$, según la descomposición primaria del operador T_A, se obtiene una suma directa de bloques (2.26). Supóngase, como primer caso, que la matriz A ya tiene esa forma. Sea N la matriz de bloques

$$N := \begin{bmatrix} N_1 & 0 & \dots & 0 \\ 0 & N_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & N_r \end{bmatrix}, \quad \text{con} \quad N_i := A_i - \alpha_i \, 1_{k_i}; \quad i = 1, \dots, r.$$

Entonces para $k := \text{máx}\{k_1, \dots, k_r\}$ se ve que $N^k = 0$; lo cual dice que N es nilpotente. Sea B la suma directa de los bloques diagonales $\alpha_i \, 1_{k_i}$; esta es una matriz diagonal cuyos autovalores son los α_i, repetidas con multiplicidades respectivas k_i. Es obvio que $A = B + N$. En cada bloque, el producto BN se reduce al producto de la *matriz escalar* $\alpha_i \, 1_{k_i}$ con la matriz $N_i \in M_{k_i}(\mathbb{C})$, de donde sigue $BN = NB$.

En el caso general, donde hace falta hacer un cambio de base de la forma $A \mapsto P^{-1}AP$, el párrafo anterior dice que $P^{-1}AP = \widetilde{B} + \widetilde{N}$ donde $\widetilde{B}$ es diagonal; $\widetilde{N}$ es nilpotente con $\widetilde{N}^k = 0$; y $\widetilde{B}\widetilde{N} = \widetilde{N}\widetilde{B}$. Colóquese $B := P\widetilde{B}P^{-1}$ y $N := P\widetilde{N}P^{-1}$. Entonces:

$$B + N = P(\widetilde{B} + \widetilde{N})P^{-1} = P(P^{-1}AP)P^{-1} = A; \quad \text{y} \quad BN = P\widetilde{B}\widetilde{N}P^{-1} = P\widetilde{N}\widetilde{B}P^{-1} = NB.$$

La matriz B es diagonalizable por definición, pues $\widetilde{B}$ es diagonal; y N es nilpotente, porque

$$N^k = (P\widetilde{N}P^{-1})^k = P\widetilde{N}P^{-1}\,P\widetilde{N}P^{-1} \cdots P\widetilde{N}P^{-1} = P\widetilde{N}^kP^{-1} = P\,0\,P^{-1} = 0. \qquad \square$$

► La estructura de una matriz diagonal es clara: después de reordenar los elementos diagonales (al permutar los vectores de la base dada), se obtiene un juego de valores diagonales, cada uno repetida con su multiplicidad. Falta averiguar la estructura de una matriz nilpotente.

Sea, entonces, $N \in M_n(\mathbb{C})$ una matriz tal que $N^k = 0$ pero $N^{k-1} \neq 0$ para algún entero positivo k (que depende de N). Hay al menos un vector $\boldsymbol{x} \in \mathbb{C}^n$, $\boldsymbol{x} \neq \boldsymbol{0}$, tal que $N^{k-1}\boldsymbol{x} \neq \boldsymbol{0}$. Para cada $l = 0, 1, \dots, k$, considérese el subespacio

$$\underline{V_l := \ker T_{N^l}} = \{\, \boldsymbol{y} \in \mathbb{C}^n : N^l\boldsymbol{y} = \boldsymbol{0} \,\}.$$

Entonces $V_0 = \{\boldsymbol{0}\}$, $V_k = \mathbb{C}^n$ y además $V_{l-1} \subseteq V_l$ para $l = 1, \dots, k$, porque

$$N^{l-1}\boldsymbol{x} = \boldsymbol{0} \implies N^l\boldsymbol{x} = N(N^{l-1}\boldsymbol{x}) = N\boldsymbol{0} = \boldsymbol{0}.$$

De este modo, los V_l forman una *cadena creciente de subespacios*:[18]

$$\{\boldsymbol{0}\} = V_0 \subseteq V_1 \subseteq \dots \subseteq V_{k-1} \subseteq V_k = \mathbb{C}^n. \tag{2.27}$$

Sea $m_l := \dim V_l$, de modo que

$$0 = m_0 \leqslant m_1 \leqslant \dots \leqslant m_{k-1} \leqslant m_k = n.$$

También es evidente que $\boldsymbol{y} \in V_l \implies N\boldsymbol{y} \in V_{l-1}$ para cada l, porque $N^{l-1}(N\boldsymbol{y}) = N^l\boldsymbol{y} = \boldsymbol{0}$.

Se debe elegir una base conveniente para $\mathbb{C}^n$, que será la unión creciente de bases para los V_l. Se requiere un lema auxiliar, a continuación.

Lema 2.50. *Sea V un espacio $\mathbb{F}$-vectorial de dimensión n y sea $W \leqslant V$ un subespacio de dimensión m. Se puede elegir vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{n-m} \in V$ que son* ***linealmente independientes sobre*** *W, lo cual significa que*

$$c_1\boldsymbol{x}_1 + \cdots + c_{n-m}\boldsymbol{x}_{n-m} \in W \quad \textit{solo si} \quad c_1 = \cdots = c_{n-m} = 0.$$

[18] En la geometría algebraica, una cadena de subespacios como (2.27) se llama una **bandera**, por analogía con una bandera corriente $V_0 \subset V_1 \subset V_2$, donde un rectángulo de tela V_2 viene apoyado por un asta V_1, coronada por un botón u ornamento puntual V_0.

Demostración. El espacio cociente V/W tiene dimensión $n-m$. Sea $\{z_1,\dots,z_{n-m}\}$ una base de V/W. Cada z_i es una coclase de la forma $z_i = x_i + W$ para algún $x_i \in V$.

Una relación de la forma $c_1\boldsymbol{x}_1 + \cdots + c_{n-m}\boldsymbol{x}_{n-m} \in W$ implica que

$$\begin{aligned} c_1 z_1 + \cdots + c_{n-m} z_{n-m} &= c_1(\boldsymbol{x}_1 + W) + \cdots + c_{n-m}(\boldsymbol{x}_{n-m} + W) \\ &= (c_1\boldsymbol{x}_1 + \cdots + c_{n-m}\boldsymbol{x}_{n-m}) + W = W. \end{aligned}$$

Pero la coclase trivial $W = \mathbf{0} + W$ es el elemento cero de V/W. Luego, la independencia lineal de los z_i en V/W conlleva $c_1 = \cdots = c_{n-m} = 0$. □

Proposición 2.51. *Sea $N \in M_n(\mathbb{C})$ una matriz nilpotente. Si k es el menor entero positivo tal que $N^k = 0$, sea $V_l := \ker T_{N^l}$ para $l = 0, 1, \dots, k$. Entonces V posee una base $\mathcal{B}$ tal que:*

(a) *$\mathcal{B}$ incluye una base de V_l para cada $l = 1, \dots, k$;*

(b) *si $\boldsymbol{x} \in \mathcal{B}$, entonces bien $N\boldsymbol{x} \in \mathcal{B}$ o bien $N\boldsymbol{x} = 0$.*

Demostración. Escríbase $\underline{r_1} := m_k - m_{k-1} = n - \dim V_{k-1}$. Por el Lema 2.50 anterior, hay vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{r_1} \in \mathbb{C}^n = V_k$ que son linealmente independientes sobre V_{k-1}.

Los vectores $N\boldsymbol{x}_1, \dots, N\boldsymbol{x}_{r_1}$ están en V_{k-1}, porque $N^{k-1}(N\boldsymbol{x}_j) = N^k\boldsymbol{x}_j = \mathbf{0}$ para cada j. Resulta que estos vectores son linealmente independientes sobre V_{k-2}. En efecto,

$$\begin{aligned} c_1 N\boldsymbol{x}_1 + \cdots + c_{r_1} N\boldsymbol{x}_{r_1} \in V_{k-2} &\implies N(c_1\boldsymbol{x}_1 + \cdots + c_{r_1}\boldsymbol{x}_{r_1}) \in V_{k-2} \\ &\implies c_1\boldsymbol{x}_1 + \cdots + c_{r_1}\boldsymbol{x}_{r_1} \in V_{k-1}, \\ &\implies c_1 = \cdots = c_{r_1} = 0. \end{aligned}$$

Se deduce que $r_1 \leqslant \dim(V_{k-1}/V_{k-2}) = m_{k-1} - m_{k-2}$. Escríbase $\underline{r_2} := m_{k-1} - m_{k-2}$; se ha mostrado que $r_2 \geqslant r_1$.

Si $r_2 > r_1$, en vista del Lema 2.50, hay vectores $\boldsymbol{x}_{r_1+1}, \dots, \boldsymbol{x}_{r_2}$ tales que el conjunto $\{N\boldsymbol{x}_1, \dots, N\boldsymbol{x}_{r_1}, \boldsymbol{x}_{r_1+1}, \dots, \boldsymbol{x}_{r_2}\} \subset V_{k-1}$ es linealmente independiente sobre V_{k-2}. Si $r_2 = r_1$, el conjunto $\{N\boldsymbol{x}_1, \dots, N\boldsymbol{x}_{r_1}\}$ juega el mismo papel.

Ahora los vectores $N^2\boldsymbol{x}_1, \dots, N^2\boldsymbol{x}_{r_1}, N\boldsymbol{x}_{r_1+1}, \dots, N\boldsymbol{x}_{r_2} \in V_{k-2}$ son linealmente independientes sobre V_{k-3}. En efecto, fíjese que

$$\begin{aligned} &b_1 N^2\boldsymbol{x}_1 + \cdots + b_{r_1} N^2\boldsymbol{x}_{r_1} + b_{r_1+1} N\boldsymbol{x}_{r_1+1} + \cdots + b_{r_2} N\boldsymbol{x}_{r_2} \in V_{k-3} \\ &\quad\implies b_1 N\boldsymbol{x}_1 + \cdots + b_{r_1} N\boldsymbol{x}_{r_1} + b_{r_1+1}\boldsymbol{x}_{r_1+1} + \cdots + b_{r_2}\boldsymbol{x}_{r_2} \in V_{k-2} \\ &\quad\implies b_1 = \cdots = b_{r_1} = b_{r_1+1} = \cdots = b_{r_2} = 0. \end{aligned}$$

Se deduce que $\underline{r_3} := m_{k-2} - m_{k-3}$ cumple $r_3 \geqslant r_2$. Si $r_3 > r_2$, hay vectores $\boldsymbol{x}_{r_2+1}, \dots, \boldsymbol{x}_{r_3}$ tales que el conjunto $\{N^2\boldsymbol{x}_1, \dots, N^2\boldsymbol{x}_{r_1}, N\boldsymbol{x}_{r_1+1}, \dots, N\boldsymbol{x}_{r_2}, \boldsymbol{x}_{r_2+1}, \dots, \boldsymbol{x}_{r_3}\} \subset V_{k-2}$ es linealmente independiente sobre V_{k-3}.

Ad (a): Al repetir este proceso k veces, se obtiene el siguiente juego de vectores en $\mathbb{C}^n$:

$$
\begin{aligned}
&\boldsymbol{x}_1, \dots, \boldsymbol{x}_{r_1}, \\
&N\boldsymbol{x}_1, \dots, N\boldsymbol{x}_{r_1}, \quad \boldsymbol{x}_{r_1+1}, \dots, \boldsymbol{x}_{r_2}, \\
&N^2\boldsymbol{x}_1, \dots, N^2\boldsymbol{x}_{r_1}, \quad N\boldsymbol{x}_{r_1+1}, \dots, N\boldsymbol{x}_{r_2}, \quad \boldsymbol{x}_{r_2+1}, \dots, \boldsymbol{x}_{r_3}, \\
&\quad \vdots \\
&N^{k-1}\boldsymbol{x}_1, \dots, N^{k-1}\boldsymbol{x}_{r_1}, \quad N^{k-2}\boldsymbol{x}_{r_1+1}, \dots, N^{k-2}\boldsymbol{x}_{r_2}, \quad N^{k-3}\boldsymbol{x}_{r_2+1}, \dots, N^{k-3}\boldsymbol{x}_{r_3}, \\
&\qquad \dots, \quad N\boldsymbol{x}_{r_{k-2}+1}, \dots, N\boldsymbol{x}_{r_{k-1}}, \quad \boldsymbol{x}_{r_{k-1}+1}, \dots, \boldsymbol{x}_{r_k}.
\end{aligned}
\tag{2.28}
$$

Aquí se ha escrito $r_j := m_{k-j+1} - m_{k-j}$ para $j = 1, \dots, k$. Todos los vectores en esta tabla son linealmente independientes, por construcción. El número total de vectores en esta lista es

$$
\begin{aligned}
r_1 + r_2 + \cdots + r_k &= (m_k - m_{k-1}) + (m_{k-1} - m_{k-2}) + \cdots + (m_2 - m_1) + m_1 \\
&= m_k = n.
\end{aligned}
$$

Por lo tanto, estos vectores forman una **base** $\mathcal{B}$ de $\mathbb{C}^n$. La construcción muestra que las últimas $(k - l)$ filas de la lista (2.28) pertenecen a V_l; y la suma telescópica para las filas anteriores:

$$
r_1 + \cdots + r_{k-l} = m_k - m_l = n - \dim V_l
$$

dice que hay m_l entradas en las últimas $(k - l)$ filas; por ende, ellas forman una base de V_l.

Ad (b): La propiedad (b) es inmediato de la forma explícita (2.28) de la base $\mathcal{B}$. De hecho, si $\boldsymbol{y}$ es un vector de la última fila, entonces $\boldsymbol{y} \in V_1 = \ker T_N$, así que $N\boldsymbol{y} = \boldsymbol{0}$. Si $\boldsymbol{z}$ es un vector de cualquier otra fila, entonces $N\boldsymbol{z}$ es un miembro de la fila siguiente. □

En la lista de vectores (2.28), la primera entrada fila genera un subespacio invariante para T_N, puesto que $N^k = 0$. De hecho, este subespacio *reduce* T_N porque las demás entradas generan un subespacio suplementario, también invariante. Conviene reordenar la base $\mathcal{B}$ de la siguiente forma (leyendo las filas de (2.28) de abajo para arriba):

$$
\begin{aligned}
\mathcal{B} = \{\, &N^{k-1}\boldsymbol{x}_1, \dots, N\boldsymbol{x}_1, \boldsymbol{x}_1, \quad N^{k-1}\boldsymbol{x}_2, \dots, N\boldsymbol{x}_2, \boldsymbol{x}_2, \ \dots, \quad N^{k-1}\boldsymbol{x}_{r_1}, \dots, N\boldsymbol{x}_{r_1}, \boldsymbol{x}_{r_1}, \\
&N^{k-2}\boldsymbol{x}_{r_1+1}, \dots, N\boldsymbol{x}_{r_1+1}, \boldsymbol{x}_{r_1+1}, \ \dots, \quad N^{k-2}\boldsymbol{x}_{r_2}, \dots, N\boldsymbol{x}_{r_2}, \boldsymbol{x}_{r_2}, \\
&N^{k-3}\boldsymbol{x}_{r_2+1}, \dots, N\boldsymbol{x}_{r_2+1}, \boldsymbol{x}_{r_2+1}, \ \dots, \quad N^{k-3}\boldsymbol{x}_{r_3}, \dots, N\boldsymbol{x}_{r_3}, \boldsymbol{x}_{r_3}, \\
&\quad \vdots \\
&N\boldsymbol{x}_{r_{k-2}+1}, \boldsymbol{x}_{r_{k-2}+1}, \ \dots, \quad N\boldsymbol{x}_{r_{k-1}}, \boldsymbol{x}_{r_{k-1}}, \\
&\boldsymbol{x}_{r_{k-1}+1}, \dots, \boldsymbol{x}_{r_k} \,\}.
\end{aligned}
\tag{2.29}
$$

Al agrupar los términos en este orden, se obtiene un juego de r_k subespacios que reducen T_N. La matriz $[T_N]^{\mathcal{B}}_{\mathcal{B}}$ así obtenida es semejante a N (desde luego) y conforma una *suma directa de bloques*.

En esta suma directa, hay r_1 bloques $k \times k$; seguido de $r_2 - r_1$ bloques $(k-1) \times (k-1)$; etc., hasta $r_{k-1} - r_{k-2}$ bloques 2×2; y por último un bloque de ceros que corresponde al subespacio $V_1 = \ker T_N = \operatorname{lin}\langle \boldsymbol{x}_{r_{k-1}+1}, \dots, \boldsymbol{x}_{r_k} \rangle$.

Para determinar las entradas de estos bloques, basta examinar el primero de ellos, determinado por las igualdades:

$$T_N(N^{k-j}\boldsymbol{x}_1) = N^{k-j+1}\boldsymbol{x}_1, \quad \text{para} \quad j = 1, \dots, k.$$

Al poner $\boldsymbol{y}_j := N^{k-j}\boldsymbol{x}_1$, se ve que $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_k\} = \{N^{k-1}\boldsymbol{x}_1, \dots, N\boldsymbol{x}_1, \boldsymbol{x}_1\}$ es la base (ordenada) del primer subespacio invariante, y además:

$$T_N(\boldsymbol{y}_1) = \boldsymbol{0}, \quad T_N(\boldsymbol{y}_2) = \boldsymbol{y}_1, \quad T_N(\boldsymbol{y}_3) = \boldsymbol{y}_2, \quad \dots, \quad T_N(\boldsymbol{y}_k) = \boldsymbol{y}_{k-1}, \cdot$$

La matriz correspondiente es el bloque

$$J_k(0) := \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix} \tag{2.30}$$

Esta es una matriz triangular con ceros en la diagonal y entradas 1 en la *subdiagonal superior*: $a_{12} = a_{23} = \cdots = a_{k-1,k} = 1$.

La matriz de T_N es la base $\mathcal{B}$ es entonces la suma directa de r_1 bloques $J_k(0)$, seguido por $(r_2 - r_1)$ bloques $J_{k-1}(0)$, etc., hasta $(r_{k-1} - r_{k-2})$ bloques $J_2(0)$, más un bloque nulo de lado $(r_k - r_{k-1}) = \dim \ker T_N$.

Definición 2.52. Sea $k \in \{2, 3, \dots\}$ y sea $\lambda \in \mathbb{C}$. El **bloque de Jordan** $J_k(\lambda)$ es la matriz triangular superior en $M_k(\mathbb{C})$ dada por

$$\underline{J_k(\lambda)} := \lambda\, 1_k + J_k(0) = \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & \lambda & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & \lambda \end{bmatrix}. \tag{2.31}$$

Conviene escribir $\underline{J_1(\lambda)} := [\lambda] \in M_1(\mathbb{C}) = \mathbb{C}$, como caso trivial. ◊

Lema 2.53. *El polinomio mínimo de un bloque de Jordan es igual a su polinomio característico: si $A = J_k(\lambda)$, entonces $q_A(t) = p_A(t) = (t - \lambda)^k$.*

Demostración. Si $A = J_k(\lambda)$, es obvio que $p_A(t) = \det J_k(t - \lambda) = (t - \lambda)^k$.

Basta entonces comprobar que $(J_k(\lambda) - \lambda\, 1_k)^l \neq 0$ cuando $k \geqslant 2$ y $l < k$.

Ahora $J_k(\lambda) - \lambda\, 1_k = J_k(0)$ es la matriz triangular nilpotente (2.30). Al renombrarla $B = J_k(0)$, se ve que las únicas entradas no nulas de $C = B^2$ son

$$c_{13} = b_{12}b_{23} = 1, \quad c_{24} = b_{23}b_{34} = 1, \quad \dots, \quad c_{k-2,k} = b_{k-2,k-1}b_{k-1,k} = 1.$$

Por inducción sobre l, se ve que en $R = B^l$ las únicas entradas no ceros son $r_{1,l+1} = r_{2,l+2} = \cdots = r_{k-l,k} = 1$. Por ejemplo, para $B = J_4(0)$ se ve que

$$B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad B^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad B^3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix},$$

y después, $B^4 = 0$. En tamaño $k \times k$, se obtiene $B^l \neq 0$ para $l = 1, 2, \dots, k - 1$, pero $B^l = 0$. Luego $q_B(t) = t^k$ y en seguida $q_A(t) = (t - \lambda)^k$. □

► Al aplicar las consideraciones anteriores sobre matrices complejas nilpotentes al caso de una matriz compleja cualquiera, se obtiene el siguiente teorema que identifica todas las clases de equivalencia de matrices complejas bajo la relación de semejanza.

Teorema 2.54. *Cualquier matriz $A \in M_n(\mathbb{C})$ es semejante a una suma directa de bloques de Jordan de la forma $J_l(\alpha_i)$, donde $\{\alpha_1, \dots, \alpha_r\}$ son los autovalores* distintos *de A. Para cada α_i, el mayor tamaño $l \times l$ de los bloques $J_l(\alpha_i)$ en esta suma directa es el exponente l_i del factor $(t - \alpha_i)^{l_i}$ en el polinomio mínimo $q_A(t)$ de la matriz A.*

Demostración. El polinomio mínimo $p_A(t)$ escinde en $\mathbb{C}[t]$ y por ende es de la forma (2.24):

$$p_A(t) = (t - \alpha_1)^{k_1}(t - \alpha_2)^{k_2} \cdots (t - \alpha_r)^{k_r},$$

donde $\alpha_1, \dots, \alpha_r \in \mathbb{C}$ son distintos y $k_1 + \cdots + k_r = n$. Su polinomio mínimo tiene la forma (2.25):

$$q_A(t) = (t - \alpha_1)^{l_1}(t - \alpha_2)^{l_2} \cdots (t - \alpha_r)^{l_r},$$

donde $1 \leqslant l_i \leqslant k_i$ en cada caso, en vista del Corolario 2.44.

Sea $\mathbb{C}^n = U_1 \oplus \cdots \oplus U_r$ la descomposición primaria debida a esta factorización de $p_A(t)$. Elíjase una base $\mathcal{B}_i$ para cada U_i, cuya unión disjunta $\mathcal{B} = \mathcal{B}_i \uplus \cdots \uplus \mathcal{B}_r$ es una base de $\mathbb{C}^n$. Al cambiar la base estándar $\mathcal{E}$ de $\mathbb{C}^n$ a esta base $\mathcal{B}$, se deduce que la matriz A es semejante a una suma directa $A = A_1 \oplus \cdots \oplus A_r$, como en (2.26).

En esta suma directa, cada bloque A_i tiene la forma $A_i = \alpha_i\,1_{k_i} + N_i$ donde N_i es una matriz cuadrada nilpotente, con $N_i^{l_i} = 0$ en $M_{k_i}(\mathbb{C})$.

Ahora, la matriz escalar $\alpha_i\,1_{k_i}$ no sufre cambio alguno al reemplazar la base $\mathcal{B}_i$ por cualquier otra base de U_i. Se puede entonces suponer que $\mathcal{B}_i$ es aquella que expresa T_{N_i} como suma directa de bloques de Jordan $J_l(0)$, como en (2.30).

Por la construcción de esta base en la demostración de la Proposición 2.51, se ve que $l \leqslant l_i$ en cada caso; y que hay al menos un bloque de lado l_i. Al sumarles los bloques escalares $\alpha_i\,1_l$, resulta que cada A_i es una suma directa de bloques de Jordan $J_l(\alpha_i)$. □

Definición 2.55. Sea $A \in M_n(\mathbb{C})$ una matriz cuadrada compleja. La suma directa de bloques de Jordan (en orden decreciente de tamaños de bloques) dada por el Teorema 2.54 anterior, se llama la **forma normal de Jordan** de la matriz A. ◊

El Teorema 2.54 proporciona una descripción completa de la estructura de una matriz cuadrada compleja, o bien la de un operador lineal sobre un espacio $\mathbb{C}$-vectorial de dimensión finita. Esta descripción es aplicable a matrices u operadores con otros cuerpos $\mathbb{F}$ de escalares, toda vez que sus polinomios característicos escinden en $\mathbb{F}[t]$.

2.6. Ejercicios sobre operadores lineales y matrices

Ejercicio 2.1. Calcular los tres autovalores de la matriz

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

Ejercicio 2.2. Calcular los tres autovalores distintos $\lambda_1, \lambda_2, \lambda_3$ de la matriz

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

Resolver las ecuaciones $(\lambda_j 1_3 - A)\boldsymbol{x}_j = \boldsymbol{0}$ para $j = 1, 2, 3$, para así obtener tres autovectores $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ de A. Sea $P := [\boldsymbol{x}_1\ \boldsymbol{x}_2\ \boldsymbol{x}_3]$ la matriz cuyas columnas son estos autovectores (en orden). Verificar que la matriz $P^{-1}AP$ es diagonal y que sus elementos diagonales son los autovalores de A.

Ejercicio 2.3. Un **cuadrado mágico** de lado n es una matriz $n \times n$ cuyas entradas son los enteros $1, 2, \dots, n^2$ dispuestos de tal manera que la suma de las entradas de cada fila; de cada columna; y de las dos diagonales, es la misma. Comprobar que $\frac{1}{2}n(n^2 + 1)$ es un *autovalor* de esta matriz.

Ejercicio 2.4. Calcular los autovalores de la matriz

$$A := \begin{bmatrix} 2 & -2 & 3 \\ 1 & 1 & 1 \\ 1 & 3 & -1 \end{bmatrix}.$$

Obtener una matriz invertible P cuyas columnas son autovectores de A. En seguida, verificar que las transpuestas (en columnas) de las filas de P^{-1} son autovectores de A^t.

Ejercicio 2.5. Calcular los autovalores de la matriz simétrica

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

Obtener una matriz invertible P, también simétrica, cuyas columnas son autovectores de A.

Ejercicio 2.6. Si λ es un autovalor de una matriz $A \in M_n(\mathbb{F})$, comprobar que λ^k es autovalor de la matriz A^k. Si además A es invertible, mostrar que $1/\lambda$ es un autovalor de A^{-1}.

Ejercicio 2.7. Calcular los polinomios característicos y determinar los autovalores de las matrices

$$A = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}, \qquad B = \begin{bmatrix} \cosh t & \operatorname{senh} t \\ \operatorname{senh} t & \cosh t \end{bmatrix}, \qquad C = \begin{bmatrix} 1 & 1 & 1 \\ 1 & \omega & \omega^2 \\ 1 & \omega^2 & \omega \end{bmatrix},$$

donde $-\pi < \theta \leqslant \pi$; $t \in \mathbb{R}$; y $\omega = e^{2\pi i/3} = \frac{1}{2}(-1 + i\sqrt{3})$.

Ejercicio 2.8. Calcular el polinomio característico de la **matriz compañera** $n \times n$:

$$A = \begin{bmatrix} 0 & 1 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & 1 \\ -a_0 & -a_1 & -a_2 & \cdots & -a_{n-2} & -a_{n-1} \end{bmatrix}$$

para verificar así que $p_A(t) = t^n + a_{n-1}t^{n-1} + \cdots + a_1 t + a_0$. Concluir que todo polinomio $p(t)$ es el polinomio característico de alguna matriz.[19]

[19]Esta matriz A es la “compañera” del polinomio $p_a(t)$.

Ejercicio 2.9. (a) Si $P^{-1}AP = D$ es una matriz diagonal, demostrar que $A^k = PD^kP^{-1}$ para todo $k = 0, 1, 2, 3, \dots$

(b) Calcular los dos autovalores de la matriz $A = \begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix}$ y obtener un par de autovectores correspondientes.

(c) Usar los resultados de las partes (a) y (b) para comprobar que

$$\begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix}^9 = \begin{bmatrix} -1025 & 513 \\ -1026 & 514 \end{bmatrix}.$$

Ejercicio 2.10. Tómese $\mathbb{F} = \mathbb{R}$ ó $\mathbb{C}$ en este Ejercicio. Las fórmulas $A^k = PD^kP^{-1}$ son casos particulares de la receta:

$$f(A) = Pf(D)P^{-1} \quad \text{toda vez que} \quad D = P^{-1}AP.$$

Esta receta es válida, no solo para polinomios $f(t)$ sino también para series de potencias (con radio de convergencia infinita), por aplicación de las fórmulas para A^k a cada potencia. Si $D = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$ es diagonal, la matriz $f(D)$ es también diagonal: en efecto, se ve que $f(D) = \operatorname{diag}[f(\lambda_1), \dots, f(\lambda_n)]$. Se define la **exponencial** de una matriz real o compleja por

$$f(t) = e^t \equiv \sum_{k=0}^{\infty} \frac{1}{k!}\, t^k \implies \underline{\exp A} := \sum_{k=0}^{\infty} \frac{1}{k!}\, A^k.$$

Comprobar los cálculos siguientes:

$$\text{(a):} \quad \exp \begin{bmatrix} 0 & c \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 1 & c \\ 0 & 1 \end{bmatrix}, \qquad \text{(b):} \quad \exp \begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix} = \begin{bmatrix} 2e^{-2} - e & e - e^{-2} \\ 2e^{-2} - 2e & 2e - e^{-2} \end{bmatrix}.$$

Ejercicio 2.11. Una **cadena de Markov** es un proceso probabilístico con un número finito n de "estados", caracterizado por números reales no negativos a_{ij} [el cual representa la probabilidad de un cambio (estado i) $\mapsto$ (estado j) *en un solo paso* del proceso]. Se impone la condición de que $\sum_{j=1}^{n} a_{ij} = 1$. Si $A = [a_{ij}] \in M_n(\mathbb{R})$ es la llamada *matriz de transición* de la cadena de Markov, resulta que la probabilidad de un cambio (estado i) $\mapsto$ (estado j) *en k pasos* es la entrada (i, j) de la matriz A^k. Comprobar que

$$A = \begin{bmatrix} \frac12 & \frac12 & 0 \\ \frac14 & \frac12 & \frac14 \\ 0 & \frac12 & \frac12 \end{bmatrix} \implies A^k = \begin{bmatrix} \frac14 + \left(\frac12\right)^{k+1} & \frac12 & \frac14 - \left(\frac12\right)^{k+1} \\ \frac14 & \frac12 & \frac14 \\ \frac14 - \left(\frac12\right)^{k+1} & \frac12 & \frac14 + \left(\frac12\right)^{k+1} \end{bmatrix} \quad \text{para} \quad k \geqslant 1.$$

Ejercicio 2.12. Sean A y B dos matrices complejas $n \times n$. Mostrar que las matrices AB y BA tienen los mismos autovalores.

⟦ Indicación: considerar un autovector de AB. ⟧

Ejercicio 2.13. Sea $A, B \in M_n(\mathbb{C})$ dos matrices cualesquiera. Demostrar que los polinomios característicos $p_{AB}(t)$ y $p_{BA}(t)$ coinciden.[20]

⟦ Indicación: si $\det A \neq 0$, entonces $BA = A^{-1}(AB)A$. En cambio, si $\det A = 0$, fíjese que $\det(A - \mu) \neq 0$ cuando $\mu \in \mathbb{C}$ no es un autovalor de A; concluir que la expresión

$$h_\lambda(t) := \det(\lambda 1_n - (A - t)B) - \det(\lambda 1_n - B(A - t)),$$

para cada λ fijo, es un polinomio en $\mathbb{C}[t]$ con más de n raíces. ⟧

Ejercicio 2.14. Calcular los polinomios característico y mínimo de las matrices:

$$A = \begin{bmatrix} 3 & 2 & 2 \\ 1 & 4 & 1 \\ -2 & -4 & -1 \end{bmatrix}, \qquad B = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{bmatrix}.$$

Ejercicio 2.15. Sea $T \in \mathcal{L}(M_2(\mathbb{F}))$ el *operador de transposición*, es decir $\underline{T(B) := B^{\mathrm{t}}}$ (véase el Ejercicio 1.12). Calcular los polinomios característico y mínimo de T. Exhibir una base de autovectores para el operador T.

Ejercicio 2.16. (a) Sea $A \in M_n(\mathbb{F})$ una matriz con n autovalores $\lambda_1, \dots, \lambda_n$, no necesariamente distintos. Si $f(t) \in \mathbb{F}[t]$ es un polinomio cualquiera, demostrar que los autovalores de la matriz $f(A)$ son $f(\lambda_1), \dots, f(\lambda_n)$.

(b) Comprobar que la traza de A^k obedece $\underline{\operatorname{tr}(A^k)} = \lambda_1^k + \cdots + \lambda_n^k$ para todo $k \in \mathbb{N}$.

Ejercicio 2.17. Sea $A \in M_n(\mathbb{F})$ una matriz invertible. Demostrar que el coeficiente de t en el polinomio característico $p_A(t)$ es $(-1)^{n-1}(\det A) \operatorname{tr}(A^{-1})$.

Ejercicio 2.18. Una matriz $B \in M_n(\mathbb{C})$ es **idempotente** si $B^2 = B$. Comprobar que la matriz $(1_n - B)$ es también idempotente. Si $B \neq 0$ y $B \neq 1_n$, demostrar que el conjunto de los autovalores distintos de B es $\{0, 1\}$.

¿Qué se puede afirmar acerca de la forma de Jordan de una matriz idempotente B?

Verificar que $r(B) = \operatorname{tr} B$ cuando B es idempotente.

Ejercicio 2.19. Calcular los polinomios característico y mínimo de la matriz

$$A = \begin{bmatrix} 5 & -1 & 0 & 0 & 0 \\ 6 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

[20] Esto no sigue directamente del Ejercicio 2.12 anterior, porque puede haber autovalores repetidos.

Ejercicio 2.20. Obtener la forma normal de Jordan de la matriz triangular siguiente:

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

Para hallarla, se debe proceder así:

(a) identificar un autovector $\boldsymbol{y}$ para el autovalor -3;

(b) identificar los subespacios V_1, V_2, V_3 anulados por $(A - 1_4)$, $(A - 1_4)^2$, $(A - 1_4)^3$, respectivamente;

(c) hallar un vector $\boldsymbol{x} \in V_3 \setminus V_2$ tal que $\{(A - 1_4)^2\boldsymbol{x},\ (A - 1_4)\boldsymbol{x},\ \boldsymbol{x},\ \boldsymbol{y}\}$ sea una base de $\mathbb{F}^4$;

(d) si P es la matriz cuyas columnas son los vectores de esta base, calcular P^{-1};

(e) verificar que la matriz $P^{-1}AP$ es una suma directa de bloques de Jordan.

Ejercicio 2.21. Sea $T : V \to V$ un operador lineal sobre un espacio $\mathbb{C}$-vectorial V. Si λ es un autovalor de T, un **autovector generalizado** de T para λ es un vector $\boldsymbol{x} \in V$ tal que

$$(T - \lambda\, 1_V)^k \boldsymbol{x} = \boldsymbol{0} \quad \text{pero} \quad (T - \lambda\, 1_V)^{k-1}\boldsymbol{x} \neq \boldsymbol{0},$$

para algún $k \geqslant 1$. (Cuando $k = 1$, $\boldsymbol{x}$ es un autovector “ordinario” de T.)

Identificar los autovectores generalizados del operador T_A en el Ejercicio 2.20 anterior.

Demostrar que V posee una base que consiste de autovectores generalizados de T.

⟦ Indicación: usar la forma normal de Jordan de la matriz de T. ⟧

Ejercicio 2.22. Si $A \in M_n(\mathbb{R})$ la matriz simétrica con entradas $a_{jk} = [\![j = k+1]\!] + [\![k = j+1]\!]$, es decir, tiene entradas 1 en las dos subdiagonales principales, y las demás entradas cero. Para $n = 5$, por ejemplo:

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

Sea $B \in M_n(\mathbb{R})$ la matriz con entradas $b_{jk} = \operatorname{sen}\left(\dfrac{jk\pi}{n+1}\right)$. Verificar que las columnas de B son autovectores de A. ¿Cuáles son los autovalores correspondientes? es cierto o falso que la matriz A es diagonalizable?

Ejercicio 2.23. Una matriz $A \in M_n(\mathbb{F})$, sobre un cuerpo $\mathbb{F}$ cualquiera, se llama **semisimple** si su polinomio mínimo $q_A(t)$ es un producto de *factores irreducibles distintos*. Comprobar que una matriz compleja (el caso $\mathbb{F} = \mathbb{C}$) es semisimple si y solo si es diagonalizable.

Encontrar una matriz $B \in M_4(\mathbb{R})$ que es semisimple pero no diagonalizable.

Ejercicio 2.24. Sea $C(t) := C_r t^r + C_{r-1} t^{r-1} + \cdots + C_1 t + C_0 \in M_n(\mathbb{F}[t])$ una matriz $n \times n$ con entradas polinomiales, o lo que es lo mismo, un polinomio con coeficientes C_i en $M_n(\mathbb{F})$. Mostrar que hay otro polinomio matricial $Q(t)$ tal que

$$C(t) = Q(t)\,(t - A) + C(A);$$

es decir, que el *resto* de la "división a la derecha" de $C(t)$ por $(t - A)$ es la matriz

$$C(A) := C_r A^r + C_{r-1} A^{r-1} + \cdots + C_1 A + C_0.$$

Ejercicio 2.25. Sea $A \in M_n(\mathbb{F})$ una matriz con polinomio característico $p_A(t)$ y polinomio mínimo $q_A(t)$. Sea $d_{n-1}(t)$ el máximo común divisor de todos los menores $(n-1) \times (n-1)$ de la matriz $(t\,1_n - A)$, esto es, el máximo común divisor de las entradas de $\operatorname{adj}(t\,1_n - A)$. Por definición, el polinomio $d_{n-1}(t)$ es mónico.

(a) Comprobar que $d_{n-1}(t)$ divide $p_A(t)$. Por ende, hay un polinomio mónico

$$\tilde{q}(t) := \frac{p_A(t)}{d_{n-1}(t)}\,.$$

Verificar que $\tilde{q}(A) = 0$ y como consecuencia, que $q_A(t)$ divide $\tilde{q}(t)$. ⟦ Indicación: usar el Ejercicio 2.24 anterior. ⟧

(b) Si $\tilde{q}(t) = s(t)\,q_A(t)$ en $\mathbb{F}[t]$, demostrar que $s(t) \equiv 1$ y concluir que el polinomio mínimo $q_A(t)$ satisface la fórmula[21]

$$q_A(t) = \frac{p_A(t)}{d_{n-1}(t)}\,. \tag{2.32}$$

Ejercicio 2.26. Usar la fórmula (2.32) del Ejercicio anterior para *calcular* el polinomio mínimo $q_A(t)$ de la matriz siguiente, en forma normal de Jordan:

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

[21] Este ejercicio proporciona una *fórmula* para $q_A(t)$, haciendo constar que existe un proceso algorítmico para obtener el polinomio mínimo. Véase la sección IV.6 del libro: Feliks Gantmacher, *The Theory of Matrices*, tomo 1, Chelsea, New York, 1959.

3 Ortogonalidad y teoría espectral

In the right hands, integration by parts and the Schwarz inequality are still among the most powerful tools of mathematics.

— Mark Kac

Hasta ahora, el cuerpo $\mathbb{F}$ de escalares ha sido arbitrario; los conceptos principales han sido la independencia lineal de vectores y la semejanza de matrices cuadradas. Con una notable excepción: cuando fue necesario suponer que los polinomios característicos escinden en $\mathbb{F}[t]$ para obtener la forma normal de Jordan, se tomó $\mathbb{F} = \mathbb{C}$, los números complejos. De ahora en adelante, para poder abordar los conceptos de *ortogonalidad* entre vectores y *positividad* de matrices, se usará solamente *escalares reales o complejos*.

Así pues, en este capítulo el cuerpo de base será $\mathbb{R}$, los números reales; o bien $\mathbb{C}$, los números complejos. Cuando una discusión es aplicable en los dos casos, se usará la letra $\mathbb{F}$ para denotar $\mathbb{F} = \mathbb{R}$ ó $\mathbb{F} = \mathbb{C}$ indiferentemente.

Si $\alpha \in \mathbb{C}$, $\underline{\alpha = s + it}$ con $s, t \in \mathbb{R}$, denótese por $\underline{\bar{\alpha} = s - it}$ su **conjugado complejo**; desde luego, vale $\bar{\alpha} = \alpha$ si y solo si $\alpha \in \mathbb{R}$.

3.1. Productos escalares reales y complejos

Definición 3.1. Sea V un espacio vectorial sobre $\mathbb{F} = \mathbb{R}$ o $\mathbb{C}$. Un **producto escalar** en V es una operación que a cada par de vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ asocia un escalar $\underline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{F}}$, con las siguientes propiedades: si $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$ y $\alpha \in \mathbb{F}$, entonces

(a) $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \overline{\langle \boldsymbol{y}, \boldsymbol{x} \rangle}$,

(b) $\langle \boldsymbol{x}, \boldsymbol{y} + \boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$,

(c) $\langle \boldsymbol{x}, \alpha\boldsymbol{y} \rangle = \alpha\langle \boldsymbol{x}, \boldsymbol{y} \rangle$,

(d) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geqslant 0$, con igualdad solo si $\boldsymbol{x} = \boldsymbol{0}$ en V. ◊

Algunos libros emplean el término *producto interno* como sinónimo de *producto escalar*.[1]

Ejemplo 3.2. Si $\mathbb{F} = \mathbb{R}$ y $V = \mathbb{R}^n$, el *producto punto* de dos vectores (de columna) es

$$\boxed{\langle \boldsymbol{x}, \boldsymbol{y} \rangle \equiv \boldsymbol{x} \cdot \boldsymbol{y} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.} \tag{3.1a}$$

Esto es un producto escalar real.

[1] En estos apuntes, se prefiere el término *producto escalar*; pero eso es cuestión de gustos, y bien se ha dicho que *de gustibus non est disputandum*. En la literatura matemática, de hecho, abundan los productos internos y externos, como también los productos interiores y exteriores. Para no complicar las cosas antes de tiempo, se recomienda evitar esta terminología.

Si $\mathbb{F} = \mathbb{C}$ y $V = \mathbb{C}^n$, se define análogamente

$$\boxed{\langle \boldsymbol{z}, \boldsymbol{w} \rangle \equiv \bar{\boldsymbol{z}} \cdot \boldsymbol{w} := \bar{z}_1 w_1 + \bar{z}_2 w_2 + \cdots + \bar{z}_n w_n.} \tag{3.1b}$$

Esto es un producto escalar complejo. ◊

Definición 3.3. Una aplicación $T\colon V \to W$ entre dos espacios vectoriales sobre $\mathbb{C}$ se llama **semilineal** (o *antilineal*) si

$$T(\alpha\,\boldsymbol{x} + \beta\,\boldsymbol{y}) = \bar{\alpha}\,\boldsymbol{x} + \bar{\beta}\,\boldsymbol{y} \quad \text{para todo} \quad \boldsymbol{x}, \boldsymbol{y} \in V;\ \alpha, \beta \in \mathbb{C}.$$

Si V, W, Z son tres espacios vectoriales sobre un cuerpo $\mathbb{F}$ cualquiera, se dice que una aplicación $T\colon V \times W \to Z$ es **bilineal** si es *lineal en cada variable por separado*; esto es,

- $\boldsymbol{x} \mapsto T(\boldsymbol{x}, \boldsymbol{y})$ queda en $\mathcal{L}(V, Z)$ para cada $\boldsymbol{y} \in W$;
- $\boldsymbol{y} \mapsto T(\boldsymbol{x}, \boldsymbol{y})$ queda en $\mathcal{L}(W, Z)$ para cada $\boldsymbol{x} \in V$.

Si V, W, Z son espacios vectoriales *sobre* $\mathbb{C}$, se dice que una aplicación $T\colon V \times W \to Z$ es **sesquilineal** si T es semilineal en una variable y lineal en la otra.[2] ◊

Las propiedades (a), (b), (c) de la Definición 3.1 muestran que el producto escalar, considerado como aplicación $V \times V \to \mathbb{F}$, es *bilineal en el caso real* $\mathbb{F} = \mathbb{R}$ pero *sesquilineal en el caso complejo* $\mathbb{F} = \mathbb{C}$. Está claro que (a) y (b) implican que $\langle \boldsymbol{y} + \boldsymbol{z}, \boldsymbol{x} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle + \langle \boldsymbol{z}, \boldsymbol{x} \rangle$ cuando $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$.

Además, la sesquilinealidad, según la propiedad (c) arriba, dice que $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ es lineal en la *segunda* variable, pero semilineal en la *primera* variable. Este convenio, establecido por los trabajos de Dirac en los albores de la mecánica cuántica, tiene diversas ventajas.[3] Sin embargo, se debe advertir que la mayoría de los libros de texto en matemática – en contraste con los de física – adoptan el convenio opuesto, en donde el producto escalar es lineal en la primera variable y semilineal en la segunda. *Caveat lector.*

Ejemplo 3.4. Si $V = C_{\mathbb{R}}[a, b]$ es el espacio vectorial real de todas las funciones continuas $f\colon [a, b] \to \mathbb{R}$, defínase

$$\langle f, g \rangle := \int_a^b f(t)\, g(t)\, dt.$$

Es fácil verificar que este es un producto escalar real.

[2] El prefijo *sesqui-* significa "$\frac{3}{2}$ veces".

[3] Paul Adrien Maurice Dirac (1902–1984), físico inglés, obtuvo una ecuación que describe el comportamiento relativista del electrón y predijo la existencia del positrón. En 1930, publicó un libro *Principles of Quantum Mechanics*, que sentó el formalismo básico de la física cuántica (e incluye sus convenios notacionales).

Por otro lado, en el espacio vectorial complejo $C_{\mathbb{C}}[a,b]$ de todas las funciones continuas $f\colon [a,b] \to \mathbb{C}$, se puede definir

$$\langle f, g\rangle := \int_a^b \overline{f(t)}\, g(t)\, dt.$$

En particular, vale $\langle f, f\rangle = \int_a^b |f(t)|^2\, dt \geqslant 0$, con igualdad si y solo si la función continua f es idénticamente nula en el intervalo $[a,b]$. ◊

Los espacios vectoriales del ejemplo anterior son infinitodimensionales. Para no complicar el panorama con asuntos de análisis tales como la convergencia de las integrales y series, *en adelante se asumirá que todos los espacios vectoriales son de dimensión finita*. Aun así, buena parte de lo que sigue es directamente extensible al caso infinitodimensional.

Ejemplo 3.5. Si $V = \mathbb{R}^{m\times n}$ es el espacio vectorial de matrices $m \times n$ *reales*, defínase

$$\langle A, B\rangle := \operatorname{tr}(A^{\mathrm{t}} B).$$

Esta expresión es evidentemente lineal en A y en B. Para verificar su positividad, es cuestión de notar que

$$\operatorname{tr}(A^{\mathrm{t}} A) = \sum_{j=1}^{n} \sum_{i=1}^{m} a_{ij}^2 \geqslant 0,$$

con igualdad si y solo si todo $a_{ij} = 0$, es decir, $A = 0$. ◊

► La positividad del producto escalar, Definición 3.1(d), permite introducir el concepto de la *longitud de un vector*. En adelante, en este capítulo, V será un espacio vectorial finitodimensional, sobre $\mathbb{R}$ o $\mathbb{C}$ según el contexto, dotado de un producto escalar $\langle -, -\rangle$ fijo.

Definición 3.6. Se define la **norma** de un vector $\boldsymbol{x} \in V$ por

$$\boxed{\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}.} \tag{3.2}$$

También se dice que $\|\boldsymbol{x}\|$ es la **longitud** de $\boldsymbol{x}$ (así, *norma* y *longitud* son sinónimos.) ◊

Proposición 3.7. *Se verifica la* ***desigualdad de Schwarz****:*

$$\boxed{\bigl|\langle \boldsymbol{x}, \boldsymbol{y}\rangle\bigr| \leqslant \|\boldsymbol{x}\|\, \|\boldsymbol{y}\|} \quad \textit{para todo} \quad \boldsymbol{x}, \boldsymbol{y} \in V. \tag{3.3}$$

con igualdad si y solo si $\boldsymbol{x}, \boldsymbol{y}$ *son proporcionales.*

Demostración. Está claro que (3.3) se cumple con igualdad si $\boldsymbol{x} = \mathbf{0}$ o bien si $\boldsymbol{y} = \mathbf{0}$.[4] Supóngase entonces que $\boldsymbol{x} \neq \mathbf{0}$, $\boldsymbol{y} \neq \mathbf{0}$; y además, en el caso $\mathbb{F} = \mathbb{C}$, que $\langle \boldsymbol{x}, \boldsymbol{y}\rangle$ es *real*.

[4] El vector $\mathbf{0}$ es proporcional a cualquier vector $\boldsymbol{x}$ porque $\mathbf{0} = 0\boldsymbol{x}$.

Para $t \in \mathbb{R}$, colóquese $f(t) := \|\boldsymbol{x} + t\boldsymbol{y}\|^2$. Entonces

$$\begin{aligned} f(t) &= \langle \boldsymbol{x} + t\boldsymbol{y}, \boldsymbol{x} + t\boldsymbol{y} \rangle \\ &= \langle \boldsymbol{x}, \boldsymbol{x} \rangle + 2t\langle \boldsymbol{x}, \boldsymbol{y} \rangle + t^2\langle \boldsymbol{y}, \boldsymbol{y} \rangle \quad \text{porque} \quad \langle \boldsymbol{y}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{R}, \\ &= \|\boldsymbol{y}\|^2\, t^2 + 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle\, t + \|\boldsymbol{x}\|^2 \\ &=: at^2 + bt + c, \end{aligned}$$

la cual es una función cuadrática real de t, con $a > 0$. Como $f(t) \geqslant 0$ para todo t por hipótesis, el discriminante de la ecuación cuadrática $at^2 + bt + c = 0$ no puede ser positivo. (De hecho, si t_1, t_2 fueran dos raíces distintas de esta ecuación, sería $f(t) < 0$ para $t_1 < t < t_2$.) Resulta entonces que $b^2 - 4ac \leqslant 0$, es decir,

$$4\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2 - 4\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \leqslant 0,$$

esto es, $\langle \boldsymbol{x}, \boldsymbol{y} \rangle^2 \leqslant \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2$. Al tomar la raíz cuadrada positiva en ambos lados, se obtiene la desigualdad deseada:

$$\big|\langle \boldsymbol{x}, \boldsymbol{y} \rangle\big| \leqslant \|\boldsymbol{x}\|\, \|\boldsymbol{y}\|,$$

con igualdad solo si $b^2 - 4ac = 0$, es decir, solo si la ecuación cuadrática $at^2 + bt + c = 0$ posee una sola raíz real $t = t_0$. Pero entonces $f(t_0) = \|\boldsymbol{x} + t_0\boldsymbol{y}\|^2 = 0$ y por eso $\boldsymbol{x} + t_0\boldsymbol{y} = \boldsymbol{0}$ en V. Luego $\boldsymbol{x} = -t_0\boldsymbol{y}$ para algún $t_0 \in \mathbb{R}$; es decir, los vectores $\boldsymbol{x}, \boldsymbol{y}$ son proporcionales.

Falta considerar el caso donde $\mathbb{F} = \mathbb{C}$ pero $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \notin \mathbb{R}$. En ese caso, vale $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = r\, e^{i\theta}$ con $r > 0$, $\theta \in \mathbb{R}$. Tómese $\boldsymbol{z} := e^{-i\theta}\boldsymbol{y}$, así que $\langle \boldsymbol{x}, \boldsymbol{z} \rangle = r \in \mathbb{R}$. De (3.3), ya verificado para ese caso, se obtiene $|\langle \boldsymbol{x}, \boldsymbol{z} \rangle| \leqslant \|\boldsymbol{x}\|\, \|\boldsymbol{z}\|$; pero $\|\boldsymbol{z}\| = |e^{-i\theta}|\, \|\boldsymbol{y}\| = \|\boldsymbol{y}\|$, así que

$$\big|\langle \boldsymbol{x}, \boldsymbol{y} \rangle\big| = \big|e^{i\theta}\langle \boldsymbol{x}, \boldsymbol{z} \rangle\big| = \big|\langle \boldsymbol{x}, \boldsymbol{z} \rangle\big| \leqslant \|\boldsymbol{x}\|\, \|\boldsymbol{z}\| = \|\boldsymbol{x}\|\, \|\boldsymbol{y}\|. \qquad \square$$

Corolario 3.8 (Desigualdad de Cauchy)**.** *Si $x_1, \dots, x_n, y_1, \dots, y_n \in \mathbb{R}$, entonces*[5]

$$(x_1y_1 + \cdots + x_ny_n)^2 \leqslant (x_1^2 + \cdots + x_n^2)\,(y_1^2 + \cdots + y_n^2). \qquad \boxminus$$

Proposición 3.9. *La norma de un vector tiene estas propiedades: si $\boldsymbol{x}, \boldsymbol{y} \in V$, $\alpha \in \mathbb{F}$, entonces*

(a) $\|\alpha\boldsymbol{x}\| = |\alpha|\, \|\boldsymbol{x}\|$ *(homogeneidad positiva),*

(b) $\|\boldsymbol{x} + \boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$ *(desigualdad triangular),*

(c) $\|\boldsymbol{x}\| \geqslant 0$, *con igualdad solo si $\boldsymbol{x} = \boldsymbol{0}$ en V.*

[5] La desigualdad (3.3) tiene varios padres. Agustin-Louis Cauchy lo mostró para el caso de n variables reales en 1821; Viktor Bunyakovsky lo mostró para integrales (ver el Ejemplo 3.4) en 1859; Hermann Schwarz mostró el caso general en 1888. Muchos libros hablan de "la desigualdad de Cauchy y Schwarz"; los rusos lo llaman "la desigualdad de Bunyakovsky".

Demostración. Ad (a): Basta notar que $\|\alpha\boldsymbol{x}\|^2 = \langle \alpha\boldsymbol{x}, \alpha\boldsymbol{x}\rangle = (\bar\alpha\alpha)\,\|\boldsymbol{x}\|^2$ y además que $|\alpha| = \sqrt{\bar\alpha\alpha}$ por definición.

Ad (b): Al usar la desigualdad de Schwarz (3.3), se obtiene

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}\rangle^2 \\
&= \langle \boldsymbol{x}, \boldsymbol{x}\rangle^2 + \langle \boldsymbol{x}, \boldsymbol{y}\rangle + \langle \boldsymbol{y}, \boldsymbol{x}\rangle + \langle \boldsymbol{y}, \boldsymbol{y}\rangle^2 \\
&\leqslant \|\boldsymbol{x}\|^2 + 2|\langle \boldsymbol{x}, \boldsymbol{y}\rangle| + \|\boldsymbol{y}\|^2 \\
&\leqslant \|\boldsymbol{x}\|^2 + 2\|\boldsymbol{x}\|\,\|\boldsymbol{y}\| + \|\boldsymbol{y}\|^2 = \big(\|\boldsymbol{x}\| + \|\boldsymbol{y}\|\big)^2.
\end{aligned}$$

Ad (c): Esto es inmediato de la Definición 3.1(d). □

Lema 3.10 (Ley del paralelogramo)**.** *Si V es un espacio vectorial con un producto escalar y si $\boldsymbol{x}, \boldsymbol{y} \in V$, entonces*

$$\boxed{\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 = 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2.}$$

Demostración. Este es un cálculo sencillo:

$$\begin{aligned}
\|\boldsymbol{x}+\boldsymbol{y}\|^2 + \|\boldsymbol{x}-\boldsymbol{y}\|^2 &= \langle \boldsymbol{x}+\boldsymbol{y}, \boldsymbol{x}+\boldsymbol{y}\rangle + \langle \boldsymbol{x}-\boldsymbol{y}, \boldsymbol{x}-\boldsymbol{y}\rangle \\
&= 2\langle \boldsymbol{x}, \boldsymbol{x}\rangle + 2\langle \boldsymbol{y}, \boldsymbol{y}\rangle \\
&= 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2.
\end{aligned}$$ □

Cualquier función $\boldsymbol{x} \mapsto \|\boldsymbol{x}\|$ sobre un espacio vectorial V que cumple las propiedades (a), (b), (c) de la Proposición 3.9 se llama una *norma* sobre V. Resulta que existen otras normas sobre $\mathbb{R}^n$ o $\mathbb{C}^n$, diferentes de (3.2), que no cumplen la ley del paralelogramo. Dos ejemplos de tales normas son

$$\begin{aligned}
\|\boldsymbol{x}\|_1 &:= |x_1| + |x_2| + \cdots + |x_n|, \\
\|\boldsymbol{x}\|_\infty &:= \text{máx}\big\{|x_1|, |x_2|, \dots, |x_n|\big\}.
\end{aligned} \tag{3.4}$$

De hecho, hay un teorema que dice que cualquier norma sobre $\mathbb{R}^n$ o $\mathbb{C}^n$ que cumple la ley del paralelogramo determina un producto escalar tal que $\langle \boldsymbol{x}, \boldsymbol{x}\rangle = \|\boldsymbol{x}\|^2$. Las normas $\|\cdot\|_1$ y $\|\cdot\|_\infty$ no provienen de un producto escalar.

Lema 3.11 (Fórmulas de polarización)**.** *Si V es un espacio vectorial con un producto escalar, la norma* (3.2) *determina el producto escalar por* polarización*:*

$$\textit{Caso } \mathbb{F} = \mathbb{R}: \quad 4\,\langle \boldsymbol{y}, \boldsymbol{x}\rangle = \|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2.$$

$$\textit{Caso } \mathbb{F} = \mathbb{C}: \quad 4\,\langle \boldsymbol{y}, \boldsymbol{x}\rangle = \|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}-\boldsymbol{y}\|^2 + i\,\|i\boldsymbol{x}+\boldsymbol{y}\|^2 - i\,\|i\boldsymbol{x}-\boldsymbol{y}\|^2. \tag{3.5}$$

Demostración. Considérese el caso $\mathbb{F} = \mathbb{C}$; el caso más sencillo $\mathbb{F} = \mathbb{R}$ se deja como ejercicio.

De la fórmula (3.2) que define la norma, se obtiene

$$\|\boldsymbol{x} \pm \boldsymbol{y}\|^2 = \langle \boldsymbol{x} \pm \boldsymbol{y}, \boldsymbol{x} \pm \boldsymbol{y} \rangle = \|\boldsymbol{x}\|^2 \pm \langle \boldsymbol{x}, \boldsymbol{y} \rangle \pm \langle \boldsymbol{y}, \boldsymbol{x} \rangle + \|\boldsymbol{y}\|^2.$$

Por lo tanto,

$$\|\boldsymbol{x} + \boldsymbol{y}\|^2 - \|\boldsymbol{x} - \boldsymbol{y}\|^2 = 2\langle \boldsymbol{x}, \boldsymbol{y} \rangle + 2\langle \boldsymbol{y}, \boldsymbol{x} \rangle.$$

De manera similar, se ve que

$$\|\boldsymbol{x} \pm i\boldsymbol{y}\|^2 = \langle \boldsymbol{x} \pm i\boldsymbol{y}, \boldsymbol{x} \pm i\boldsymbol{y} \rangle = \|\boldsymbol{x}\|^2 \pm i\langle \boldsymbol{x}, \boldsymbol{y} \rangle \mp i\langle \boldsymbol{y}, \boldsymbol{x} \rangle + \|\boldsymbol{y}\|^2,$$

y de ahí se obtiene

$$i\|\boldsymbol{x} + i\boldsymbol{y}\|^2 - i\|\boldsymbol{x} - i\boldsymbol{y}\|^2 = -2\langle \boldsymbol{x}, \boldsymbol{y} \rangle + 2\langle \boldsymbol{y}, \boldsymbol{x} \rangle.$$

La fórmula (3.5) sigue directamente. □

► La desigualdad de Schwarz tiene consecuencias inmediatas para la geometría de un espacio vectorial real.

Definición 3.12. Si V es un espacio vectorial sobre $\mathbb{R}$, la **distancia** entre dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ se define como la longitud de la diferencia, $\|\boldsymbol{x} - \boldsymbol{y}\|$. Si $\boldsymbol{x} \neq \boldsymbol{0}$ y $\boldsymbol{y} \neq \boldsymbol{0}$, el **ángulo** entre estos dos vectores no nulos se define como sigue. La desigualdad de Schwarz implica que

$$-1 \leqslant \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} \leqslant 1,$$

así que hay un único ángulo θ, con $0 \leqslant \theta \leqslant \pi$, que cumple

$$\underline{\cos\theta} := \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|}\,.$$

Dicho de otro modo, se verifica la relación $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|\,\|\boldsymbol{y}\| \cos\theta$. ◊

Definición 3.13. Dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ se llaman **ortogonales** si $\underline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0}$. (Esta terminología es aplicable en los dos casos $\mathbb{F} = \mathbb{R}$ y $\mathbb{F} = \mathbb{C}$.) Se escribe $\underline{\boldsymbol{x} \perp \boldsymbol{y}}$ para significar que $\boldsymbol{x}, \boldsymbol{y}$ son ortogonales.

Dícese que los vectores no nulos $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in V$ forman un **conjunto ortogonal** en V si $\boldsymbol{x}_i \perp \boldsymbol{x}_j$ para $i \neq j$. ◊

Lema 3.14. *Un conjunto ortogonal de vectores en V es linealmente independiente.*

Demostración. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\} \subset V$ un conjunto ortogonal de vectores no nulos. Tómese $c_1, \dots, c_m \in \mathbb{F}$ tales que $c_1\boldsymbol{x}_1 + \cdots + c_m\boldsymbol{x}_m = \boldsymbol{0}$. Para cada índice $j = 1, \dots, m$, se ve que

$$0 = \langle \boldsymbol{x}_j, \boldsymbol{0} \rangle = \langle \boldsymbol{x}_j, c_1\boldsymbol{x}_1 + \cdots + c_m\boldsymbol{x}_m \rangle = \langle \boldsymbol{x}_j, c_j\boldsymbol{x}_j \rangle = c_j \, \|\boldsymbol{x}_j\|^2,$$

y por tanto $c_j = 0$. Luego $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es linealmente independiente en V. □

Definición 3.15. Si M es un subespacio de V, su **complemento ortogonal** $M^\perp$ es el subespacio de V definido por

$$\underline{M^\perp} := \{\, \boldsymbol{y} \in V : \langle \boldsymbol{y}, \boldsymbol{x} \rangle = 0 \text{ para todo } \boldsymbol{x} \in M \,\}.$$

Obsérvese que $M^\perp$ es un *subespacio* de V porque, si $\boldsymbol{y}, \boldsymbol{z} \in M^\perp$, $\alpha \in \mathbb{F}$ y si $\boldsymbol{x} \in M$, entonces

$$\begin{aligned} \langle \boldsymbol{y} + \boldsymbol{z}, \boldsymbol{x} \rangle &= \langle \boldsymbol{y}, \boldsymbol{x} \rangle + \langle \boldsymbol{z}, \boldsymbol{x} \rangle = 0 + 0 = 0, \\ \langle \alpha\boldsymbol{y}, \boldsymbol{x} \rangle &= \bar{\alpha}\langle \boldsymbol{y}, \boldsymbol{x} \rangle = \bar{\alpha}\, 0 = 0. \end{aligned}$$ ◊

3.2. Bases ortonormales

En esta sección, V es un espacio vectorial (sobre $\mathbb{R}$ o $\mathbb{C}$) de dimensión finita n, dotado de un producto escalar $\langle -, - \rangle$.

Definición 3.16. Se dice que un juego de n vectores $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es una **base ortonormal** de V si se verifica las siguientes dos propiedades:

(a) $\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = 0$ cuando $j \neq k$;

(b) $\langle \boldsymbol{e}_k, \boldsymbol{e}_k \rangle = 1$ para $k = 1, \dots, n$.

Alternativamente, estas dos condiciones pueden combinarse en una:

$$\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = [\![j = k]\!] \quad \text{para} \quad j, k = 1, \dots, n. \tag{3.6}$$

En otras palabras, $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es una base ortonormal de V si es un conjunto ortogonal de n vectores tales que $\|\boldsymbol{e}_k\| = 1$ para todo k. (Por el Lema 3.14, una base ortonormal es efectivamente una base de V.) ◊

Si $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es una base ortonormal de V, dos vectores cualesquiera $\boldsymbol{x}, \boldsymbol{y} \in V$ definen combinaciones lineales $\boldsymbol{x} = \sum_{j=1}^n x_j\boldsymbol{e}_j$, $\boldsymbol{y} = \sum_{k=1}^n y_k\boldsymbol{e}_k$ de los vectores básicos. Entonces

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} \rangle &= \left\langle \sum_{j=1}^n x_j\boldsymbol{e}_j, \sum_{k=1}^n y_k\boldsymbol{e}_k \right\rangle = \sum_{j=1}^n \sum_{k=1}^n \langle x_j\boldsymbol{e}_j, y_k\boldsymbol{e}_k \rangle \\ &= \sum_{j=1}^n \sum_{k=1}^n \bar{x}_j y_k \, \langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = \sum_{k=1}^n \bar{x}_k y_k = \bar{x}_1 y_1 + \bar{x}_2 y_2 + \cdots + \bar{x}_n y_n, \end{aligned}$$

porque los términos de la doble suma con $j \neq k$ se anulan. De este modo, se obtiene las formas explícitas (3.1) del Ejemplo 3.2 al usar *coordenadas con respecto a una base ortonormal.*

► Una base ortonormal permite identificar el *espacio vectorial dual* V^* con el *espacio vectorial original* V. En efecto, sea $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ una base ortonormal de V; defínase un juego de formas lineales $\mathcal{F} := \{f_1, \dots, f_n\} \subset V^*$ por

$$\boxed{f_k(\boldsymbol{x}) := \langle \boldsymbol{e}_k, \boldsymbol{x}\rangle} \quad \text{para cada} \quad k = 1, \dots, n.$$

Si $\boldsymbol{x} = \sum_{j=1}^n x_j \boldsymbol{e}_j$, entonces $f_k(\boldsymbol{x}) = x_k$, así que $\mathcal{F}$ es la base de V^* dual a la base $\mathcal{E}$ de V.

Defínase una aplicación *semilineal* $J\colon V \to V^*$ por

$$J\bigl(\textstyle\sum_{k=1}^n y_k \boldsymbol{e}_k\bigr) := \textstyle\sum_{k=1}^n \bar{y}_k f_k.$$

Por un análogo del Escolio 1.14, la aplicación semilineal J está determinada por sus valores sobre una base de V; es decir, basta saber que $\underline{J\boldsymbol{e}_k := f_k}$ para $k = 1, 2, \dots, n$. Entonces

$$J\boldsymbol{y} = \sum_{k=1}^n \bar{y}_k\, f_k, \qquad J\boldsymbol{y}(\boldsymbol{x}) = \sum_{k=1}^n \bar{y}_k\, f_k(\boldsymbol{x}) = \sum_{k=1}^n \bar{y}_k x_k = \langle \boldsymbol{y}, \boldsymbol{x}\rangle. \tag{3.7}$$

Resulta, entonces, que $J\boldsymbol{y}$ es la forma *lineal* $\underline{\boldsymbol{x} \mapsto \langle \boldsymbol{y}, \boldsymbol{x}\rangle}$. El resultado final de este cálculo aclara que J tiene una descripción que *no depende de la base ortonormal* específica $\mathcal{E}$ de V. De este modo, la aplicación J proporciona una identificación "canónica" de V^* con V, en presencia de un producto escalar.

Lema 3.17. *Bajo la identificación* (3.7) *de* V^* *con* V*, el complemento ortogonal* $M^\perp$ *de un subespacio* $M \subseteq V$ *coincide con el* anulador $M^\perp \leqslant V^*$ *de la Definición 1.21.*

En consecuencia, resulta que $\underline{M \oplus M^\perp = V}$.

Demostración. Es cuestión de notar lo siguiente:

$$\begin{aligned} J\boldsymbol{y} \in M^\perp \text{ (anulador)} &\iff J\boldsymbol{y}(\boldsymbol{x}) = 0 \quad \text{para todo } \boldsymbol{x} \in M \\ &\iff \quad \langle \boldsymbol{y}, \boldsymbol{x}\rangle = 0 \quad \text{para todo } \boldsymbol{x} \in M \\ &\iff \boldsymbol{y} \in M^\perp \text{ (complemento ortogonal).} \end{aligned}$$

Entonces vale $\dim M^\perp = \dim V - \dim M$. Como $M \cap M^\perp = \{\boldsymbol{0}\}$ obviamente, y por conteo de dimensiones $\dim(M \oplus M^\perp) = \dim M + \dim M^\perp = \dim V$, se obtiene $M \oplus M^\perp = V$. □

Para comprobar la *existencia* de una base ortonormal para un espacio vectorial con producto escalar, hay un algoritmo que la *construye* a partir de una base cualquiera. Se trata de un proceso iterativo que toma cada vector de la base original y lo proyecta sobre una recta que es ortogonal a cada uno de los vectores anteriores. Este proceso se conoce como el **algoritmo de Gram y Schmidt**.[6]

[6]Este algoritmo aparece en un libro de Laplace, *Théorie Analytique des Probabilités* (Paris, 1816) y una versión modificada aparece en un trabajo de Jørgen Pedersen Gram en 1883. La versión moderna del algoritmo se debe a Erhard Schmidt, un estudiante de Hilbert y autor de notables trabajos sobre las ecuaciones integrales, en: "Zur Theorie der linearen und nichtlinearen Integralgleichungen, I: Entwicklung willkürlicher Funktionen nach Systemen vorgeschriebener", *Mathematische Annalen* **63** (1907), 433–476.

Proposición 3.18. *Sea V un espacio vectorial de dimensión n con un producto escalar, y sea $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base de V. Defínase los vectores $\boldsymbol{e}_1, \dots, \boldsymbol{e}_n \in V$ como sigue:*

$$
\begin{aligned}
& & \boldsymbol{e}_1 &:= \boldsymbol{x}_1/\|\boldsymbol{x}_1\|; \\
\boldsymbol{y}_2 &:= \boldsymbol{x}_2 - \langle \boldsymbol{e}_1, \boldsymbol{x}_2\rangle\, \boldsymbol{e}_1, & \boldsymbol{e}_2 &:= \boldsymbol{y}_2/\|\boldsymbol{y}_2\|; \\
\boldsymbol{y}_3 &:= \boldsymbol{x}_3 - \langle \boldsymbol{e}_1, \boldsymbol{x}_3\rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_3\rangle\, \boldsymbol{e}_2, & \boldsymbol{e}_3 &:= \boldsymbol{y}_3/\|\boldsymbol{y}_3\|; \\
&\ \vdots & &\ \vdots \\
\boldsymbol{y}_k &:= \boldsymbol{x}_k - \langle \boldsymbol{e}_1, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_2 - \cdots - \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_{k-1}, & \boldsymbol{e}_k &:= \boldsymbol{y}_k/\|\boldsymbol{y}_k\|. \\
&\ \vdots & &\ \vdots \\
\boldsymbol{y}_n &:= \boldsymbol{x}_n - \langle \boldsymbol{e}_1, \boldsymbol{x}_n\rangle\, \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_n\rangle\, \boldsymbol{e}_2 - \cdots - \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_n\rangle\, \boldsymbol{e}_{n-1}, & \boldsymbol{e}_n &:= \boldsymbol{y}_n/\|\boldsymbol{y}_n\|.
\end{aligned}
\tag{3.8}
$$

Entonces $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_k\}$ es una base ortonormal del subespacio $\operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\rangle \leqslant V$; y para $k = n$, el conjunto $\mathcal{E} := \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es una base ortonormal de V mismo.

Demostración. Por inducción sobre k. Fíjese que $\boldsymbol{x}_1 \neq \boldsymbol{0}$ (por la independencia lineal de $\mathcal{B}$), así que $\|\boldsymbol{x}_1\| \neq 0$. Luego $\{\boldsymbol{e}_1\}$ es una base ortonormal de $\operatorname{lin}\langle \boldsymbol{x}_1\rangle = \{\, \alpha \boldsymbol{x}_1 : \alpha \in \mathbb{F}\,\}$. En efecto, $\boldsymbol{e}_1$ es un múltiplo de $\boldsymbol{x}_1$ tal que $\|\boldsymbol{e}_1\| = 1$.

Supóngase entonces que se ha elegido $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}$ por el procedimiento indicado, y que ellos forman una base ortonormal de $\operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}\rangle$. Para que $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_k\}$ sea base ortonormal de $\operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_k\rangle$, basta comprobar que $\boldsymbol{e}_k \perp \boldsymbol{e}_j$ para $j < k$ y que $\|\boldsymbol{e}_k\| = 1$. Si $j < k$, entonces

$$
\begin{aligned}
\langle \boldsymbol{e}_j, \boldsymbol{y}_k\rangle = \Big\langle \boldsymbol{e}_j, \boldsymbol{x}_k - \sum_{i=1}^{k-1} \langle \boldsymbol{e}_i, \boldsymbol{x}_k\rangle\, \boldsymbol{e}_i \Big\rangle &= \langle \boldsymbol{e}_j, \boldsymbol{x}_k\rangle - \sum_{i=1}^{k-1} \langle \boldsymbol{e}_i, \boldsymbol{x}_k\rangle\, \langle \boldsymbol{e}_j, \boldsymbol{e}_i\rangle \\
&= \langle \boldsymbol{e}_j, \boldsymbol{x}_k\rangle - \sum_{i=1}^{k-1} \langle \boldsymbol{e}_i, \boldsymbol{x}_k\rangle\, [\![j = i]\!] \\
&= \langle \boldsymbol{e}_j, \boldsymbol{x}_k\rangle - \langle \boldsymbol{e}_j, \boldsymbol{x}_k\rangle = 0.
\end{aligned}
$$

En consecuencia, cualquier múltiplo de $\boldsymbol{y}_k$ es también ortogonal a $\boldsymbol{e}_j$ cuando $j < k$. Basta entonces comprobar que $\boldsymbol{y}_k \neq \boldsymbol{0}$, para que se pueda dividir $\boldsymbol{y}_k$ por el número positivo $\|\boldsymbol{y}_k\|$ y así definir $\boldsymbol{e}_k$ como un múltiplo de $\boldsymbol{y}_k$ que cumple $\|\boldsymbol{e}_k\| = 1$.

Ahora, si fuera $\boldsymbol{y}_k = \boldsymbol{0}$, $\boldsymbol{x}_k$ sería una combinación lineal de $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}$ debido a (3.8); y estos vectores $\boldsymbol{e}_j$ son a su vez combinaciones lineales de $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{k-1}$. Pero en tal caso los vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ serían linealmente *dependientes*, lo cual es falso porque $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V.

Se concluye que $\boldsymbol{y}_k \neq \boldsymbol{0}$. Luego $\boldsymbol{e}_k$ está bien definida; vale $\|\boldsymbol{e}_k\| = 1$; y $\boldsymbol{e}_k$ es ortogonal a $\operatorname{lin}\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}\rangle$. □

Corolario 3.19. *Una base ortonormal* $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\}$ *para un subespacio* $W \leqslant V$ *puede ser completado para obtener así una base ortonormal de* V.

Demostración. Los vectores $\boldsymbol{e}_1, \dots, \boldsymbol{e}_m$ son linealmente independientes y generan el subespacio W. Luego, por el Escolio 1.9, es posible hallar otros vectores $\boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n \in V$ tales que $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m, \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n\}$ sea una base de V (no necesariamente ortonormal). Ahora aplíquese el algoritmo de Gram y Schmidt a esta base: los primeros m vectores no sufren cambio alguno y el resultado es una base ortonormal $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de V cuyos primeros m elementos son los vectores $\boldsymbol{e}_1, \dots, \boldsymbol{e}_m$ originales. □

3.3. Aplicaciones transpuestas y adjuntas

La **transpuesta** de una aplicación lineal $T \colon V \to W$, según la Definición 1.18, es la aplicación lineal $T^{\mathrm{t}} \colon W^* \to V^*$ dada por $T^{\mathrm{t}} g := g \circ T$, para $g \in W^*$. De manera más explícita: $T^{\mathrm{t}} g(\boldsymbol{x}) := g(T\boldsymbol{x})$, cuando $g \in W^*$ y $\boldsymbol{x} \in V$.

Si V y W son espacios vectoriales *reales*, dotados de productos escalares, la fórmula (3.7) proporciona dos identificaciones canónicas $J \colon V \to V^*$ y $J' \colon W \to W^*$ dadas por

$$\begin{aligned} Jz &: \boldsymbol{x} \mapsto \langle z, \boldsymbol{x} \rangle_V \quad \text{para} \quad \boldsymbol{x} \in V, \\ J'\boldsymbol{y} &: \boldsymbol{y} \mapsto \langle \boldsymbol{y}, \boldsymbol{w} \rangle_W \quad \text{para} \quad \boldsymbol{w} \in W. \end{aligned}$$

Estos isomorfismos $\mathbb{R}$-lineales permiten reemplazar $T^{\mathrm{t}} \colon W^* \to V^*$ por una aplicación lineal de W en V. En efecto, se ve que

$$\langle J^{-1} T^{\mathrm{t}} J' \boldsymbol{y}, \boldsymbol{x} \rangle_V = T^{\mathrm{t}} J' \boldsymbol{y}(\boldsymbol{x}) = J' \boldsymbol{y}(T\boldsymbol{x}) = \langle \boldsymbol{y}, T\boldsymbol{x} \rangle_W \,. \tag{3.9}$$

En este cálculo se ha puesto unas etiquetas $_{V,W}$ a los productos escalares, para énfasis; es usual omitirlas cuando el contexto deja claro a cuál de los dos productos escalares se refiere.

Definición 3.20. Dados productos escalares para los espacios $\mathbb{R}$-vectoriales V y W, se identifica la aplicación $T^{\mathrm{t}} \colon W^* \to V^*$ con la aplicación $J^{-1} T^{\mathrm{t}} J' \colon W \to V$. De este modo se *redefine* la transpuesta de $T \colon V \to W$ como la aplicación $T^{\mathrm{t}} \colon W \to V$ determinada por la fórmula:

$$\boxed{\langle T^{\mathrm{t}} \boldsymbol{y}, \boldsymbol{x} \rangle := \langle \boldsymbol{y}, T\boldsymbol{x} \rangle.} \tag{3.10}$$

Dicho de otra manera: dada $T \colon V \to W$, hay exactamente una aplicación $\mathbb{R}$-lineal $T^{\mathrm{t}} \colon W \to V$ que hace cumplir (3.10) para todo $\boldsymbol{x} \in V$, $\boldsymbol{y} \in W$. ◊

Para el caso $V = \mathbb{R}^n$ y $W = \mathbb{R}^m$, es más cómodo (y más usual) emplear la notación del *producto punto* en vez de los corchetes angulares para denotar productos escalares. Los elementos del espacio vectorial "original" $\mathbb{R}^n$ son *vectores de columna*, considerados como matrices $n \times 1$. El espacio vectorial dual $(\mathbb{R}^n)^*$ se identifica con los *vectores de fila*, es decir, los matrices $1 \times n$. De ahora en adelante *se escribirá* $\underline{\boldsymbol{x}^{\mathrm{t}}}$ *para denotar el vector de fila* que es la transpuesta del vector de columna $\boldsymbol{x}$. El producto escalar se convierte en $\underline{\boldsymbol{x}^{\mathrm{t}}\,\boldsymbol{y}} = \boldsymbol{x} \cdot \boldsymbol{y}$. La simetría del producto escalar real se expresa así:

$$\boldsymbol{x}^{\mathrm{t}}\,\boldsymbol{y} = \boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{y} \cdot \boldsymbol{x} = \boldsymbol{y}^{\mathrm{t}}\,\boldsymbol{x}.$$

La *base estándar* de $\mathbb{R}^n$ es una base ortonormal con respecto al producto punto. Si $A \in \mathbb{R}^{m\times n}$ es la matriz de la aplicación lineal $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ con respecto a las bases estándares, la fórmula (3.10) que define la aplicación transpuesta es entonces equivalente a la fórmula matricial:

$$\underline{\boldsymbol{x} \cdot A^{\mathrm{t}}\boldsymbol{y}} = \boldsymbol{x}^{\mathrm{t}}A^{\mathrm{t}}\boldsymbol{y} = (A^{\mathrm{t}}\boldsymbol{y})^{\mathrm{t}}\,\boldsymbol{x} = \boldsymbol{y}^{\mathrm{t}}A\boldsymbol{x} = \underline{\boldsymbol{y} \cdot A\boldsymbol{x}}\,. \tag{3.11}$$

► Para espacios vectoriales *complejos* (el caso $\mathbb{F} = \mathbb{C}$), se debe tomar en cuenta que las identificaciones canónicas $J : V \to V^*$ y $J' : W \to W^*$ no son $\mathbb{C}$-lineales, sino solo *semilineales*. Esta circunstancia modifica la aplicación transpuesta en el caso complejo.

Definición 3.21. Sean V, W espacios vectoriales sobre $\mathbb{C}$. Si $S : V \to W$ es $\mathbb{C}$-lineal, se define la **aplicación adjunta** $S^* : W \to V$ por $\underline{S^*} := J^{-1}S^{\mathrm{t}}J' \in \mathcal{L}(W, V)$. (Debido a la semilinealidad de J y J', la identificación de la Definición 3.20 ya no es aplicable.)

Si $\boldsymbol{y} \in W$ y $\boldsymbol{x} \in V$, la transferencia (3.9) funciona del mismo modo que el caso real:

$$\langle J^{-1}S^{\mathrm{t}}J'\boldsymbol{y}, \boldsymbol{x}\rangle_V = S^{\mathrm{t}}J'\boldsymbol{y}(\boldsymbol{x}) = J'\boldsymbol{y}(S\boldsymbol{x}) = \langle \boldsymbol{y}, S\boldsymbol{x}\rangle_W,$$

Entonces la aplicación adjunta $S^* : W \to V$ queda determinada por la fórmula clave:

$$\boxed{\langle S^*\boldsymbol{y}, \boldsymbol{x}\rangle := \langle \boldsymbol{y}, S\boldsymbol{x}\rangle.} \tag{3.12}$$

Al tomar el conjugado complejo de ambos lados, se obtiene una fórmula equivalente: $\langle S\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, S^*\boldsymbol{y}\rangle$. En consecuencia, se puede mover una aplicación lineal de un lado a otro de un producto escalar, reemplazándola por su aplicación adjunta al otro lado. ◊

► Sea $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ una base ortonormal para V y $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_m\}$ una base ortonormal para W. Una aplicación lineal $S : V \to W$ y su adjunta $S^* : W \to V$ determinan dos matrices $A = [S]^{\mathcal{U}}_{\mathcal{E}}$ y $B = [S^*]^{\mathcal{E}}_{\mathcal{U}}$ con respecto a estas bases. De acuerdo con la fórmula (1.9), se obtiene

$$S\boldsymbol{e}_j = \sum_{l=1}^{m} a_{lj}\,\boldsymbol{u}_l, \quad \text{así que} \quad a_{lj} = \langle \boldsymbol{u}_l, S\boldsymbol{e}_j\rangle \quad \text{para } j = 1, \dots, n,$$

$$S^*\boldsymbol{u}_k = \sum_{i=1}^{n} b_{ik}\,\boldsymbol{e}_i, \quad \text{así que} \quad b_{ik} = \langle \boldsymbol{e}_i, S^*\boldsymbol{u}_k\rangle \quad \text{para } k = 1, \dots, m.$$

Por lo tanto,

$$b_{ik} := \langle \boldsymbol{e}_i, S^*\boldsymbol{u}_k\rangle = \langle S\boldsymbol{e}_i, \boldsymbol{u}_k\rangle = \overline{\langle \boldsymbol{u}_k, S\boldsymbol{e}_i\rangle} = \overline{a_{ki}}\,.$$

En resumen: la matriz $B \in \mathbb{C}^{n\times m}$ es la *transpuesta conjugada* de la matriz $A \in \mathbb{C}^{m\times n}$.

Definición 3.22. Sea $A = [a_{ij}]$ una matriz compleja $m \times n$. Se denota por $\overline{A} := [\bar{a}_{ij}]$ su **matriz conjugada**, también $m \times n$, obtenida de A al tomar el conjugado complejo de cada uno de sus elementos. Obsérvese que $\overline{A} = A$ si y solo si todas las entradas de A son reales.

La **matriz adjunta** de A es la matriz $n \times m$ dada por $\underline{A^* := [\bar{a}_{ji}]}$; esta es la matriz transpuesta de $\overline{A}$ y a la vez la matriz conjugada de A^{t}:

$$\boxed{A^* := (\overline{A})^{\mathsf{t}} = \overline{A^{\mathsf{t}}}.} \tag{3.13}$$

Las transformaciones $A \mapsto \overline{A}$ y $A \mapsto A^*$ son semilineales. La primera conserva el orden de multiplicación: $\overline{AB} = \overline{A}\,\overline{B}$, mientras la segunda revierte el orden:

$$(AB)^* = B^* A^*$$

en general. ⟦ Si $A \in \mathbb{C}^{m\times n}$ y $B \in \mathbb{C}^{n\times k}$, entonces $B^* \in \mathbb{C}^{k\times n}$ y $A^* \in \mathbb{C}^{n\times m}$, así que estos productos de matrices son consistentes. ⟧ ◊

Definición 3.23. Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada sobre un cuerpo cualquiera. Dícese que A es una **matriz simétrica** si $\underline{A^{\mathsf{t}} = A}$.

En el caso $\mathbb{F} = \mathbb{C}$, una matriz cuadrada $A \in M_n(\mathbb{C})$ es una **matriz hermítica** si $\underline{A^* = A}$. Una matriz simétrica real es también hermítica.[7] ◊

Ejemplo 3.24. Si $A \in \mathbb{R}^{m\times n}$, las matrices cuadradas $A^{\mathsf{t}}A \in M_n(\mathbb{R})$ y $AA^{\mathsf{t}} \in M_m(\mathbb{R})$ son simétricas. En efecto,

$$(A^{\mathsf{t}}A)^{\mathsf{t}} = A^{\mathsf{t}}A^{\mathsf{tt}} = A^{\mathsf{t}}A, \qquad (AA^{\mathsf{t}})^{\mathsf{t}} = A^{\mathsf{tt}}A^{\mathsf{t}} = AA^{\mathsf{t}}.$$

De igual modo, si $B \in \mathbb{C}^{m\times n}$, la matrices cuadradas $B^*B \in M_n(\mathbb{C})$ y $BB^* \in M_m(\mathbb{C})$ son hermíticas. De hecho,

$$(B^*B)^* = B^*B^{**} = B^*B, \qquad (BB^*)^* = B^{**}B^* = BB^*. \qquad ◊$$

Proposición 3.25. *Si $A \in M_n(\mathbb{R})$ es una matriz real simétrica, o bien si $A \in M_n(\mathbb{C})$ es una matriz hermítica, entonces* los autovalores de A son todos reales. *Además, si dos autovectores $\boldsymbol{x}$, $\boldsymbol{y}$ de A corresponden a autovalores distintos, entonces $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$.*

Si A posee n autovalores distintos $\{\lambda_1, \dots, \lambda_n\}$, *sean $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ unos autovectores de A tales que $A\boldsymbol{x}_j = \lambda_j \boldsymbol{x}_j$ y además $\|\boldsymbol{x}_j\| = 1$ para $j = 1, \dots, n$. Estos autovectores $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ entonces forman una base ortonormal de $\mathbb{R}^n$ o $\mathbb{C}^n$, según el caso.*

Demostración. Es suficiente considerar el caso $\mathbb{F} = \mathbb{C}$, con $A = A^* \in M_n(\mathbb{C})$. Esto implica que $\langle A\boldsymbol{y}, \boldsymbol{x} \rangle = \langle \boldsymbol{y}, A\boldsymbol{x} \rangle$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$.

[7] La palabra *hermítica* celebra el matemático francés *Charles Hermite* quien fue el primero que mostró, en 1855, que una matriz compleja A que satisface $A = A^*$ tiene autovalores reales.

Si $\boldsymbol{x} \in \mathbb{C}^n$ es un autovector para el autovalor λ de A, entonces

$$\begin{aligned}\lambda\,\|\boldsymbol{x}\|^2 &= \lambda\,\langle \boldsymbol{x},\boldsymbol{x}\rangle = \langle \boldsymbol{x},\lambda\boldsymbol{x}\rangle \\ &= \langle \boldsymbol{x},A\boldsymbol{x}\rangle = \langle A\boldsymbol{x},\boldsymbol{x}\rangle \\ &= \langle \lambda\boldsymbol{x},\boldsymbol{x}\rangle = \bar\lambda\,\langle \boldsymbol{x},\boldsymbol{x}\rangle = \bar\lambda\,\|\boldsymbol{x}\|^2,\end{aligned}$$

así que $\bar\lambda = \lambda$ porque $\boldsymbol{x} \neq \boldsymbol{0}$ (y por ende $\|\boldsymbol{x}\| > 0$) por ser $\boldsymbol{x}$ un autovector. Esto comprueba que $\lambda \in \mathbb{R}$.

Ahora sean $A\boldsymbol{x} = \lambda\,\boldsymbol{x}$, $A\boldsymbol{y} = \mu\,\boldsymbol{y}$ con $\lambda \neq \mu$. Entonces

$$\begin{aligned}\mu\,\langle \boldsymbol{x},\boldsymbol{y}\rangle &= \langle \boldsymbol{x},\mu\boldsymbol{y}\rangle = \langle \boldsymbol{x},A\boldsymbol{y}\rangle \\ &= \langle A\boldsymbol{x},\boldsymbol{y}\rangle = \langle \lambda\boldsymbol{x},\boldsymbol{y}\rangle = \lambda\,\langle \boldsymbol{x},\boldsymbol{y}\rangle, \quad \text{pues} \quad \lambda \in \mathbb{R}.\end{aligned}$$

Como $\mu \neq \lambda$, se concluye que $\langle \boldsymbol{x},\boldsymbol{y}\rangle = 0$.

Si A posee n autovalores distintos, hay autovectores $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ tales que $A\boldsymbol{x}_j = \lambda_j\boldsymbol{x}_j$ para $j = 1,\dots,n$. Estos vectores no nulos son ortogonales entre sí, por lo que se acaba de probar. Al multiplicar cada $\boldsymbol{x}_j$ por un escalar positiva si fuera necesario, se puede suponer que $\|\boldsymbol{x}_j\| = 1$ para cada j. Entonces $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ consta de n vectores mutuamente ortogonales de norma 1, y por lo tanto es una base ortonormal de $\mathbb{C}^n$. □

Proposición 3.26. *Cada matriz hermítica $A = A^* \in M_n(\mathbb{C})$ es diagonalizable. De hecho, A posee una base ortonormal de autovectores.*

Demostración. De acuerdo con la Proposición 2.45 y la observación de que cualquie polinomio complejo escinde, para mostrar que $A = A^*$ es diagonalizable basta comprobar que su polinomio mínimo no tiene factores repetidos.

La descomposición primaria de $V = \mathbb{C}^n$ que corresponde al operador T_A dice que hay una suma directa de subespacios invariantes $\mathbb{C}^n = U_1 \oplus \cdots \oplus U_r$, etiquetadas con los autovalores distintos $\alpha_1,\dots,\alpha_r$ de A. Este es el Corolario 2.44; las fórmulas (2.24) y (2.25) para los polinomios característico y mínimo para $T = T_A$ toman las formas

$$\begin{aligned}p_A(t) &= (t-\alpha_1)^{k_1}(t-\alpha_2)^{k_2}\cdots(t-\alpha_r)^{k_r}, \\ q_A(t) &= (t-\alpha_1)^{l_1}(t-\alpha_2)^{l_2}\cdots(t-\alpha_r)^{l_r},\end{aligned}$$

cada cada $\alpha_i \in \mathbb{R}$; y los subespacios invariantes son $U_i := \{\,\boldsymbol{x} \in \mathbb{C}^n : (A-\alpha_i\,1_n)^{l_i}\boldsymbol{x} = \boldsymbol{0}\,\}$.

Tómese $\boldsymbol{x} \in U_i$. Hay un único $m \in \mathbb{N}$ tal que $2^{m-1} < l_i \leqslant 2^m$, entonces también vale $(A-\alpha_i\,1_n)^{2^m}\boldsymbol{x} = \boldsymbol{0}$. Como $\alpha_i \in \mathbb{R}$, la matriz $(A-\alpha_i\,1_n)$ es también hermítica. Por lo tanto,

$$\begin{aligned}0 = \langle \boldsymbol{x},\boldsymbol{0}\rangle = \langle \boldsymbol{x},(A-\alpha_i\,1_n)^{2^m}\boldsymbol{x}\rangle &= \langle (A-\alpha_i\,1_n)^{2^{m-1}}\boldsymbol{x},(A-\alpha_i\,1_n)^{2^{m-1}}\boldsymbol{x}\rangle \\ &= \left\|(A-\alpha_i\,1_n)^{2^{m-1}}\boldsymbol{x}\right\|^2.\end{aligned}$$

Esto implica que $(A-\alpha_i\,1_n)^{2^{m-1}}\boldsymbol{x} = \boldsymbol{0}$ en $\mathbb{C}^n$. Al repetir este argumento m veces, se deduce que $(A-\alpha_i\,1_n)\boldsymbol{x} = \boldsymbol{0}$ para cada $\boldsymbol{x} \in U_i$.

Esto dice que $l_i = 1$ y que todo vector no nulo en U_i es un *autovector* de A.

Para terminar, en la descomposición $\mathbb{C}^n = U_1 \oplus \cdots \oplus U_r$, la Proposición 3.25 anterior muestra que estos subespacios son *mutuamente ortogonales*: $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$ para $\boldsymbol{x} \in U_i$, $\boldsymbol{y} \in U_j$ con $i \neq j$. En resumen, $U_1 \oplus \cdots \oplus U_r$ es una *suma directa ortogonal*. Se puede encontrar una base ortonormal en cada subespacio U_i (si fuere necesario, tómese una base cualquiera de U_i y modifíquelo con el algoritmo de Gram y Schmidt). Su unión es una base ortonormal de $\mathbb{C}^n$, formado por autovectores de A. $\square$

3.4. Matrices ortogonales, unitarias y positivas

La ubicuidad de bases ortonormales en la teoría de las matrices simétricas reales y las matrices hermíticas complejas justifica la introducción de las siguientes dos clases de matrices.

Definición 3.27. (a) Una matriz $Q \in M_n(\mathbb{R})$ es una **matriz ortogonal** si sus columnas forman una base ortonormal de $\mathbb{R}^n$.

(b) Una matriz $U \in M_n(\mathbb{C})$ es una **matriz unitaria** si sus columnas forman una base ortonormal de $\mathbb{C}^n$. $\Diamond$

Proposición 3.28. (a) *Una matriz $Q \in M_n(\mathbb{R})$ es ortogonal si y solo si* $\underline{Q^{\mathrm{t}} Q = 1_n}$.

(b) *Una matriz $U \in M_n(\mathbb{C})$ es unitaria si y solo si* $\underline{U^* U = 1_n}$.

Demostración. Es suficiente demostrar la parte (b), porque una matriz *real* es ortogonal si y solo si es unitaria.

Ad (b): Sean $\boldsymbol{u}_1, \dots, \boldsymbol{u}_n$ las columnas de la matriz $U \in M_n(\mathbb{C})$. Por definición, U es unitaria si y solo si $\langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle = 1$ para $i = 1, \dots, n$ y además $\langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle = 0$ para $i \neq j$.

La entrada (i, j) de la matriz $M = U^*U$ cumple

$$m_{ij} = \sum_{k=1}^{n} \overline{u_{ki}}\, u_{kj} = \langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle,$$

así que U es unitaria si y solo si $m_{ij} = [\![i = j]\!]$ para todo i, j; si y solo si $M = 1_n$. $\square$

Corolario 3.29. (a) *$Q \in M_n(\mathbb{R})$ es ortogonal si y solo si Q es invertible, con $Q^{-1} = Q^{\mathrm{t}}$.*

(b) *$U \in M_n(\mathbb{C})$ es unitaria si y solo si U es invertible, con $U^{-1} = U^*$.*

Demostración. Ad (a): Si $Q \in M_n(\mathbb{R})$ es ortogonal, entonces $r(Q) = n$ porque sus n columnas son linealmente independientes. Por lo tanto, Q es invertible. La fórmula $Q^{\mathrm{t}} Q = 1_n$ dice que Q^{-1} es necesariamente igual a Q^{t}; en consecuencia, se ve que $\underline{Q\, Q^{\mathrm{t}} = 1_n}$ también.

A la inversa, si Q es invertible con $Q^{-1} = Q^{\mathrm{t}}$, entonces $Q^{\mathrm{t}} Q = 1_n = Q\, Q^{\mathrm{t}}$ y la Proposición 3.28(a) muestra que Q es ortogonal.

Ad (b): Si $U \in M_n(\mathbb{C})$ es unitaria, entonces $r(U) = n$ porque sus n columnas son linealmente independientes. Por lo tanto, U es invertible. La fórmula $U^* U = 1_n$ dice que U^{-1} es necesariamente igual a U^* ; en consecuencia, se ve que $U U^* = 1_n$ también.

A la inversa, si U es invertible con $U^{-1} = U^*$, entonces $U^* U = 1_n = U U^*$ y la Proposición 3.28(b) muestra que U es unitaria. □

► La parte (a) del Teorema 1.23 de rango y nulidad, dice que $r(T^{\mathrm{t}}) = r(T)$ para una aplicación lineal T cualquiera. Dada una matriz real $B \in \mathbb{R}^{m\times n}$, tómese $T = T_B : \boldsymbol{x} \mapsto B\boldsymbol{x}$ para obtener $r(B^{\mathrm{t}}) = r(B)$.

En el caso de una aplicación adjunta con $\mathbb{F} = \mathbb{C}$, según la Definición 3.21, $S^* := J^{-1}S^{\mathrm{t}}J'$ donde J y J' son semilineales (en particular, $\mathbb{R}$-lineales) y biyectivas. La composición con una aplicación $\mathbb{R}$-lineal y biyectiva conserva la independencia lineal de vectores, lo cual implica que $r(J^{-1}S^{\mathrm{t}}J') = r(S^{\mathrm{t}}J') = r(S^{\mathrm{t}})$. Se concluye que $r(S^*) = r(S^{\mathrm{t}}) = r(S)$.

Lema 3.30. *Si $A \in \mathbb{C}^{m\times n}$, entonces $r(A) = r(\overline{A}) = r(A^*) = r(A^{\mathrm{t}})$.*

Demostración. Cualquier relación de dependencia lineal entre las columnas de $\overline{A}$,

$$c_1\, \bar{\boldsymbol{a}}_1 + c_2\, \bar{\boldsymbol{a}}_2 + \cdots + c_n\, \bar{\boldsymbol{a}}_n = \boldsymbol{0},$$

es el conjugado complejo de otra relación de dependencia lineal entre las columnas de A:

$$\bar{c}_1\, \boldsymbol{a}_1 + \bar{c}_2\, \boldsymbol{a}_2 + \cdots + \bar{c}_n\, \boldsymbol{a}_n = \boldsymbol{0}.$$

Por lo tanto, las matrices $\overline{A}$ y A tienen el mismo número de columnas linealmente independientes; este número es $r(\overline{A}) = r(A)$.

La igualdad $r(A^{\mathrm{t}}) = r(A)$ viene del teorema de rango y nulidad, al usar $T = T_A$. Además, como $A^* = \overline{A^{\mathrm{t}}}$ en $\mathbb{C}^{n\times m}$ por (3.13), se obtiene también $r(A^*) = r(A^{\mathrm{t}})$. □

Resulta que las matrices $A^{\mathrm{t}}A$ o AA^{t} (en el caso real) o bien A^*A o AA^* (en el caso complejo) también tienen el mismo rango que la matriz A, como se verá a continuación.

Proposición 3.31. *Si $A \in \mathbb{C}^{m\times n}$, entonces $r(A^*A) = r(A)$.*

Demostración. Si $\boldsymbol{x} \in \mathbb{C}^n$, entonces $A\boldsymbol{x} = \boldsymbol{0} \implies A^*A\boldsymbol{x} = \boldsymbol{0}$; e inversamente,

$$A^*A\boldsymbol{x} = \boldsymbol{0} \implies \langle \boldsymbol{x}, A^*A\boldsymbol{x}\rangle = 0 \implies \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = 0 \implies A\boldsymbol{x} = \boldsymbol{0} \tag{3.14}$$

porque $\langle \boldsymbol{x}, A^*A\boldsymbol{x}\rangle = \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle$ por la fórmula (3.12).

De ahí se concluye que $A\boldsymbol{x} = \boldsymbol{0}$ en $\mathbb{C}^m$ si y solo si $A^*A\boldsymbol{x} = \boldsymbol{0}$ en $\mathbb{C}^n$. Fíjese que los operadores $T_A : \mathbb{C}^n \to \mathbb{C}^m$ y $T_{A^*A} : \mathbb{C}^n \to \mathbb{C}^n$ tienen el mismo dominio $\mathbb{C}^n$. La relación (3.14) dice que los núcleos $\ker T_A$ y $\ker T_{A^*A}$ de $\mathbb{C}^n$ coinciden; así que estos operadores tienen la misma nulidad: $n(A) = n(A^*A)$.

Del teorema de rango y nulidad, se obtiene $r(A^*A) = n - n(A^*A) = n - n(A) = r(A)$. □

Corolario 3.32. *Si $A \in \mathbb{C}^{m\times n}$, entonces $\underline{r(AA^*) = r(A)}$.*

Demostración. Al aplicar la Proposición 3.31 a la matriz compleja A^* en vez de A, se obtiene $r(AA^*) = r(A^*)$. Del Lema 3.30 sigue $r(AA^*) = r(A^*) = r(A)$. □

Escolio 3.33. *Si $A \in \mathbb{R}^{m\times n}$, entonces $r(A^{\mathrm{t}}A) = r(A) = r(AA^{\mathrm{t}})$.* ⊟

Lema 3.34. *Una matriz hermítica $A = A^* \in M_n(\mathbb{C})$ es nula (esto es, $A = 0$) si y solo si $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = 0$ para todo $\boldsymbol{x} \in \mathbb{C}^n$.*

Demostración. Fíjese primero que cualquier $A \in M_n(\mathbb{C})$ es nula si y solo si $\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = 0$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$. De hecho, el caso particular $\boldsymbol{y} = A\boldsymbol{x}$ de esta condición dice que $\langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = 0$ y por ende $A\boldsymbol{x} = \boldsymbol{0}$ para todo $\boldsymbol{x} \in \mathbb{C}^n$, lo cual implica que $A = 0$.

Ahora sea A una matriz hermítica y supóngase que $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = 0$ para todo $\boldsymbol{x}$. Entonces, para cada par de vectores $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$, resulta que

$$\langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle \boldsymbol{x}, A\boldsymbol{y}\rangle = \langle (\boldsymbol{x}+\boldsymbol{y}), A(\boldsymbol{x}+\boldsymbol{y})\rangle - \langle \boldsymbol{x}, A\boldsymbol{x}\rangle - \langle \boldsymbol{y}, A\boldsymbol{y}\rangle = 0.$$

Por ser $A = A^*$, la fórmula (3.12). convierte esto en

$$2\Re\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = \langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle A\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle \boldsymbol{x}, A\boldsymbol{y}\rangle = 0. \tag{3.15}$$

Esto dice que $\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = 0$ siempre y cuando $\langle \boldsymbol{y}, A\boldsymbol{x}\rangle$ es *real*.

Aun cuando $\langle \boldsymbol{y}, A\boldsymbol{x}\rangle$ no sea real, es posible escribir $\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = re^{i\theta} \in \mathbb{C}$. En seguida,

$$|\langle \boldsymbol{y}, A\boldsymbol{x}\rangle| = r = e^{-i\theta}\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = \langle e^{i\theta}\boldsymbol{y}, A\boldsymbol{x}\rangle.$$

Pero ahora $\langle e^{i\theta}\boldsymbol{y}, A\boldsymbol{x}\rangle$ sí es real (y no negativo); al aplicar la relación (3.15) con $e^{i\theta}\boldsymbol{y}$ en lugar de $\boldsymbol{y}$, se deduce que $|\langle \boldsymbol{y}, A\boldsymbol{x}\rangle| = 0$ y por ende $\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = 0$ para dos vectores $\boldsymbol{x}, \boldsymbol{y}$ cualesquiera. Luego $A = 0$. □

Proposición 3.35. *Una matriz $A \in M_n(\mathbb{C})$ es hermítica si y solo si $\underline{\langle \boldsymbol{x}, A\boldsymbol{x}\rangle \in \mathbb{R}}$ para todo $\boldsymbol{x}$.*

Demostración. Si $A = A^*$, entonces $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle$ es real para cada $\boldsymbol{x} \in \mathbb{C}^n$, porque

$$\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \langle A^*\boldsymbol{x}, \boldsymbol{x}\rangle = \langle A\boldsymbol{x}, \boldsymbol{x}\rangle = \overline{\langle \boldsymbol{x}, A\boldsymbol{x}\rangle}.$$

Inversamente, si $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle \in \mathbb{R}$ para todo $\boldsymbol{x} \in \mathbb{C}^n$, entonces

$$\langle \boldsymbol{x}, (A - A^*)\boldsymbol{x}\rangle = \langle \boldsymbol{x}, A\boldsymbol{x}\rangle - \langle \boldsymbol{x}, A^*\boldsymbol{x}\rangle = \langle \boldsymbol{x}, A\boldsymbol{x}\rangle - \langle A\boldsymbol{x}, \boldsymbol{x}\rangle = 0 \quad \text{para todo } \boldsymbol{x}.$$

Por el Lema 3.34 anterior, se concluye que $i(A - A^*) = 0$ en $M_n(\mathbb{C})$, así que $A^* = A$. □

► El concepto de *positividad* para matrices cuadradas reales o complejos. En primer lugar, dícese que una matriz $A \in M_n(\mathbb{R})$ es **no negativa** si todas sus entradas son números no negativos: $a_{ij} \geqslant 0$ para cada $i, j = 1, \dots, n$. Este es ciertamente una clasificación útil: por ejemplo, las matrices de transicion de una cadema de Markov (véase el Ejercicio 2.11) son matrices no negativas. Sin embargo, algunas matrices no negativas tienen autovalores negativas: un fenómeno desagradable.

Afortunadamente, hay otra noción de positividad que garantiza autovalores no negativos, aun cuando algunas entradas pueden ser negativas o complejas. A la luz de las Proposiciones 3.25 y 3.35, se trata de la clase de matrices hermíticas cuyos autovalores son no negativos, o cuyos valores vectoriales $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle$ son no negativos. (Como se verá luego, estas dos condiciones son equivalentes.)

Definición 3.36. Dícese que $A \in M_n(\mathbb{C})$ es una **matriz positiva** (a veces: *matriz positiva semidefinida*)[8] si se cumple

$$\underline{\langle \boldsymbol{x}, A\boldsymbol{x} \rangle \geqslant 0} \quad \text{para todo} \quad \boldsymbol{x} \in \mathbb{C}^n. \tag{3.16}$$

Por la Proposición 3.35 anterior, una tal A es hermítica. En particular, $A \in M_n(\mathbb{R})$ es positiva (semidefinida) si y solo si es *simétrica* y cumple (3.16), es decir, $\boldsymbol{x}^{\mathrm{t}} A\boldsymbol{x} \geqslant 0$ para todo $\boldsymbol{x} \in \mathbb{R}^n$.

Dícese que $A \in M_n(\mathbb{C})$ es **definida positiva** si la desigualdad en (3.16) es estricta para vectores no nulos: $\underline{\langle \boldsymbol{x}, A\boldsymbol{x} \rangle > 0}$ para $\boldsymbol{x} \neq \boldsymbol{0}$ en $\mathbb{C}^n$. ◊

Ejemplo 3.37. Si $\underline{A = B^*B}$ para alguna matriz $B \in M_n(\mathbb{C})$, entonces A es una matriz positiva. En efecto,[9]

$$\langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \langle \boldsymbol{x}, B^*B\boldsymbol{x} \rangle = \langle B\boldsymbol{x}, B\boldsymbol{x} \rangle = \|B\boldsymbol{x}\|^2 \geqslant 0, \quad \text{para} \quad \boldsymbol{x} \in \mathbb{C}^n.$$

Del mismo modo, cada matriz de la forma $\underline{A = CC^*}$ es positiva: tómese $B := C^*$. ◊

Proposición 3.38. *Una matriz $A \in M_n(\mathbb{C})$ es definida positiva si y solo si es positiva (semidefinida) e invertible.*

Demostración. La matriz A del enunciado es positiva semidefinida en ambos casos; en particular, A es hermítica. Por la Proposición 3.26, $\mathbb{C}^n$ posee una base ortonormal $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ de autovectores de A. Sean $\lambda_1, \dots, \lambda_n$ los autovalores correspondientes. La matriz A es invertible si y solo si su forma diagonal $\operatorname{diag}[\lambda_1, \dots, \lambda_n]$ es invertible, si y solo si $\lambda_i \neq 0$ para $i = 1, \dots, n$.

Entonces, si A no es invertible, hay al menos un autovalor nulo $\lambda_i = 0$. En este caso, el autovector $\boldsymbol{u}_i$ cumple $\langle \boldsymbol{u}_i, A\boldsymbol{u}_i \rangle = \langle \boldsymbol{u}_i, \boldsymbol{0} \rangle = 0$, y por eso A no es definida positiva.

[8]En particular, la matriz nula 0 es positiva (semidefinida). Como el término *matriz no negativa* ha sido reservado para otro concepto, no queda más que admitir la matriz nula y el número cero con entes "positivos" de manera honoraria.

[9]Resulta que este ejemplo es universal: dada una matriz positiva A, existe otra matriz B tal que $A = B^*B$. Esto se verificará más adelante, en el Teorema 3.58.

Supóngase, por el contrario, que A es positiva e invertible, así que $\lambda_i \neq 0$ para $i = 1, \dots, n$. De hecho, los autovalores son números positivos: $\lambda_i = \langle \boldsymbol{u}_i, A\boldsymbol{u}_i \rangle > 0$. Obsérvese que si $\boldsymbol{x} = c_1\boldsymbol{u}_1 + \cdots + c_n\boldsymbol{u}_n$, entonces

$$\begin{aligned} A\boldsymbol{x} &= c_1\lambda_1\boldsymbol{u}_1 + \cdots + c_n\lambda_n\boldsymbol{u}_n, \\ \langle \boldsymbol{x}, A\boldsymbol{x} \rangle &= |c_1|^2\lambda_1 + \cdots + |c_n|^2\lambda_n. \end{aligned}$$

Las desigualdades $\lambda_i > 0$ implican que

$$\langle \boldsymbol{x}, A\boldsymbol{x} \rangle = 0 \iff c_1 = \cdots = c_n = 0 \iff \boldsymbol{x} = \boldsymbol{0}.$$

Por lo tanto, $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle > 0$ para $\boldsymbol{x} \neq \boldsymbol{0}$; es decir, A es definida positiva. □

► Cada matriz real simétrica es hermítica. Por tanto, la Proposición 3.38 anterior da un criterio para detectar si una matriz dada, que ya es real y simétrica, es definida positiva o no. En primer lugar, se debe determinar si A es invertible (por ejemplo, al evaluar $\det A$); después es necesario comprobar las relaciones de positividad (3.16). La eliminación gaussiana simple, *sin intercambios de filas*, permite resolver las dos inquietudes simultáneamente.

Proposición 3.39. *Una matriz $A \in M_n(\mathbb{R})$ es definida positiva si y solo si A es simétrica y en la eliminación gaussiana simple los pivotes sucesivos $a_{kk}^{(k)}$ son todos positivos.*

Demostración. En ambos lados de esta equivalencia, A es una matriz real simétrica.

Cuando la eliminación gaussiana simple es aplicable con éxito a una matriz cuadrada invertible A, su efecto es una factorización $\underline{A = LDU}$: en ella, D es una matriz diagonal, L es triangular inferior unipotente y U es triangular superior unipotente. Además, ésta descomposición es única. La matriz transpuesta A^{t} tiene una factorización correspondiente, $\underline{A^{\mathsf{t}} = U^{\mathsf{t}}DL^{\mathsf{t}}}$. Como A es simétrica, $A^{\mathsf{t}} = A$, esta unicidad implica que $U = L^{\mathsf{t}}$.

Ad ($\Rightarrow$): Supóngase que A es definida positiva.

La eliminación gaussiana *simple* funciona cuando ningún pivote $a_{kk}^{(k)}$ se anula durante el proceso. Para $k = 1$, vale

$$a_{11}^{(1)} = a_{11} = \langle \boldsymbol{e}_1, A\,\boldsymbol{e}_1 \rangle > 0.$$

Cada operación de fila de tipo (b) – en el lenguaje de la Sección 1.4 – se ejecuta al premultiplicar A por varias matrices $R_{ij}(c)$ que son triangulares inferiores, véase la fórmula (1.14b); un producto de tales matrices es también triangular inferior. Entonces, en el paso número k de la eliminación, ya se ha hecho una conversión $A \mapsto L_kA$ donde la matriz cuadrada L_k es triangular inferior. El elemento elemento (k, k) de una matriz L_kA es el siguiente pivote $a_{kk}^{(k)}$; y hay ceros debajo de la diagonal en las primeras $(k-1)$ columnas de la nueva matriz L_kA.

Como A es simétrica, la *posmultiplicación* $L_k A \mapsto L_k A L_k^{\mathsf{t}}$ ejecuta las operaciones *de columna* que son análogas a las operaciones de fila anteriores; el resultado es una matriz que tiene ceros *a la derecha* de la diagonal en sus primeras $(k-1)$ *filas*. Dichas operaciones de columna no cambian los elementos diagonales $a_{jj}^{(j)}$ ya calculados, para $j = 1, \dots, k$, porque en cada paso su efecto es $a_{jj}^{(j)} \mapsto a_{jj}^{(j)} + 0$. Al inicio del paso número k, entonces, la matriz tiene el siguiente aspecto:

$$L_k A L_k^{\mathsf{t}} = \begin{bmatrix} a_{11} & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & a_{22}^{(2)} & \dots & 0 & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & a_{kk}^{(k)} & * & \dots & * \\ 0 & 0 & \dots & * & * & \dots & * \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & * & * & \dots & * \end{bmatrix}$$

donde cada $*$ representa una entrada desconocida.

Entonces $a_{kk}^{(k)}$ es el elemento (k,k) de la matriz simétrica $L_k A L_k^{\mathsf{t}}$, así que

$$a_{kk}^{(k)} = \langle \boldsymbol{e}_k, L_k A L_k^{\mathsf{t}} \boldsymbol{e}_k \rangle = \langle L_k^{\mathsf{t}} \boldsymbol{e}_k, A L_k^{\mathsf{t}} \boldsymbol{e}_k \rangle > 0, \tag{3.17}$$

ya que A es definida positiva. Fíjese que $L_k^{\mathsf{t}} \boldsymbol{e}_k \neq \boldsymbol{0}$ porque $\langle L_k^{\mathsf{t}} \boldsymbol{e}_k, \boldsymbol{e}_k \rangle = \langle \boldsymbol{e}_k, L_k \boldsymbol{e}_k \rangle = 1$. En efecto, al aplicar las mismas operaciones *de fila* a la matriz identidad, no cambian la entrada 1 en la posición (k,k) de la matriz 1_n.

Después de $(n-1)$ pasos, se llega a una matriz diagonal

$$D := L_n A L_n^{\mathsf{t}} = \operatorname{diag}\big[a_{11}, a_{22}^{(2)}, \dots, a_{nn}^{(n)}\big].$$

Al definir $L := L_n^{-1}$, también triangular inferior, se ha obtenido la factorización $A = LDL^{\mathsf{t}}$ sin necesidad de hacer intercambios de filas. La fórmula (3.17), válida ahora para $k = 1, \dots, n$, dice que todos los pivotes son positivos.

Ad ($\Leftarrow$): Ahora sea A una matriz simétrica e invertible que se factoriza como $A = LDL^{\mathsf{t}}$ por eliminación gaussiana simple. La matriz D es diagonal, $D = \operatorname{diag}[d_{11}, d_{22}, \dots, d_{nn}]$. Los pivotes son sus elementos diagonales d_{kk}. Si cada $d_{kk} > 0$, entonces $D = D^{1/2} D^{1/2}$, donde la **raíz cuadrada positiva** de D se define como $D^{1/2} := \operatorname{diag}[\sqrt{d_{11}}, \dots, \sqrt{d_{nn}}]$. Luego

$$A = LDL^{\mathsf{t}} = LD^{1/2} D^{1/2} L^{\mathsf{t}} = B^{\mathsf{t}} B, \quad \text{al poner} \quad B := D^{1/2} L^{\mathsf{t}}.$$

Así, A es invertible y positiva; por la Proposición 3.38, A es definida positiva. □

Definición 3.40. Una matriz definida positiva $A \in M_n(\mathbb{R})$ tiene la factorización $A = LDL^{\mathsf{t}}$ por eliminación gaussiana. Sea $\underline{C := LD^{1/2}}$; esta es una matriz triangular inferior. La presentación $\underline{A = CC^{\mathsf{t}}}$ se llama la **factorización de Cholesky** de A. ◊

Esta factorización viene de la siguiente circunstancia: por ser A definida positiva, se sabe que $A = BB^{\mathrm{t}}$ para *alguna* matriz B. Pero tal matriz B no es única, pues es una entre varias posibilidades: para empezar, se puede reemplazar B por $(-B)$. Entre todas estas posibilidades, hay exactamente una, $B = C$, que es triangular inferior con entradas positivas en la diagonal. Se puede calcular el factor C en forma directa por una variante de la eliminación gaussiana.

3.5. Operadores autoadjuntos

Un espacio vectorial V sobre $\mathbb{C}$, de dimensión finita, dotado de un producto escalar $\langle -, - \rangle$, se llama un **espacio de Hilbert** (finitodimensional); o a veces *espacio hilbertiano*.[10] En esta sección, V y W denotarán dos espacios de Hilbert finitodimensionales.

Proposición 3.41. *Sea $T \in \mathcal{L}(V, W)$ una aplicación lineal; su adjunto es $T^* \in \mathcal{L}(W, V)$. Entonces* $\ker T^*$ *es el complemento ortogonal $T(V)^\perp$ de la imagen de T; mientras* $\ker T$ *es el complemento ortogonal $T^*(W)^\perp$ de la imagen de T^*.*

En consecuencia, T^* es uno-a-uno si y solo si T es sobreyectivo; y viceversa.

Demostración. Obsérvese que $T(V)^\perp \leqslant W$ y que $T^*(W)^\perp \leqslant V$.

Para cualquier vector $\boldsymbol{y} \in W$, la fórmula (3.12) muestra que

$$\begin{aligned} \boldsymbol{y} \in \ker T^* &\iff T^*\boldsymbol{y} = \boldsymbol{0} \quad \text{en } V \\ &\iff \langle T^*\boldsymbol{y}, \boldsymbol{x} \rangle = 0 \quad \text{para todo} \quad \boldsymbol{x} \in V \\ &\iff \langle \boldsymbol{y}, T\boldsymbol{x} \rangle = 0 \quad \text{para todo} \quad T\boldsymbol{x} \in T(V) \\ &\iff \boldsymbol{y} \in T(V)^\perp. \end{aligned}$$

De igual manera, para cualquier vector $\boldsymbol{x} \in V$, se obtiene

$$\begin{aligned} \boldsymbol{x} \in \ker T &\iff T\boldsymbol{x} = \boldsymbol{0} \quad \text{en } W \\ &\iff \langle \boldsymbol{y}, T\boldsymbol{x} \rangle = 0 \quad \text{para todo} \quad \boldsymbol{y} \in W \\ &\iff \langle T^*\boldsymbol{y}, \boldsymbol{x} \rangle = 0 \quad \text{para todo} \quad T^*\boldsymbol{y} \in T^*(W) \\ &\iff \boldsymbol{x} \in T^*(W)^\perp. \end{aligned}$$

□

Definición 3.42. Un operador lineal $T \in \mathcal{L}(V)$ es **autoadjunto** si $\underline{T^* = T}$ en $\mathcal{L}(V)$.

Si $A = [T]^{\mathcal{E}}_{\mathcal{E}} \in M_n(\mathbb{C})$ es la matriz de T con respecto a una base ortonormal $\mathcal{E}$ de V, entonces A es *hermítica*, es decir, $\underline{A^* = A}$ en $M_n(\mathbb{C})$. ◊

[10]Los trabajos de *David Hilbert* y su estudiante *Erhard Schmidt* sobre ecuaciones integrales en los años 1906–08 resaltaran el producto escalar (del Ejemplo 3.4) en espacios de funciones de cuadrado integrable. Esos espacios también entraron en la formulación matemática de la mecánica cuántica en los años 1925–30. El término “espacio de Hilbert” fue introducido por otro de sus colaboradores, *John von Neumann* en 1929, para describir espacios vectoriales infinitodimensionales (completos) con producto escalar. Hoy en día, se usa ese término para al caso finitodimensional también.

Lema 3.43. *Todo $T \in \mathcal{L}(V)$ puede expresarse, de forma única, como $\underline{T = T_1 + i\,T_2}$ donde T_1 y T_2 son operadores autoadjuntos.*

Demostración. Si $T = T_1 + iT_2$ con T_1, T_2 autoadjuntos, entonces $T^* = T_1^* - iT_2^* = T_1 - iT_2$ puesto que la correspondencia $T \mapsto T^*$ es semilineal. De ahí se obtiene

$$\boxed{T_1 = \tfrac{1}{2}(T^* + T), \qquad T_2 = \tfrac{i}{2}(T^* - T).} \tag{3.18}$$

Esto establece la *unicidad* de la "parte real" T_1 y de la "parte imaginaria" T_2 del operador T.

Por otro lado, (3.18) permite *definir* T_1 y T_2 por $T_1 := \frac{1}{2}(T^* + T)$, $T_2 := \frac{i}{2}(T^* - T)$. Es inmediato que

$$T_1^* = \tfrac{1}{2}(T + T^*) = T_1 \quad \text{y} \quad T_2^* = -\tfrac{i}{2}(T - T^*) = T_2.$$

Esto establece la *existencia* de la descomposición deseada. □

Lema 3.44. *Si $T \in \mathcal{L}(V)$ es autoadjunto, sus autovalores son reales.*

Demostración. Esto recapitula la primera parte de la Proposición 3.25, con una demostración paralela. Si $T^* = T$ y si $\lambda \in \mathbb{C}$ es un autovalor de T con un autovector correspondiente $\boldsymbol{x} \neq \boldsymbol{0}$, entonces

$$\lambda \langle \boldsymbol{x}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, \lambda \boldsymbol{x} \rangle = \langle \boldsymbol{x}, T\boldsymbol{x} \rangle = \langle T\boldsymbol{x}, \boldsymbol{x} \rangle = \langle \lambda \boldsymbol{x}, \boldsymbol{x} \rangle = \bar{\lambda} \langle \boldsymbol{x}, \boldsymbol{x} \rangle$$

así que $\bar{\lambda} = \lambda$, es decir, $\lambda \in \mathbb{R}$. □

Fíjese también que si $T^* = T$, entonces $\underline{\langle \boldsymbol{x}, T\boldsymbol{x} \rangle \in \mathbb{R}}$ para cada $\boldsymbol{x} \in V$, porque

$$\overline{\langle \boldsymbol{x}, T\boldsymbol{x} \rangle} = \langle T\boldsymbol{x}, \boldsymbol{x} \rangle = \langle T^*\boldsymbol{x}, \boldsymbol{x} \rangle = \langle \boldsymbol{x}, T\boldsymbol{x} \rangle.$$

Definición 3.45. Un operador lineal $T \in \mathcal{L}(V)$ se llama **idempotente** si $\underline{T^2 = T}$. Nótese que también $T^k = T$, para $k = 3, 4, \dots$

Un operador lineal $P \in \mathcal{L}(V)$ es un **proyector**, o a veces *proyector ortogonal*,[11] si P es *idempotente y autoadjunto* a la vez, es decir, si $\underline{P^2 = P = P^*}$. ◊

Sea $P \in \mathcal{L}(V)$ un proyector. La restricción de P a su imagen $P(V)$ coincide con el operador identidad sobre $P(V)$, porque $P(P(\boldsymbol{x})) = P(\boldsymbol{x})$ para todo $\boldsymbol{x} \in V$. Por otro lado, se cumple $P(V)^\perp = \ker P^* = \ker P$ por la Proposición 3.41, así que la restricción de P al complemento ortogonal de su imagen es el operador nulo.

En consecuencia, $P(V)^\perp = \ker P^*$ es obviamente un subespacio invariante para P. Como $P(V) \oplus P(V)^\perp = V$ por el Lema 3.17, el subespacio $P(V)$ *reduce* el proyector P.

[11] Cualquier operador idempotente $T \in \mathcal{L}(V)$ se restringe al operador identidad sobre su imagen $T(V)$, pero su núcleo $\ker T$ no es necesariamente ortogonal a $T(V)$. Si $W = T(V)$, la aplicación sobreyectiva $\boldsymbol{x} \mapsto T(\boldsymbol{x})$ de V en W se llama la *proyección* de V sobre W *a lo largo de* $\ker T$. Conviene distinguir las palabras "proyección" (aplicación sobreyectiva) y "proyector" (elemento de $\mathcal{L}(V)$ que cumple $P^2 = P = P^*$), aunque muchos autores las confunden. *Caveat lector*.

Los proyectores abundan en $\mathcal{L}(V)$, porque están en correspondencia biunívoca con los subespacios de V, en vista del resultado siguiente.

Proposición 3.46. *Si M es un subespacio de V, existe* un único proyector $P_M \in \mathcal{L}(V)$ *tal que* $P_M(V) = M$.

Demostración. Sea $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\}$ una base ortonormal del subespacio M (obtenida al aplicar el algoritmo de Gram y Schmidt a una base cualquiera de M). Por el Corolario 3.19, hay otros vectores $\boldsymbol{e}_{m+1}, \dots, \boldsymbol{e}_n$ tales que $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es una base ortonormal de V.

Es evidente que $\boldsymbol{e}_k \in M^\perp$ para $k > m$; y si $\boldsymbol{x} \in M^\perp$ con la expansión $\boldsymbol{x} = c_1\,\boldsymbol{e}_1 + \cdots + c_n\,\boldsymbol{e}_n$, entonces $c_j = \langle \boldsymbol{e}_j, \boldsymbol{x} \rangle = 0$ para $j \leqslant m$. Se deduce que $\{\boldsymbol{e}_{m+1}, \dots, \boldsymbol{e}_n\}$ es una base ortonormal de $M^\perp$. (Esto proporciona una segunda demostración de la relación $M \oplus M^\perp = V$, ya notado en el Lema 3.17.)

Ahora defínase $P_M \in \mathcal{L}(V)$ por

$$P_M(\boldsymbol{x}) := \langle \boldsymbol{e}_1, \boldsymbol{x} \rangle\, \boldsymbol{e}_1 + \cdots + \langle \boldsymbol{e}_m, \boldsymbol{x} \rangle\, \boldsymbol{e}_m. \tag{3.19}$$

Está claro que $P_M(V) = M$ y que $P_M^2 = P_M$.

La igualdad $P_M^* = P_M$ sigue del cálculo siguiente: si $\boldsymbol{x}, \boldsymbol{y} \in V$, entonces

$$\begin{aligned} \langle \boldsymbol{y}, P_M \boldsymbol{x} \rangle &= \left\langle \sum_{j=1}^n \langle \boldsymbol{e}_j, \boldsymbol{y} \rangle\, \boldsymbol{e}_j, \sum_{i=1}^m \langle \boldsymbol{e}_i, \boldsymbol{x} \rangle\, \boldsymbol{e}_i \right\rangle = \sum_{j=1}^n \sum_{i=1}^m \overline{\langle \boldsymbol{e}_j, \boldsymbol{y} \rangle}\, \langle \boldsymbol{e}_i, \boldsymbol{x} \rangle\, \langle \boldsymbol{e}_j, \boldsymbol{e}_i \rangle \\ &= \sum_{k=1}^m \overline{\langle \boldsymbol{e}_k, \boldsymbol{y} \rangle}\, \langle \boldsymbol{e}_k, \boldsymbol{x} \rangle = \left\langle \sum_{j=1}^m \langle \boldsymbol{e}_j, \boldsymbol{y} \rangle\, \boldsymbol{e}_j, \sum_{i=1}^n \langle \boldsymbol{e}_i, \boldsymbol{x} \rangle\, \boldsymbol{e}_i \right\rangle = \langle P_M \boldsymbol{y}, \boldsymbol{x} \rangle. \end{aligned}$$

Para la unicidad, obsérvese que cualquier proyector P con $P(V) = M$ cumple $P\boldsymbol{z} = \boldsymbol{z}$ para $\boldsymbol{z} \in M$ y además $P\boldsymbol{y} = \boldsymbol{0}$ para $\boldsymbol{y} \in M^\perp$. Como $V = M \oplus M^\perp$, cualquier $\boldsymbol{x} \in V$ puede escribirse de manera única como

$$\boldsymbol{x} = \boldsymbol{z} + \boldsymbol{y}, \quad \text{con} \quad \boldsymbol{z} \in M,\ \boldsymbol{y} \in M^\perp.$$

Luego $P\boldsymbol{x} = \boldsymbol{z}$. Por otro lado, vale $P_M\boldsymbol{x} = P_M\boldsymbol{z} + P_M\boldsymbol{y} = \boldsymbol{z} + \boldsymbol{0} = \boldsymbol{z}$. Se concluye que $P\boldsymbol{x} = P_M\boldsymbol{x}$ para todo $\boldsymbol{x} \in V$, así que $P = P_M$. □

La fórmula (3.19) dice que la matriz de P_M, con respecto a la base ortonormal $\mathcal{E}$ de la demostración anterior, es una matriz de bloques:

$$[P_M]_{\mathcal{E}}^{\mathcal{E}} = \begin{bmatrix} 1_m & 0 \\ 0 & 0 \end{bmatrix} = \operatorname{diag}[\underbrace{1, \dots, 1}_{m}, \underbrace{0, \dots, 0}_{n-m}].$$

(En particular, los subespacios M y $M^\perp$ reducen el proyector P_M.)

Lema 3.47. *Dos subespacios M y N de V son* ortogonales *(escrito $\underline{M \perp N}$: $\langle z, \boldsymbol{x}\rangle = 0$ para $\boldsymbol{x} \in M$, $z \in N$), si y solo si $P_M P_N = 0$ en $\mathcal{L}(V)$.*

Demostración. Si $\boldsymbol{x} \in M$ y $z \in N$, entonces

$$\langle z, \boldsymbol{x}\rangle = \langle P_N z, P_M \boldsymbol{x}\rangle = \langle P_M^* P_N z, \boldsymbol{x}\rangle = \langle P_M P_N z, \boldsymbol{x}\rangle.$$

Luego $P_M P_N = 0$ implica que $\langle z, \boldsymbol{x}\rangle = 0$ para $\boldsymbol{x} \in M$, $z \in N$, así que $M \perp N$.

Inversamente, si $M \perp N$, entonces $N \subseteq M^\perp$. Si $\boldsymbol{y} \in V$, entonces $P_N \boldsymbol{y} \in N$ y por ende $P_N \boldsymbol{y} \in M^\perp$, y luego $P_M(P_N \boldsymbol{y}) = \mathbf{0}$. Como $\boldsymbol{y}$ es arbitrario, esto dice que $P_M P_N = 0$. ◻

► En un espacio de Hilbert, la *semejanza* de matrices $A \sim P^{-1}AP$ no es la clasificación más natural, pues la mera invertibilidad de la matriz P no toma en cuenta el producto escalar. Las matrices cuadradas en $M_n(\mathbb{C})$ se clasifican mejor por **semejanza unitaria**, es decir, por la relación $\underline{A \sim U^{-1}AU = U^*AU}$ donde U es una *matriz unitaria*. La siguiente Proposición muestra que cualquier matriz en $M_n(\mathbb{C})$ es unitariamente semejante a una matriz triangular, es decir, es **trigonalizable** por una semejanza unitaria.

Proposición 3.48. *Si $T \in \mathcal{L}(V)$, entonces hay una cadena de subespacios invariantes*

$$\{\mathbf{0}\} \leqslant R_1 \leqslant \cdots \leqslant R_{n-1} \leqslant R_n = V$$

tal que $\dim R_k = k$ *y* $T(R_k) \subseteq R_k$ *para* $k = 1, \dots, n$.

En consecuencia, hay una base ortonormal *en V respecto del cual T posee una* matriz triangular.

Demostración. Esto se demuestra por inducción sobre $n = \dim V$; el resultado es evidente si $\dim V = 1$. Supóngase, entonces, que la Proposición es válida para espacios vectoriales de dimensión $(n-1)$.

El operador adjunto T^* tiene al menos un autovalor $\mu \in \mathbb{C}$ [el cual es una raíz compleja de su polinomio característico $p_{T^*}(t)$.] Sea $\boldsymbol{y} \in V \setminus \{\mathbf{0}\}$ un autovector, de modo que $T^*\boldsymbol{y} = \mu \boldsymbol{y}$. El subespacio unidimensional $\mathbb{C}\boldsymbol{y} = \{\, \alpha \boldsymbol{y} : \alpha \in \mathbb{C}\,\}$ es invariante bajo T^*. Ahora defínase $R_{n-1} := (\mathbb{C}\boldsymbol{y})^\perp$. Nótese que $\dim R_{n-1} = n - \dim(\mathbb{C}\boldsymbol{y}) = n-1$. Si $\boldsymbol{x} \in R_{n-1}$, entonces

$$\langle T\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{x}, T^*\boldsymbol{y}\rangle = \mu \langle \boldsymbol{x}, \boldsymbol{y}\rangle = 0,$$

así que $T\boldsymbol{x} \in R_{n-1}$. Esto muestra que $T(R_{n-1}) \subseteq R_{n-1}$.

Sea $S \in \mathcal{L}(R_{n-1})$ la *restricción* de T a este subespacio: esto es, $S\boldsymbol{x} := T\boldsymbol{x} \in R_{n-1}$ para cada $\boldsymbol{x} \in R_{n-1}$. Al reemplazar V por R_{n-1} y T por S, la hipótesis inductiva muestra que hay una cadena de subespacios $\{\mathbf{0}\} \leqslant R_1 \leqslant \cdots \leqslant R_{n-1}$ con $\dim R_k = k$ y además $S(R_k) \subset R_k$ para $k = 1, \dots, n-1$. Como $T\boldsymbol{x} = S\boldsymbol{x}$ para cada $\boldsymbol{x} \in R_k$, el resultado ya está demostrado para $\dim V = n$.

Para obtener la matriz triangular de T, es cuestión de encontrar una base ortonormal apropiada $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ de V.

Tómese $\boldsymbol{u}_1 \in R_1$ tal que $\|\boldsymbol{u}_1\| = 1$; tómese $\boldsymbol{u}_2 \in R_2 \cap R_1^\perp$ tal que $\|\boldsymbol{u}_2\| = 1$, etc. Después de elegir $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\} \subset R_k$, se debe tomar $\boldsymbol{u}_{k+1} \in R_{k+1} \cap R_k^\perp$ tal que $\|\boldsymbol{u}_{k+1}\| = 1$. La existencia de ese vector $\boldsymbol{u}_{k+1}$ está garantizado por el Corolario 3.19 (compleción de una base ortonormal parcial).

Como $\dim R_k = k$ para cada k, se ve que $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ es una base ortonormal para R_k, para cada $k = 1, \dots, n$. La invariancia $T(R_k) \subseteq R_k$ significa que

$$T(\boldsymbol{u}_k) = \sum_{i=1}^{k} a_{ik}\, \boldsymbol{u}_i \quad \text{para} \quad k = 1, \dots, n. \tag{3.20}$$

Esta expansión determina la matriz de T – se puede compararla con la fórmula (1.9) – y no aparecen términos a_{ij} con $i > j$; la matriz de T con respecto a esta base es triangular:

$$[T]_{\mathcal{U}}^{\mathcal{U}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ 0 & 0 & a_{33} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{n-1,n} \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}. \qquad \square$$

Corolario 3.49. *Sea $A \in M_n(\mathbb{C})$. Entonces hay una matriz unitaria U tal que la matriz $U^*AU = U^{-1}AU$ es triangular.*

Demostración. Sea $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ la *base estándar* de $\mathbb{C}^n$, de modo que $A = [T_A]_{\mathcal{E}}^{\mathcal{E}}$. Además, sea $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ la base ortonormal de $\mathbb{C}^n$ construida en la demostración de la Proposición 3.48 anterior para el operador $T_A \in \mathcal{L}(\mathbb{C}^n)$. La matriz de cambio de base dada por (1.12) es

$$U := [1]_{\mathcal{U}}^{\mathcal{E}} = [\boldsymbol{u}_1\ \boldsymbol{u}_2\ \dots \boldsymbol{u}_n], \tag{3.21}$$

la cual es unitaria porque sus columnas forman una base ortonormal de $\mathbb{C}^n$. En vista de la expansión (3.20) para T_A en esta base,

$$U^*AU = U^{-1}AU = [1]_{\mathcal{E}}^{\mathcal{U}}\, [T_A]_{\mathcal{E}}^{\mathcal{E}}\, [1]_{\mathcal{U}}^{\mathcal{E}} = [T_A]_{\mathcal{U}}^{\mathcal{U}}$$

es una matriz triangular. $\square$

► En el caso real, el resultado análogo no es cierto: dada una matriz $A \in M_N(\mathbb{R})$, no siempre existe una matriz ortogonal V tal que $V^t AV = V^{-1}AV$ sea triangular. La razón es la posible ausencia de autovalores: como $T_A(R_1) \subseteq R_1$, un vector no nulo en R_1 debería ser un autovector de A, con un autovalor real.

3.6. El teorema espectral y sus consecuencias

Cualquier operador lineal sobre un espacio de Hilbert finitodimensional V tiene una matriz triangular respecto de alguna base ortonormal; en consecuencia, cualquier matriz $A \in M_n(\mathbb{C})$ es *trigonalizable* por cambio de base ortonormal. Hace falta mejorar ese resultado, al obtener un criterio para que A sea (unitariamente) *diagonalizable*. Se busca, entonces, una condición sobre una matriz A que garantiza que $D := U^*AU$ es una matriz diagonal para alguna matriz unitaria U.

En constraste con el caso general del Capítulo 2, no se trata de analizar el polinomio mínimo, porque eso conduce a cambios de base cualesquiera, sin tomar en cuenta el producto escalar. En su lugar, se destaca el uso de proyectores, como luego se verá.

► Para empezar, se debe abordar primero el caso de los operadores *autoadjuntos* sobre V y las matrices *hermíticas* correspondientes $A = A^* \in M_n(\mathbb{C})$. En este caso la diagonalización $U^*AU = D$ sí es factible, por el lema que sigue.

Lema 3.50. *Una matriz hermítica $A = A^* \in M_n(\mathbb{C})$ es diagonalizable, mediante conjugación $\underline{A \mapsto U^*AU}$ por una matriz unitaria U.*

Demostración. Se sabe, por la Proposición 3.26, que la matriz hermítica A tiene una base ortonormal de autovectores $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\} \subset \mathbb{C}^n$. La matriz U de (3.21) cuyas columnas son esos autovectores cumple $AU = UD$, donde $D = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$ es la matriz diagonal que recoge los autovalores de A. □

Conviene ahora reformular esa diagonalización usando una colección de proyectores. Además, es oportuno expresar el resultado en términos de la estructura abstracta de operadores autoadjuntos, mediante el **teorema espectral** (primera versión), que se demuestra a continuación.

Teorema 3.51 (Teorema Espectral 1)**.** *Sea V un espacio de Hilbert finitodimensional y sea $T = T^* \in \mathcal{L}(V)$ un operador autoadjunto. Entonces se puede escribir*

$$\boxed{T = \alpha_1 P_1 + \cdots + \alpha_r P_r,} \tag{3.22}$$

donde los $\alpha_j \in \mathbb{R}$ son los autovalores distintos de T, los P_j son proyectores no nulos tales que

$$\boxed{P_i P_j = 0 \quad si \quad i \neq j; \quad y \quad P_1 + \cdots + P_r = 1_V\,.} \tag{3.23}$$

Demostración. La Proposición 3.48 ha construido una base ortonormal $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ de V para la cual T tiene una matriz triangular: $A = [T]^{\mathcal{U}}_{\mathcal{U}}$, con $a_{ij} = 0$ para $i > j$. La matriz de T^* es $A^* = [T^*]^{\mathcal{U}}_{\mathcal{U}}$. La condición $T^* = T$ conlleva $A^* = A$, así que $a_{ij} = 0$ para $i < j$ también:

la matriz A es *diagonal*. Sus elementos diagonales cumplen $\bar{a}_{kk} = a_{kk}$ para $k = 1, \dots, n$, es decir, son reales.

Dichos elementos diagonales son autovalores de la matriz A y también del operador T. Sean $\{\alpha_1, \dots, \alpha_r\}$ los elementos *distintos* de la lista $(a_{11}, \dots, a_{nn})$. El subespacio generado por los $\boldsymbol{u}_k$ tales que $a_{kk} = \alpha_j$ será denotado por $M_j := \ker(T - \alpha_j 1_V)$. Sea P_j el (único) proyector tal que $P_j(V) = M_j$, dado por la Proposición 3.46.

Si $i \neq j$ en $\{1, \dots, r\}$, los subespacios M_i y M_j son ortogonales, ya que son generados por dos partes disjuntas de la base ortonormal $\mathcal{U}$. El Lema 3.47 muestra que $P_i P_j = 0$.

En consecuencia, la suma $P_1 + \cdots + P_r$ es también un proyector. De hecho, la suma es autoadjunto, y es idempotente por el cálculo:

$$(P_1 + \cdots + P_r)^2 = (P_1^2 + \cdots + P_r^2) + \sum_{i<j} (P_i P_j + P_j P_i) = P_1^2 + \cdots + P_r^2 = P_1 + \cdots + P_r.$$

Cada $\boldsymbol{u}_k$ pertenece a un solo subespacio M_j, así que $(P_1 + \cdots + P_r)(\boldsymbol{u}_k) = P_j(\boldsymbol{u}_k) = \boldsymbol{u}_k$. Esto dice que $P_1 + \cdots + P_r = 1_V$ en $\mathcal{L}(V)$.

La expansión (3.22) se verifica así: para cada $k = 1, \dots, n$, se obtiene

$$\begin{aligned}(\alpha_1 P_1 + \cdots + \alpha_r P_r)(\boldsymbol{u}_k) &= \sum_{j=1}^r \alpha_j P_j(\boldsymbol{u}_k) = \sum_{j=1}^r \alpha_j \, [\![\boldsymbol{u}_k \in M_j]\!] \, \boldsymbol{u}_k \\ &= \sum_{j=1}^r \alpha_j \, [\![\alpha_j = a_{kk}]\!] \, \boldsymbol{u}_k = a_{kk} \boldsymbol{u}_k = T\boldsymbol{u}_k \,,\end{aligned}$$

porque $\alpha_j = a_{kk}$ para un solo índice j. Por linealidad, $T\boldsymbol{x} = (\alpha_1 P_1 + \cdots + \alpha_r P_r)(\boldsymbol{x})$ para todo $\boldsymbol{x} \in V$. □

Definición 3.52. El conjunto de autovalores distintos $\{\alpha_1, \dots, \alpha_r\}$ de un operador lineal $T \in \mathcal{L}(V)$ se llama[12] el **espectro** de T.

La expansión $T = \alpha_1 P_1 + \cdots + \alpha_r P_r$ de T dado por el Teorema 3.51 se llama la **descomposición espectral** de T. ◊

► En general, la búsqueda de una base ortonormal de autovectores para una determinada matriz hermítica brinda poca información, porque la base depende de esa matriz. Dadas dos

[12] La palabra *espectro* (literalmente, un fantasma) fue introducido por *Isaac Newton* en 1674 para denotar la banda de colores en que la luz blanca se divide en el arco iris, o bajo su separación por un prisma de vidrio. En un famoso experimento, Newton logró esa separación al proyectar la banda de colores como una aparición fantasmagórica en la pared de su cuarto oscuro. Resulta que el espectro óptico solar es una superposición de líneas delgadas correspondientes a distintas frecuencias (es decir, colores) de la luz. A partir de los trabajos de *Max Born* y *Werner Heisenberg* en los años 1925–26, la energía de los fotones de un determinado color, que es proporcional a su frecuencia, viene dado por un autovalor de cierto operador lineal autoadjunto sobre un espacio de Hilbert (infinitodimensional). De ahí viene la costumbre de llamar "espectro" al conjunto de autovalores (o de autovalores generalizados) de un operador lineal cualquiera.

matrices hermíticas distintas, la base ortonormal que diagonaliza la primera no necesariamente diagonaliza la segunda. Es deseable tener un criterio para la existencia de una sola base ortonormal en la cual las dos matrices tienen forma diagonal. Resulta que esto es posible si y solo si las dos matrices *conmutan*.

Proposición 3.53. *Dos matrices hermíticas $A, B \in M_n(\mathbb{C})$ son* simultáneamente diagonalizables*, es decir, hay una matriz unitaria U tal que tanto U^*AU como U^*BU son matrices diagonales, si y solo si $\underline{AB = BA}$.*

Demostración. Dos matrices diagonales $C = \operatorname{diag}[\kappa_1, \dots, \kappa_n]$ y $D = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$ conmutan, porque $CD = \operatorname{diag}[\kappa_1\lambda_1, \dots, \kappa_n\lambda_n] = DC$. Ahora, si U es una matriz unitaria tal que U^*AU y U^*BU son diagonales, entonces

$$U^*ABU = (U^*AU)(U^*BU) = (U^*BU)(U^*AU) = U^*BAU,$$

así que $AB = U(U^*ABU)U^* = U(U^*BAU)U^* = BA$.

Por otro lado, si $A = A^*$, $B = B^*$ y además $AB = BA$, sea $T_A = \alpha_1 P_1 + \cdots + \alpha_r P_r$ la descomposición espectral de T_A. Sea $M_j = P_j(\mathbb{C}^n) = \ker(T_A - \alpha_j 1_n)$, de manera que $\mathbb{C}^n = M_1 \oplus \cdots \oplus M_r$ es la descomposición primaria[13] de $\mathbb{C}^n$ para el operador T_A. Si $\boldsymbol{x} \in M_j$, entonces

$$AB\boldsymbol{x} = BA\boldsymbol{x} = B(\alpha_j \boldsymbol{x}) = \alpha_j\, B\boldsymbol{x},$$

lo cual dice que el vector $B\boldsymbol{x}$ pertenece al subespacio M_j; se deduce que $T_B(M_j) \subseteq M_j$.

De hecho, el subespacio M_j *reduce* T_B, porque $M_j^\perp = \bigoplus_{i \neq j} M_i$ es también un subespacio invariante para T_B. Si $\widetilde{\mathcal{U}} = \{\tilde{\boldsymbol{u}}_1, \dots, \tilde{\boldsymbol{u}}_n\}$ es una base ortonormal de autovectores de A formada por la unión de bases ortonormales de los subespacios $M_1, \dots, M_r$, y si $\widetilde{U} := [\tilde{\boldsymbol{u}}_1\ \tilde{\boldsymbol{u}}_2\ \dots \tilde{\boldsymbol{u}}_r]$ es la matriz unitaria correspondiente, entonces $\widetilde{U}^*A\widetilde{U}$ es diagonal y $\widetilde{U}^*B\widetilde{U} = B_1 \oplus \cdots \oplus B_r$ es una suma directa de bloques. Cada bloque B_j es una matriz hermítica en $M_{m_j}(\mathbb{C})$, donde $m_j = \dim M_j$.

Se diagonaliza cada bloque B_j por un cambio de base ortonormal en cada subespacio M_j por separado. Entonces, hay una matriz unitaria de la forma $W = W_1 \oplus \cdots \oplus W_r$ tal que $W^*(\widetilde{U}^*B\widetilde{U})W$ es diagonal. Sea $U := \widetilde{U}W$ y nótese que U es unitaria por ser el producto de dos matrices unitarias. Cada W_j conmuta con el bloque escalar $\alpha_j 1_{m_j}$ de la matriz diagonal $\widetilde{U}^*A\widetilde{U}$, así que $U^*AU = W^*(\widetilde{U}^*A\widetilde{U})W = \widetilde{U}^*A\widetilde{U}$. Al final, ambas U^*BU y U^*AU son matrices diagonales. □

Corolario 3.54. *Dos operadores autoadjuntos $S, T \in \mathcal{L}(V)$ poseen matrices diagonales respecto de una misma base ortonormal si y solo si $ST = TS$.* ⊟

[13] En la demostración de la Proposición 3.26 se vio que la descomposición primaria de una matriz hermítica es una suma directa ortogonal.

► Un operador lineal cualquiera en $\mathcal{L}(V)$ no necesariamente posee una descomposición espectral análogo a (3.22) – porque su matriz sería diagonalizable. Sin embargo, esa expansión sigue en vigor para una clase de operadores más amplia que los autoadjuntos. Considérese un operador $T \in \mathcal{L}(V)$ que posee una base ortonormal de autovectores con autovalores distintos $\alpha_1, \dots, \alpha_r \in \mathbb{C}$, que no son necesariamente reales. En tal caso se puede escribir:

$$T = \alpha_1 P_1 + \cdots + \alpha_r P_r, \tag{3.24a}$$

$$T^* = \bar{\alpha}_1 P_1 + \cdots + \bar{\alpha}_r P_r. \tag{3.24b}$$

Si algún $\alpha_j \notin \mathbb{R}$, entonces T no es autoadjunto, pero se cumple la relación

$$T^*T = |\alpha_1|^2 P_1 + \cdots + |\alpha_r|^2 P_r = TT^*.$$

Definición 3.55. Un operador lineal T sobre un espacio de Hilbert V se llama un **operador normal** si $\underline{T^*T = TT^*}$.

Una matriz cuadrada compleja $A \in M_n(\mathbb{C})$ es una **matriz normal** si $\underline{A^*A = AA^*}$.

En particular, los operadores unitarios o autoadjuntos [respectivamente, las matrices unitarias o hermíticas] son normales. ◊

Teorema 3.56 (Teorema Espectral 2)**.** *Sea V un espacio de Hilbert finitodimensional y sea $T \in \mathcal{L}(V)$ un* operador normal. *Entonces T y T^* obedecen los desarrollos* (3.24)*, donde los $\alpha_j \in \mathbb{C}$ son los autovalores distintos de T y $P_1, \dots, P_r$ son proyectores no nulos que cumplen* (3.23)*.*

Hay dos maneras alternativas de demostrar este resultado.

Demostración 1. El Lema 3.43 dice que $T = T_1 + iT_2$ donde T_1 y T_2 son autoadjuntos. Es obvio que $T^* = T_1 - iT_2$. Luego T es normal si y solo si las dos partes conmutan: $T_1T_2 = T_2T_1$.

Por el Corolario 3.54, T es normal si y solo si hay una base ortonormal $\mathcal{U}$ de V para la cual las dos matrices $[T_1]^{\mathcal{U}}_{\mathcal{U}}$ y $[T_2]^{\mathcal{U}}_{\mathcal{U}}$ son diagonales; luego la matriz $[T]^{\mathcal{U}}_{\mathcal{U}} = [T_1]^{\mathcal{U}}_{\mathcal{U}} + i\,[T_2]^{\mathcal{U}}_{\mathcal{U}}$ es también diagonal. Sean $\alpha_1, \dots, \alpha_r$ los autovalores distintos de T y sea $P_j \in \mathcal{L}(V)$ el proyector cuya imagen es $M_j := \ker(T - \alpha_j 1_V)$. Las propiedades (3.23) de los proyectores P_j se verifican de igual manera que en la demostración del Teorema 3.51. □

Demostración 2. Obsérvese que en la demostración del Teorema 3.51 se usó la hipótesis $T = T^*$ únicamente para mostrar que la matriz triangular de T es necesariamente diagonal.

La Proposición 3.48 construye, para cualquier $T \in \mathcal{L}(V)$, una base ortonormal $\mathcal{U}$ de V tal que la matriz $A := [T]^{\mathcal{U}}_{\mathcal{U}}$ sea triangular superior: $a_{ij} = 0$ para $i > j$. Entonces $A^* = [T^*]^{\mathcal{U}}_{\mathcal{U}}$ es una matriz triangular inferior. Las entradas diagonales de las matrices A^*A y AA^* son

$$(A^*A)_{kk} = \sum_{i \leqslant k} \bar{a}_{ik} a_{ik} = \sum_{i \leqslant k} |a_{ik}|^2, \qquad (AA^*)_{kk} = \sum_{j \geqslant k} a_{kj} \bar{a}_{kj} = \sum_{j \geqslant k} |a_{kj}|^2. \tag{3.25}$$

Como T es normal, se verifica $A^*A = AA^*$. El caso $k = 1$ de (3.25) entonces implica que

$$|a_{11}|^2 = |a_{11}|^2 + |a_{12}|^2 + |a_{13}|^2 + \cdots + |a_{1n}|^2,$$

y por lo tanto $a_{12} = a_{13} = \cdots = a_{1n} = 0$. El caso $k = 2$ produce la igualdad

$$0 + |a_{22}|^2 = |a_{22}|^2 + |a_{23}|^2 + \cdots + |a_{2n}|^2,$$

y luego $a_{23} = a_{24} = \cdots = a_{2n} = 0$.

Al repetir este argumento para $k = 3, \dots, n$, se obtiene $a_{kj} = 0$ toda vez que $j > k$; la matriz A es diagonal. El resto de la demostración del Teorema 3.51 es aplicable sin cambio. □

Para cualquier *matriz normal* $A \in M_n(\mathbb{C})$, sea $T_A = \alpha_1 P_1 + \cdots + \alpha_r P_r$ la descomposición espectral del operador $T_A : \boldsymbol{x} \mapsto A\boldsymbol{x}$. Si $\mathcal{E}$ es la base estándar de $\mathbb{C}^n$, sea $E_j := [P_j]^{\mathcal{E}}_{\mathcal{E}}$ la matriz del proyector P_j. Estas matrices cumplen $E_j^2 = E_j = E_j^*$ en $M_n(\mathbb{C})$. La fórmula

$$\boxed{A = \alpha_1 E_1 + \cdots + \alpha_r E_r} \quad \text{con} \quad \begin{cases} E_j^2 = E_j = E_j^* & \text{para } j = 1, \dots, r; \\ E_i E_j = 0 & \text{si } i \neq j; \\ E_1 + \cdots + E_r = 1_n \,. \end{cases} \tag{3.26}$$

es la **descomposición espectral** de la matriz normal A.

► El teorema espectral proporciona información valiosa acerca de las matrices positivas en $M_n(\mathbb{C})$ – y también en $M_n(\mathbb{R})$, desde luego. Esto permite completar la lista de caracterizaciones de positividad semidefinida y definida.

Proposición 3.57. *Una matriz $A \in M_n(\mathbb{C})$ es positiva (semidefinida) si y solo si A es hermítica y todas sus autovalores son números no negativos. La matriz A es* definida positiva *si y solo si A es hermítica y todas sus autovalores son números* positivos.

Demostración. Si A es una matriz positiva, entonces A es hermítica por la Proposición 3.35. Sea $A = \alpha_1 E_1 + \cdots + \alpha_r E_r$ su descomposición espectral, con $\alpha_1, \dots, \alpha_r \in \mathbb{R}$.

Si $\boldsymbol{x} \in \mathbb{C}^n$ y $j \in \{1, \dots, r\}$, entonces

$$\langle \boldsymbol{x}, E_j \boldsymbol{x} \rangle = \langle \boldsymbol{x}, E_j^2 \boldsymbol{x} \rangle = \langle \boldsymbol{x}, E_j^* E_j \boldsymbol{x} \rangle = \langle E_j \boldsymbol{x}, E_j \boldsymbol{x} \rangle = \|E_j \boldsymbol{x}\|^2$$

y en consecuencia

$$\langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \sum_{j=1}^{r} \alpha_j \langle \boldsymbol{x}, E_j \boldsymbol{x} \rangle = \sum_{j=1}^{r} \alpha_j \, \|E_j \boldsymbol{x}\|^2. \tag{3.27}$$

Si $\alpha_j \geqslant 0$ para cada j, esta relación muestra que todo $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle \geqslant 0$, así que A es una matriz positiva. Inversamente, si $\boldsymbol{x}_j \in M_j = P_j(\mathbb{C}^n) = \{ \boldsymbol{x} \in \mathbb{C}^n : E_j \boldsymbol{x} = \boldsymbol{x} \}$, entonces $\langle \boldsymbol{x}_j, A\boldsymbol{x}_j \rangle = \alpha_j \, \|\boldsymbol{x}_j\|^2$. Luego, la positividad de A implica que $\alpha_j \geqslant 0$ para cada $j = 1, \dots, r$.

Si $\boldsymbol{x} \neq \boldsymbol{0}$ en $\mathbb{C}^n$, la relación $\boldsymbol{x} = E_1 \boldsymbol{x} + \cdots + E_r \boldsymbol{x}$ implica que $E_j \boldsymbol{x} \neq \boldsymbol{0}$ para al menos un valor de j. De (3.27) se ve que $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle > 0$ para todo $\boldsymbol{x} \neq \boldsymbol{0}$ si y solo si $\alpha_j > 0$ para cada j. □

Teorema 3.58. *Para una matriz $A \in M_n(\mathbb{C})$, las siguientes condiciones son equivalentes:*

(a) *A es una matriz positiva: $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle \geqslant 0$ para todo $\boldsymbol{x} \in \mathbb{C}^n$.*

(b) *A es normal y sus autovalores son números no negativos;*

(c) *$\underline{A = B^*B}$ para alguna matriz $B \in M_n(\mathbb{C})$;*

(d) *$\underline{A = C^2}$ para alguna matriz hermítica $C \in M_n(\mathbb{C})$;*

Además, si A es positiva, hay una única matriz positiva *$R \in M_n(\mathbb{C})$ tal que $\underline{A = R^2}$.*

Esta matriz R se llama la **raíz cuadrada positiva** de A y se denota $\underline{A^{1/2}} := R$.

Demostración. La equivalencia (a) $\iff$ (b) es la Proposición 3.57 anterior. Las implicaciones (d) $\implies$ (c) $\implies$ (a) son evidentes.

Ad (b) $\implies$ (d): Sea $A = \alpha_1 E_1 + \cdots + \alpha_r E_r$ la descomposición espectral de A con cada $\alpha_j \geqslant 0$, y defínase

$$R := \sqrt{\alpha_1}\, E_1 + \sqrt{\alpha_2}\, E_2 + \cdots + \sqrt{\alpha_r}\, E_r.$$

La matriz R cumple la condición (b) y por lo tanto es una matriz positiva (y en particular, hermítica). Su cuadrado se calcula en seguida:

$$R^2 = \sum_{j=1}^{r} \alpha_j E_j^2 + \sum_{i \neq j} \sqrt{\alpha_i \alpha_j}\, E_i E_j = \sum_{j=1}^{r} \alpha_j E_j = A.$$

Para comprobar la unicidad de R, sea S una matriz positiva tal que $S^2 = A$. A partir de su descomposición espectral $S = \beta_1 F_1 + \cdots + \beta_s F_s$, se obtiene

$$\beta_1^2 F_1 + \cdots + \beta_s^2 F_s = S^2 = A = \alpha_1 E_1 + \cdots + \alpha_r E_r\,.$$

El número de autovalores distintos de A entonces es $s = r$. Los conjuntos $\{\beta_1^2, \dots, \beta_r^2\}$ y $\{\alpha_1, \dots, \alpha_r\}$ coinciden porque ellos son todos los autovalores de A. Después de una permutación de este conjunto, se puede suponer que $\beta_j^2 = \alpha_j$ para $j = 1, \dots, r$, y entonces $\beta_j = \sqrt{\alpha_j}$ para cada j porque $\beta_j \geqslant 0$. Finalmente, los autovectores de S y de A corresponden:

$$\{\, \boldsymbol{x} \in \mathbb{C}^n : F_j \boldsymbol{x} = \boldsymbol{x} \,\} = \ker(T_S - \beta_j 1_n) = \ker(T_A - \alpha_j 1_n) = \{\, \boldsymbol{x} \in \mathbb{C}^n : E_j \boldsymbol{x} = \boldsymbol{x} \,\},$$

así que $F_j = E_j$ para cada j. Se ha comprobado que $S = R$. □

► Conviene repetir los conceptos de positividad semidefinida o definida en términos de operadores en vez de matrices.

Definición 3.59. Un operador lineal $T \in \mathcal{L}(V)$ sobre un espacio de Hilbert es un **operador positivo** si T es autoadjunto y si $\langle \boldsymbol{x}, T\boldsymbol{x}\rangle \geqslant 0$ para todo $\boldsymbol{x} \in V$.

Un operador positivo T es un **definido positivo** si $\langle \boldsymbol{x}, T\boldsymbol{x}\rangle > 0$ para todo $\boldsymbol{x} \neq \boldsymbol{0}$ en V. ◊

(La condición de que T es autoadjunto es redundante, porque una versión de la Proposición 3.35 la implica; pero para mayor énfasis se la incluye en la definición anterior.)

Por el Teorema 3.58, aplicado a la matriz de T en cualquier base ortonormal de V, un operador T es positivo si y solo si $T = T^*$ y todos sus autovalores son no negativos; si y solo si $T = S^*S$ para algún operador $S \in \mathcal{L}(V)$; si y solo si $T = Q^2$ para algún operador autoadjunto $Q \in \mathcal{L}(V)$; si y solo si $T = R^2$ para algún operador positivo $R \in \mathcal{L}(V)$.

Esta raíz cuadrada positiva es única: si $T = \alpha_1 P_1 + \cdots + \alpha_r P_r$ es la descomposición espectral de T, entonces $R = \sqrt{\alpha_1}\, P_1 + \cdots + \sqrt{\alpha_r}\, P_r$. Se escribe $\underline{T^{1/2} := R}$ para denotar la raíz cuadrada positiva.

► El teorema espectral también aporta información sobre matrices unitarias.

Proposición 3.60. *Una matriz $U \in M_n(\mathbb{C})$ es* unitaria *si y solo si U es normal y cada autovalor λ de U cumple $\underline{|\lambda| = 1}$.*

Demostración. Si U es unitaria, entonces $U^*U = 1_n = UU^*$, así que U es una matriz normal. Por lo tanto, U posee una descomposición espectral $U = \alpha_1 E_1 + \cdots + \alpha_r E_r$ con $\alpha_1, \dots, \alpha_r \in \mathbb{C}$. La matriz adjunta es $U^* = \bar{\alpha}_1 E_1 + \cdots + \bar{\alpha}_r E_r$. Luego

$$|\alpha_1|^2 E_1 + \cdots + |\alpha_r|^2 E_r = U^*U = 1_n = E_1 + \cdots + E_r.$$

Puesto que $E_i E_j = 0$ para $i \neq j$, se obtiene $|\alpha_i|^2 E_i = E_i$ al premultiplicar ambos lados de la igualdad anterior por la matriz E_i. Si $\boldsymbol{x}$ es un autovector de U para el autovalor α_i, entonces $E_i\boldsymbol{x} = \boldsymbol{x}$ y por ende $|\alpha_i|^2 \boldsymbol{x} = |\alpha_i|^2 E_i \boldsymbol{x} = E_i \boldsymbol{x} = \boldsymbol{x}$ con $\boldsymbol{x} \neq \boldsymbol{0}$. Luego $|\alpha_i| = 1$ para $i = 1, \dots, r$.

Por otro lado, si U es una matriz normal con autovalores complejos de valor absoluto 1, entonces U es de la forma $U = \alpha_1 E_1 + \cdots + \alpha_r E_r$ con cada $|\alpha_j| = 1$. Por tanto, vale

$$U^*U = UU^* = |\alpha_1|^2 E_1 + \cdots + |\alpha_r|^2 E_r = E_1 + \cdots + E_r = 1_n$$

y esto muestra que U es unitaria. □

Definición 3.61. Un operador lineal $W \in \mathcal{L}(V)$ es una **isometría parcial** si $\underline{WW^*W = W}$. ◇

Un operador unitario es una isometría parcial, pero una isometría parcial con $\ker W \neq \{\boldsymbol{0}\}$ no es unitaria.[14] Como $W = WW^*W$ implica $W^* = (WW^*W)^* = W^*WW^*$, se ve que el operador adjunto W^* es otra isometría parcial.

Los operadores WW^* y W^*W son proyectores; son obviamente autoadjuntos, $(WW^*)^2 = (WW^*W)W^* = WW^*$ y también $(W^*W)^2 = W^*(WW^*W) = W^*W$. Es fácil comprobar que WW^* es el proyector cuya imagen es $W(V)$, mientras que W^*W es el proyector cuya imagen es $W^*(V) = (\ker W)^\perp$.

[14] Sobre un espacio de Hilbert infinitodimensional, un operador W se llama una **isometría** si $W^*W = 1$. Esto es equivalente a la condición de que $\|W\boldsymbol{x}\| = \|\boldsymbol{x}\|$ para cada vector $\boldsymbol{x}$, porque $\|W\boldsymbol{x}\|^2 = \langle W\boldsymbol{x}, W\boldsymbol{x}\rangle = \langle \boldsymbol{x}, W^*W\boldsymbol{x}\rangle$. En particular, una isometría es un operador inyectivo. En el contexto infinitodimensional, hay isometrías que no son sobreyectivos, luego no son invertibles; pero en espacios de Hilbert finitodimensionales, cualquier isometría es automáticamente unitaria.

Notación. Si $T \in \mathcal{L}(V)$ es un operador lineal, el operador T^*T es positivo. Escríbase

$$\boxed{|T| := (T^*T)^{1/2}} \qquad (3.28)$$

para denotar la raíz cuadrada positiva de T^*T. El operador positivo $|T|$ se llama el **módulo** del operador T. Fíjese que $|T| = T$ si y solo si T es un operador positivo.[15]

El teorema que sigue muestra que cualquier operador lineal sobre un espacio de Hilbert es el producto de una isometría parcial y un operador positivo (su módulo). Esto es una generalización de la expresión polar de un número complejo: $\underline{z = re^{i\theta}}$, donde $r = |z| \geqslant 0$ y $e^{i\theta} = \cos\theta + i \operatorname{sen}\theta$ es un número complejo de valor absoluto 1.

Teorema 3.62 (Descomposición polar)**.** *Sea V un espacio de Hilbert finitodimensional y sea $T \in \mathcal{L}(V)$. Entonces hay una única isometría parcial $W \in \mathcal{L}(V)$ tal que*

$$\underline{\ker W = \ker T} \quad y \quad \boxed{T = W|T|.} \qquad (3.29)$$

El operador T es invertible si y solo si W es unitaria y $|T|$ es definido positivo.

Demostración. Para aliviar la notación, escríbase $\underline{A := |T|} \equiv (T^*T)^{1/2}$, un operador positivo. Si $\boldsymbol{x} \in V$, se ve que

$$\|A\boldsymbol{x}\|^2 = \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = \langle \boldsymbol{x}, A^2\boldsymbol{x}\rangle = \langle \boldsymbol{x}, T^*T\boldsymbol{x}\rangle = \langle T\boldsymbol{x}, T\boldsymbol{x}\rangle = \|T\boldsymbol{x}\|^2. \qquad (3.30)$$

Por lo tanto, la correspondencia $A\boldsymbol{x} \mapsto T\boldsymbol{x}$ define una aplicación lineal biyectiva del subespacio $A(V)$ en el subespacio $T(V)$. Además, la ecuación (3.30) muestra que $\ker A = \ker T$.

De la Proposición 3.41, se obtiene $A(V)^\perp = \ker A^* = \ker A$, así que $A(V)^\perp = \ker T$.

Defínase $W \in \mathcal{L}(V)$ por

$$W\boldsymbol{y} := \begin{cases} T\boldsymbol{x}, & \text{si } \boldsymbol{y} = A\boldsymbol{x}, \\ \boldsymbol{0}, & \text{si } \boldsymbol{y} \in \ker T. \end{cases}$$

Como $V = A(V) \oplus A(V)^\perp = A(V) \oplus \ker T$, el operador W queda bien definido por esta fórmula. Es inmediato de esta definición que $WA\boldsymbol{x} = T\boldsymbol{x}$ para todo vector $\boldsymbol{x}$, o sea, $\underline{WA = T}$.

Como $T\boldsymbol{x} = \boldsymbol{0}$ si y solo si $A\boldsymbol{x} = \boldsymbol{0}$, se ve que $W\boldsymbol{y} = \boldsymbol{0}$ si y solo si $\boldsymbol{y} \in \ker T$. Se deduce que $\underline{\ker W = \ker T = \ker A}$.

Dado un vector $\boldsymbol{x} \in V$, existen $\boldsymbol{z} \in V$, $\boldsymbol{u} \in \ker T$ tales que $\boldsymbol{x} = A\boldsymbol{z} + \boldsymbol{u}$. Entonces, si $\boldsymbol{y} \in V$ también, se obtiene

$$\begin{aligned} \langle \boldsymbol{x}, W^*W(A\boldsymbol{y})\rangle &= \langle W\boldsymbol{x}, WA\boldsymbol{y}\rangle = \langle WA\boldsymbol{z} + W\boldsymbol{u}, WA\boldsymbol{y}\rangle = \langle WA\boldsymbol{z}, WA\boldsymbol{y}\rangle \\ &= \langle T\boldsymbol{z}, T\boldsymbol{y}\rangle = \langle \boldsymbol{z}, T^*T\boldsymbol{y}\rangle = \langle \boldsymbol{z}, A^2\boldsymbol{y}\rangle = \langle A\boldsymbol{z}, A\boldsymbol{y}\rangle \\ &= \langle A\boldsymbol{z} + \boldsymbol{u}, A\boldsymbol{y}\rangle = \langle \boldsymbol{x}, A\boldsymbol{y}\rangle. \end{aligned}$$

[15]En el caso unidimensional $V = \mathbb{C}$, cualquier elemento de $\mathcal{L}(\mathbb{C})$ es de la forma $w \mapsto zw$ para algún $z \in \mathbb{C}$. El *módulo* de z es sinónimo de su *valor absoluto* $|z| := \sqrt{\bar{z}z} = (\bar{z}z)^{1/2}$.

Aquí se usó $\langle \boldsymbol{u}, A\boldsymbol{y}\rangle = 0$ porque $\ker T = A(V)^\perp$.

Se ha demostrado que $W^*W(A\boldsymbol{y}) = A\boldsymbol{y}$ para todo $\boldsymbol{y} \in V$. Por tanto, $A(V)$ es un subespacio invariante para el operador W^*W y la restricción de W^*W a este subespacio es la identidad. Además, para cada $\boldsymbol{x} \in V$, se ve que

$$WW^*W\boldsymbol{x} = WW^*W(Az + \boldsymbol{u}) = W(Az + W^*W\boldsymbol{u}) = W(Az) = W(Az + \boldsymbol{u}) = W\boldsymbol{x}.$$

Esto dice que $WW^*W = W$, es decir, W es una isometría parcial.

Si $T = UA$ para cualquier isometría parcial con $\ker U = \ker T$, entonces $U(A\boldsymbol{x}) = T\boldsymbol{x}$ para cada $\boldsymbol{x} \in V$ y también $U\boldsymbol{y} = \boldsymbol{0}$ si $\boldsymbol{y} \in \ker T$; como $A(V) \oplus \ker T = V$, se sigue que $U = W$. Esto establece la *unicidad* de la isometría parcial W que cumple (3.29).

Finalmente, T es invertible si y solo si T^*T es invertible y $\ker T = \{\boldsymbol{0}\}$, si y solo si T^*T es definido positivo y $\ker W = \{\boldsymbol{0}\}$, si y solo si A es definido positivo y la isometría parcial W es unitaria. □

Corolario 3.63. *Si $T \in \mathcal{L}(V)$, hay una única isometría parcial W' tal que $\underline{W'(V) = T(V)}$ que cumple $\underline{T = |T^*|W'}$.*

Demostración. Si $T^* = U|T^*|$ es la descomposición polar de T^*, tómese $W' := U^*$. Entonces $T = (T^*)^* = |T^*|^*U^* = |T^*|W'$ porque el operador positivo $|T^*|$ es autoadjunto. Además, vale $W'(V) = (\ker U)^\perp = (\ker T^*)^\perp = T(V)$.

La unicidad de la factorización $T = |T^*|W'$ es consecuencia de la unicidad de la descomposición polar $T^* = U|T^*|$. □

Escolio 3.64. *Un operador lineal $T \in \mathcal{L}(V)$ es normal si y solo si $|T^*| = |T|$, si y solo si los factores de su descomposición polar conmutan: $T = W|T| = |T|W$.* ⊟

► El teorema espectral no proporciona información directa acerca de las matrices ortogonales en $M_n(\mathbb{R})$, porque en el caso real no se puede asegurar la existencia de autovalores y autovectores *a priori*. Sin embargo, la diagonalizabilidad de las matrices unitarias con autovalores de valor absoluto 1 (véase la Proposición 3.60) posee una contraparte real, si en lugar de una matriz diagonal se acepta una suma directa de bloques 2×2.

Teorema 3.65. *Dada una matriz ortogonal $Q \in M_n(\mathbb{R})$, hay una colección de ángulos $\theta_1, \dots, \theta_t \in (0, \pi)$ y números $r, s \in \mathbb{N}$ tales que $2t + r + s = n$, amén de una matriz ortogonal P tal que*

$$P^{-1}QP = P^{\mathrm{t}}QP = R_{\theta_1} \oplus \cdots \oplus R_{\theta_t} \oplus (-1_r) \oplus 1_s \tag{3.31}$$

*donde cada R_{θ_j} es una **rotación** de la forma*

$$R_\theta := \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix}. \tag{3.32}$$

Demostración. La matriz ortogonal Q cumple $Q^{\mathrm{t}}Q = QQ^{\mathrm{t}} = 1_n$ en $M_n(\mathbb{R})$. Al considerar Q como elemento de $M_n(\mathbb{C})$ con entradas reales, Q es también unitaria en $M_n(\mathbb{C})$. Por el Teorema 3.56 y la Proposición 3.60, hay una base ortonormal $\mathcal{U}$ de $\mathbb{C}^n$ para la cual $[T_Q]^{\mathcal{U}}_{\mathcal{U}}$ es una matriz diagonal con autovalores $\lambda_i \in \mathbb{C}$ que cumplen $|\lambda_i| = 1$. Si la descomposición espectral de T_Q es $T_Q = \alpha_1 P_1 + \cdots + \alpha_l P_l$, entonces cada λ_i pertenece a $\{\alpha_1, \dots, \alpha_l\}$.

Los polinomios característicos de Q y de Q^{t} coinciden, así que los autovalores de $Q^* = Q^{\mathrm{t}}$ son también $\alpha_1, \dots, \alpha_l$. Un autovalor real α_j es necesariamente ± 1. Los otros autovalores son números complejos de valor absoluto 1, que forman pares conjugados $(\lambda_k, \bar{\lambda}_k) = (e^{i\theta_k}, e^{-i\theta_k})$ con $0 < \theta_k < \pi$.

Si $\boldsymbol{z} \in \mathbb{C}^n$ es un autovector de Q para un autovalor ± 1, sea $\boldsymbol{z} =: \boldsymbol{x} + i\boldsymbol{y}$ con $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Si $\boldsymbol{y} = \boldsymbol{0}$, entonces $\boldsymbol{z} = \boldsymbol{x} \in \mathbb{R}^n$; y si $\boldsymbol{x} = \boldsymbol{0}$, entonces $(-i)\boldsymbol{z} = \boldsymbol{y}$ es un autovector real para Q. En el caso $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{y} \neq \boldsymbol{0}$, se ve que $\mathbb{R}\text{-}\mathrm{lin}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ es un subespacio de autovectores de Q de dimensión 2 sobre $\mathbb{R}$. En todo caso, los subespacios $\ker(T_Q + 1) \subseteq \mathbb{R}^n$ y $\ker(T_Q - 1) \subseteq \mathbb{R}^n$ poseen bases ortonormales de autovectores reales para los autovalores respectivos -1 y $+1$. Sean $r := n(T_Q + 1)$, $s := n(T_Q - 1)$ sus dimensiones respectivas.

Ahora sea $\boldsymbol{z} \in \mathbb{C}^n$ un autovector para el autovalor complejo $e^{i\theta}$ con $0 < \theta < \pi$, con longitud $\|\boldsymbol{z}\| = \sqrt{2}$. Escríbase $\boldsymbol{z} =: \boldsymbol{x} + i\boldsymbol{y}$ con $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ y nótese que $\bar{\boldsymbol{z}} := \boldsymbol{x} - i\boldsymbol{y}$ es un autovector para el autovalor $e^{-i\theta}$ porque $Q\boldsymbol{z} = e^{i\theta}\boldsymbol{z}$ conlleva $Q\bar{\boldsymbol{z}} = e^{-i\theta}\bar{\boldsymbol{z}}$. Ahora

$$e^{i\theta}\langle \bar{\boldsymbol{z}}, \boldsymbol{z} \rangle = \langle \bar{\boldsymbol{z}}, Q\boldsymbol{z} \rangle = \langle Q^*\bar{\boldsymbol{z}}, \boldsymbol{z} \rangle = \langle Q^{-1}\bar{\boldsymbol{z}}, \boldsymbol{z} \rangle = \langle e^{i\theta}\bar{\boldsymbol{z}}, \boldsymbol{z} \rangle = e^{-i\theta}\langle \bar{\boldsymbol{z}}, \boldsymbol{z} \rangle,$$

así que $\langle \bar{\boldsymbol{z}}, \boldsymbol{z} \rangle = 0$ porque $e^{i\theta} \neq e^{-i\theta}$. Por lo tanto, vale

$$0 = \langle \bar{\boldsymbol{z}}, \boldsymbol{z} \rangle = \langle \boldsymbol{x} - i\boldsymbol{y}, \boldsymbol{x} + i\boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{x} \rangle - \langle \boldsymbol{y}, \boldsymbol{y} \rangle + 2i\, \langle \boldsymbol{x}, \boldsymbol{y} \rangle$$

y además $\langle \boldsymbol{x}, \boldsymbol{x} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle = \langle \boldsymbol{z}, \boldsymbol{z} \rangle = 2$. Se deduce que $\|\boldsymbol{x}\| = \|\boldsymbol{y}\| = 1$ y $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. En resumen, $\{\boldsymbol{x}, \boldsymbol{y}\}$ es una base ortonormal por el subespacio bidimensional $\mathbb{R}\text{-}\mathrm{lin}\langle \boldsymbol{z}, \bar{\boldsymbol{z}} \rangle$. Al tomar partes reales e imaginarios de ecuación $Q\boldsymbol{z} = e^{i\theta}\boldsymbol{z}$, se obtiene

$$\begin{aligned} Q\boldsymbol{x} &= \boldsymbol{x} \cos\theta - \boldsymbol{y} \operatorname{sen}\theta, \\ Q\boldsymbol{y} &= \boldsymbol{x} \operatorname{sen}\theta + \boldsymbol{y} \cos\theta. \end{aligned}$$

Luego $\mathrm{lin}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ es un subespacio invariante para Q, en donde la restricción de T_Q posee la matriz R_θ de (3.32).

Por una permutación de la base $\mathcal{U}$ de $\mathbb{C}^n$, los autovalores no reales de Q pueden ordenarse como $(e^{i\theta_1}, e^{-i\theta_1}, \dots, e^{i\theta_t}, e^{-i\theta_t})$ donde $2t = n - r - s$. Al repetir el proceso anterior para cada par $(e^{i\theta_k}, e^{-i\theta_k})$, se construye una nueva base ortonormal $\mathcal{V}$ de $\mathbb{R}^n$ para la cual la matriz de T_Q es el lado derecho de (3.31). La matriz de cambio de bases $P := [1]^{\mathcal{E}}_{\mathcal{V}}$ es ortogonal. □

3.7. Ejercicios sobre ortogonalidad y teoría espectral

Ejercicio 3.1. Encontrar el complemento ortogonal $M^\perp \subset \mathbb{R}^3$ del subespacio $M = \mathrm{lin}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ de $\mathbb{R}^3$ generado por

$$\boldsymbol{x} := \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \qquad \boldsymbol{y} := \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

⟦ Indicación: sea $A \in \mathbb{R}^{2\times 3}$ la matriz cuyas filas son $\boldsymbol{x}^t$ y $\boldsymbol{y}^t$; resolver la ecuación $A\boldsymbol{z} = \boldsymbol{0}$. ⟧

Ejercicio 3.2. Sea $W := \mathrm{lin}\langle \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \rangle$ el subespacio de $\mathbb{R}^4$ generado por los vectores

$$\boldsymbol{x} := \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \end{bmatrix}, \qquad \boldsymbol{y} := \begin{bmatrix} 2 \\ 1 \\ -1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{z} := \begin{bmatrix} 4 \\ 2 \\ 3 \\ 1 \end{bmatrix}.$$

Encontrar un vector $\boldsymbol{w} \in W$ tal que $\langle \boldsymbol{w}, \boldsymbol{x} \rangle = \langle \boldsymbol{w}, \boldsymbol{y} \rangle = 0$.

Ejercicio 3.3. Encontrar una base ortonormal para $U = \mathrm{lin}\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4 \rangle \leqslant \mathbb{R}^5$, donde

$$\boldsymbol{x}_1 := \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}, \qquad \boldsymbol{x}_2 := \begin{bmatrix} 2 \\ -1 \\ 0 \\ -1 \\ 1 \end{bmatrix}, \qquad \boldsymbol{x}_3 := \begin{bmatrix} 0 \\ 1 \\ 0 \\ -1 \\ 0 \end{bmatrix}, \qquad \boldsymbol{x}_4 := \begin{bmatrix} 2 \\ 1 \\ 1 \\ -1 \\ -1 \end{bmatrix}.$$

Ejercicio 3.4. El hiperplano $U := \{\, (x, y, z, w) \in \mathbb{R}^4 : x + y + z + w = 0 \,\}$ tiene dimensión 3. Encontrar (a) una base vectorial de U; y en seguida (b) una base ortonormal de U.

Ejercicio 3.5. En el espacio vectorial $\mathbb{R}[t]$ de polinomios reales, considérese la fórmula:

$$\langle p(t) \mid q(t) \rangle := \int_{-1}^{1} \frac{p(x)\, q(x)}{\sqrt{1 - x^2}}\, dx.$$

(a) Comprobar que este $\langle - \mid - \rangle$ es un producto escalar sobre $\mathbb{R}[t]$.

(b) El **polinomio de Chebyshev** $\underline{T_n(t)}$ es aquél polinomio de grado $n \in \mathbb{N}$ que cumple[16]

$$T_n(\cos\theta) \equiv \cos n\theta, \quad \text{para} \quad -\pi \leqslant \theta \leqslant \pi.$$

Demostrar que estos polinomios son ortogonales con respecto al producto escalar dado.

[16] Por ejemplo, $T_0(t) = 1$, $T_1(t) = t$, $T_2(t) = 2t^2 - 1$, $T_3(t) = 4t^3 - 3t$, etcétera.

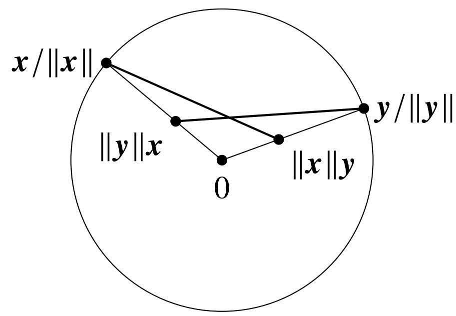

Figura 3.1: La fórmula de simetría

Ejercicio 3.6. Si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n \setminus \{0\}$, con normas $\|\boldsymbol{x}\| > 0$, $\|\boldsymbol{y}\| > 0$, comprobar esta *fórmula de simetría*:

$$\left\| \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|} - \|\boldsymbol{y}\|\boldsymbol{x} \right\| = \left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|} - \|\boldsymbol{x}\|\boldsymbol{y} \right\|.$$

⟦ Indicación: véase la Figura 3.1, ilustrando el círculo unitario en el subespacio $\operatorname{lin}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$. ⟧

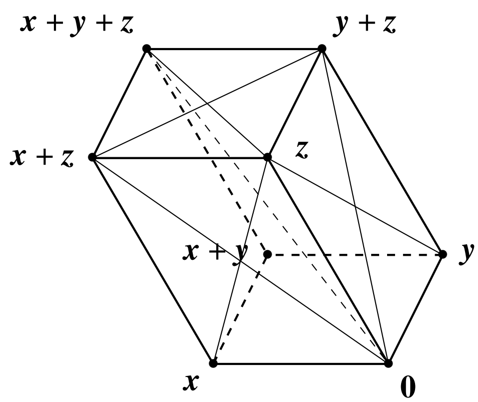

Figura 3.2: Un paralelepípedo con vértices $\mathbf{0}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$

Los tres ejercicios que siguen se refieren a una *norma* de vectores, que cumple las propiedades de la Proposición 3.9 y la ley del paralelogramo, Lema 3.10, pero *no necesariamente* proviene de un producto escalar.

Ejercicio 3.7. Si V es un espacio vectorial (real o complejo) con una *norma* que cumple la ley del paralelogramo, demostrar que siguiente relación, para $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$:

$$\|\boldsymbol{x} + \boldsymbol{y} + \boldsymbol{z}\|^2 + \|\boldsymbol{x}\|^2 + \|\boldsymbol{y}\|^2 + \|\boldsymbol{z}\|^2 = \|\boldsymbol{x} + \boldsymbol{y}\|^2 + \|\boldsymbol{y} + \boldsymbol{z}\|^2 + \|\boldsymbol{z} + \boldsymbol{x}\|^2. \qquad (*)$$

⟦ Indicación: considerar las facetas y las diagonales de un paralelepípedo (Figura 3.2) con vértices $\mathbf{0}, \boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}$. ⟧

Ejercicio 3.8. El teorema de Jordan y von Neumann (caso real) dice lo siguiente. *Si V es un espacio normado real cuya norma cumple la ley del paralelogramo, la fórmula que sigue define un producto escalar compatible con esa norma*:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \tfrac{1}{2}\big(\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}\|^2 - \|\boldsymbol{y}\|^2\big),$$

(a) Usar la fórmula $(*)$ del Ejercicio 3.7 para comprobar que $\langle \boldsymbol{x}, \boldsymbol{y}+\boldsymbol{z} \rangle = \langle \boldsymbol{x}, \boldsymbol{y} \rangle + \langle \boldsymbol{x}, \boldsymbol{z} \rangle$.

(b) Deducir que $\langle \boldsymbol{x}, a\boldsymbol{y} \rangle = a\,\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ para $a \in \mathbb{N}$, $a \in \mathbb{Z}$, $a \in \mathbb{Q}$, $a \in \mathbb{R}$ sucesivamente.

(c) Concluir que $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ cumple todas las propiedades de un producto escalar real en V; y que $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \|\boldsymbol{x}\|^2$.

Ejercicio 3.9. El teorema de Jordan y von Neumann (caso complejo) dice lo siguiente. *Si V es un espacio normado sobre $\mathbb{C}$ cuya norma cumple la ley del paralelogramo, las siguientes fórmulas definen un producto escalar compatible con esa norma*:

$$\Re\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \tfrac{1}{2}\big(\|\boldsymbol{x}+\boldsymbol{y}\|^2 - \|\boldsymbol{x}\|^2 - \|\boldsymbol{y}\|^2\big), \quad \Im\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \tfrac{1}{2}\big(\|\boldsymbol{x}-i\boldsymbol{y}\|^2 - \|\boldsymbol{x}\|^2 - \|\boldsymbol{y}\|^2\big).$$

(Es inmediato del Ejercicio 3.8 anterior que la parte real $\Re\langle -, - \rangle$ es un producto escalar real.)

Mostrar que $\Re\langle \boldsymbol{x}, i\boldsymbol{x} \rangle = 0$. Concluir que $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := \Re\langle \boldsymbol{x}, \boldsymbol{y} \rangle - i\,\Re\langle \boldsymbol{x}, i\boldsymbol{y} \rangle$ cumple todas las propiedades de un producto escalar complejo en V, tal que $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \|\boldsymbol{x}\|^2$.

Ejercicio 3.10. Encontrar una tercera columna de modo que la siguiente matriz A sea una matriz ortogonal:

$$A = \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{2}} & ? \\ \frac{1}{\sqrt{3}} & 0 & ? \\ \frac{1}{\sqrt{3}} & -\frac{1}{\sqrt{2}} & ? \end{bmatrix}.$$

Ejercicio 3.11. Decidir (con razonamiento) si la siguiente matriz es ortogonal o no:

$$B = \begin{bmatrix} \frac12 & \frac12 & \frac12 & \frac12 \\ \frac12 & \frac12 & -\frac12 & -\frac12 \\ \frac12 & -\frac12 & -\frac12 & \frac12 \\ \frac12 & -\frac12 & \frac12 & -\frac12 \end{bmatrix}.$$

Ejercicio 3.12. Demostrar que la siguiente matriz Q es una matriz ortogonal:

$$Q = \begin{bmatrix} \operatorname{sen}\theta \operatorname{sen}\phi & \operatorname{sen}\theta\cos\phi & \cos\theta\cos\phi & -\cos\theta\operatorname{sen}\phi \\ \operatorname{sen}\theta\cos\phi & -\operatorname{sen}\theta\operatorname{sen}\phi & \cos\theta\operatorname{sen}\phi & \cos\theta\cos\phi \\ \cos\theta\operatorname{sen}\phi & \cos\theta\cos\phi & -\operatorname{sen}\theta\cos\phi & \operatorname{sen}\theta\operatorname{sen}\phi \\ \cos\theta\cos\phi & -\cos\theta\operatorname{sen}\phi & -\operatorname{sen}\theta\operatorname{sen}\phi & -\operatorname{sen}\theta\cos\phi \end{bmatrix}.$$

Ejercicio 3.13. (a) Si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ son vectores de columna no nulos, demostrar que $\boldsymbol{x}\boldsymbol{y}^{\mathrm{t}}$ es una matriz en $M_n(\mathbb{R})$ cuyo rango es 1.

(b) Sea $\boldsymbol{x} \in \mathbb{R}^n$ con $\|\boldsymbol{x}\| = 1$. La **matriz de Householder** determinada por $\boldsymbol{x}$ es

$$H_{\boldsymbol{x}} := 1_n - 2\,\boldsymbol{x}\boldsymbol{x}^{\mathrm{t}}.$$

Mostrar que $H_{\boldsymbol{x}}$ es simétrica y ortogonal, con $H_{\boldsymbol{x}}^2 = 1_n$. ⟦ Indicación: $\boldsymbol{x}^{\mathrm{t}}\boldsymbol{x} = \|\boldsymbol{x}\|^2$. ⟧

Ejercicio 3.14. (a) Calcular los autovalores de la matriz

$$A := \begin{bmatrix} 5 & -6 & 2 \\ -6 & 4 & -4 \\ 2 & -4 & 0 \end{bmatrix}.$$

Encontrar una matriz *ortogonal* P cuyas columnas son autovectores de A, de modo que $P^{\mathrm{t}}AP = P^{-1}AP$ sea diagonal.

(b) Calcular A^5, usando esta forma diagonal $D = P^{\mathrm{t}}AP$.

Ejercicio 3.15. (a) Si U y V son matrices unitarias en $M_n(\mathbb{C})$, demostrar que el producto UV es también unitaria.

(b) Si $U \in M_N(\mathbb{C})$ es unitaria, demostrar que U es invertible y que U^{-1} es también unitaria.[17]

(c) Demostrar que $A \in M_n(\mathbb{C})$ es unitaria si y solo si $\|A\boldsymbol{x}\| = \|\boldsymbol{x}\|$ para todo $\boldsymbol{x} \in \mathbb{C}^n$.

Ejercicio 3.16. Determinar si cada una de estas matrices simétricas,

$$A := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \qquad B := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \qquad C := \begin{bmatrix} 9 & -6 & 2 \\ -6 & 8 & -4 \\ 2 & -4 & 4 \end{bmatrix},$$

es (a) positiva; y (b) definida positiva.

Ejercicio 3.17. (a) Sea $A := \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ una matriz simétrica en $M_2(\mathbb{R})$. Demostrar directamente que es posible factorizar $A = LDL^{\mathrm{t}}$, con L triangular inferior unipotente y D diagonal invertible, si y solo si $a \neq 0$ y $c \neq b^2/a$.

(b) Concluir que A es definida positiva si y solo si $a > 0$, $c > 0$ y $ac - b^2 > 0$.

[17] Las partes (a) y (b) dicen que la totalidad de matrices unitarias $n \times n$ es un *grupo*, llamado $\mathrm{U}(n)$.

(c) Obtener la factorización $A = LDL^{\mathrm{t}}$ para $A := \begin{bmatrix} 4 & 12 \\ 12 & 45 \end{bmatrix}$. Usar esta factorización para comprobar que

$$\boldsymbol{x}^{\mathrm{t}} A \boldsymbol{x} = (2x_1 + 6x_2)^2 + (3x_2)^2 \quad \text{si} \quad \boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2.$$

Ejercicio 3.18. Dadas m vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$, su **determinante de Gram** es

$$\det A = \begin{vmatrix} \boldsymbol{x}_1 \cdot \boldsymbol{x}_1 & \boldsymbol{x}_1 \cdot \boldsymbol{x}_2 & \dots & \boldsymbol{x}_1 \cdot \boldsymbol{x}_m \\ \boldsymbol{x}_2 \cdot \boldsymbol{x}_1 & \boldsymbol{x}_2 \cdot \boldsymbol{x}_2 & \dots & \boldsymbol{x}_2 \cdot \boldsymbol{x}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_m \cdot \boldsymbol{x}_1 & \boldsymbol{x}_m \cdot \boldsymbol{x}_2 & \dots & \boldsymbol{x}_m \cdot \boldsymbol{x}_m \end{vmatrix}.$$

Este es el determinante de la matriz $A = [a_{ij}]$ tal que $a_{ij} = \boldsymbol{x}_i \cdot \boldsymbol{x}_j$. Demostrar que $\det A = 0$ si y solo si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es linealmente dependiente, y que $\underline{\det A > 0}$ cuando $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es *linealmente independiente*.

⟦ Indicación: encontrar una matriz B tal que $A = B^{\mathrm{t}} B$. ⟧

Ejercicio 3.19. Si $J = \{j_1, j_2, \dots, j_r\} \subset \{1, 2, \dots, n\}$, el menor $m_{JJ} := \det A_{JJ}$ se llama un **menor principal** de la matriz $A \in M_n(\mathbb{F})$. Si $J_k = \{1, 2, \dots, k\}$, los menores principales $D_k := m_{J_k, J_k}$ para $k = 1, \dots, n$, se llaman los **menores principales delanteras** de la matriz cuadrada A:

$$D_1 = a_{11}, \quad D_2 = \begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix}, \quad D_3 = \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix}, \quad \dots, \quad D_n = \det A.$$

(a) Si la matriz $A \in M_n(\mathbb{R})$ es definida positiva, demostrar que todos sus menores principales delanteras son números positivos.

(b) Inversamente, si $A \in M_n(\mathbb{R})$ es simétrica, con $D_k > 0$ para $k = 1, 2, \dots, n$, demostrar que la matriz A es definida positiva. ⟦ Indicación: eliminación gaussiana simple. ⟧

Ejercicio 3.20. Considérese la siguiente matriz $A \in M_8(\mathbb{R})$:

$$A := \begin{bmatrix} b & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & b & 0 & -1 & 0 & 0 & -1 & 0 \\ -1 & 0 & c & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & c & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & b & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & b & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 0 & c & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & c \end{bmatrix}.$$

Demostrar que esta matriz[18] es definida positiva si y solo si $b > 0$, $c > 0$ y $bc > 4$.

[18] La matriz apareció en un problema de la mecánica cuántica, en la tesis de licenciatura: Ileana Castillo A., *Productos cuánticos en espacios de funciones analíticas*, UCR, 1988.

En los cinco ejercicios que siguen, V es un espacio de Hilbert de dimensión finita n.

Ejercicio 3.21. (a) Si $S \in \mathcal{L}(V)$ es un operador autoadjunto tal que $\langle \boldsymbol{x}, S\boldsymbol{x} \rangle = 0$ para todo $\boldsymbol{x} \in V$, demostrar que $S = 0$. ⟦ Indicación: Lema 3.34. ⟧

(b) Si además $T \in \mathcal{L}(V)$ es un operador lineal cualquiera, demostrar que $T = T^*$ si y solo si $\langle \boldsymbol{x}, T\boldsymbol{x} \rangle \in \mathbb{R}$ para todo $\boldsymbol{x} \in V$. ⟦ Indicación: considerar $S := i(T^* - T)$. ⟧

Ejercicio 3.22. (a) Si $S \in \mathcal{L}(V)$ es autoadjunto, usar el teorema espectral para obtener dos operadores positivos $S_+, S_- \in \mathcal{L}(V)$ tales que $S_+S_- = S_-S_+ = 0$ y $S = \underline{S_+ - S_-}$.

(b) Concluir que cualquier operador lineal $T \in \mathcal{L}(V)$ es una combinación lineal de la forma

$$T = \underline{T_1 - T_2 + iT_3 - iT_4}$$

donde las "cuatro partes" T_1, T_2, T_3, T_4 son operadores positivos.

Ejercicio 3.23. (a) Si $T \in \mathcal{L}(V)$ es un operador normal, demostrar que hay un polinomio $p(t) \in \mathbb{C}[t]$ tal que $p(T) = T^*$. ⟦ Indicación: buscar un polinomio $p(t)$ que cumple $p(\alpha_j) = \bar{\alpha}_j$ para cada autovalor α_j de T. ⟧

(b) Si $S \in \mathcal{L}(V)$ es otro operador lineal (no necesariamente normal) tal que $ST = TS$, demostrar que $S^*T = TS^*$.

(c) Si $S, T \in \mathcal{L}(V)$ son normales y $ST = TS$, concluir que el producto ST es también normal.

Ejercicio 3.24. (a) Si $T \in \mathcal{L}(V)$ es un operador positivo, demostrar que hay un polinomio $q(t) \in \mathbb{R}[t]$ tal que $q(T) = T^{1/2}$. ⟦ Indicación: buscar un polinomio $q(t)$ que cumple $q(\alpha_j) = \sqrt{\alpha_j}$ para cada autovalor α_j de T. ⟧

(b) Si $S \in \mathcal{L}(V)$ es otro operador positivo tal que $ST = TS$, demostrar que los operadores $S + T$ y ST son también positivos.

(c) Si $Q, R \in \mathcal{L}(V)$ son operadores positivos que *no conmutan*, demostrar que $Q + R$ es positivo pero QR no es necesariamente positivo. ⟦ Indicación: dar un contraejemplo de dos matrices positivas cuyo producto no es una matriz positiva. ⟧

Ejercicio 3.25. (a) Demostrar que $U \in \mathcal{L}(V)$ es unitario si y solo si $\|U\boldsymbol{x}\| = \|\boldsymbol{x}\|$ para todo $\boldsymbol{x} \in V$.

(b) Demostrar que $T \in \mathcal{L}(V)$ es normal si y solo si $\|T\boldsymbol{x}\| = \|T^*\boldsymbol{x}\|$ para todo $\boldsymbol{x} \in V$.

(c) Mostrar que $T \in \mathcal{L}(V)$ es normal si y solo si $\langle T\boldsymbol{y}, T\boldsymbol{x} \rangle = \langle T^*\boldsymbol{y}, T^*\boldsymbol{x} \rangle$ para todo $\boldsymbol{x}, \boldsymbol{y} \in V$.

Ejercicio 3.26. Sea $A \in M_n(\mathbb{C})$ y sean $\lambda_1, \dots, \lambda_n$ los autovalores de A, repetidos según su multiplicidad. Demostrar la desigualdad:

$$\operatorname{tr}(A^*A) \geqslant |\lambda_1|^2 + \cdots + |\lambda_n|^2,$$

con igualdad si y solo si A es una matriz normal, $A^*A = AA^*$. ⟦ Indicación: usar la fórmula (3.25) para calcular $\operatorname{tr}(A^*A)$. ⟧

Ejercicio 3.27. Obtener la descomposición polar QR de la matriz $B = \begin{bmatrix} -1 & -2 \\ 2 & 1 \end{bmatrix} \in M_2(\mathbb{C})$, así: (a) hallar la matriz positiva R tal que $R^2 = B^*B$; y (b) calcular $Q := BR^{-1}$.

Ejercicio 3.28. Sea $\zeta := e^{2\pi i/n}$ una **raíz n-ésima de** 1 (compleja)[19] y sea $F \in M_n(\mathbb{C})$ la matriz con entradas $f_{jk} := \zeta^{(j-1)(k-1)}/\sqrt{n}$:

$$F := \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \zeta & \zeta^2 & \dots & \zeta^{n-1} \\ 1 & \zeta^2 & \zeta^4 & \dots & \zeta^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \zeta^{n-1} & \zeta^{2(n-1)} & \dots & \zeta^{(n-1)^2} \end{bmatrix}.$$

La matriz F se llama la **transformación de Fourier finita** de orden n.

(a) Escribir F explícitamente en los casos $n = 2, 3, 4$. ⟦ Indicación: $\omega := e^{2\pi i/3}$. ⟧

(b) Demostrar que $F^*F = 1_n$ y concluir que la matriz F es *unitaria*. ⟦ Indicación: notar que $\bar{\zeta} = e^{-2\pi i/n} = \zeta^{-1}$. ⟧

(c) Calcular la matriz F^2 y mostrar que $F^4 = 1_n$. Deducir que los únicos autovalores posibles[20] para F son $\lambda = 1, -1, i, -i$.

[19] El número complejo $\zeta := e^{2\pi i/n} = \cos\frac{2\pi}{n} + i \operatorname{sen}\frac{2\pi}{n}$ cumple $\zeta^n = 1$. Sus potencias $\zeta, \zeta^2, \zeta^3, \dots, \zeta^{n-1}$ son otras raíces n-ésimas de 1, porque $(\zeta^k)^n = (\zeta^n)^k = 1^k = 1$ para $k = 0, 1, 2, \dots, n-1$. (Como $\zeta^0 = \zeta^n = 1$, hay exactamente n raíces n-ésimas de 1, las soluciones complejas distintas de la ecuación polinomial $z^n - 1 = 0$.)

[20] La *multiplicidad* de cada uno de estos autovalores, para n cualquiera, es “aproximadamente $n/4$”. La determinación exacta de las multiplicidades es un problema de la teoría de números. Por ejemplo, si $n = 4m + 1$ con $m \in \mathbb{N}$, se sabe que $\lambda = 1$ ocurre $(m + 1)$ veces, y que $\lambda = -1, i, -i$ ocurren m veces cada uno.

4 Formas bilineales

I vari rami della Matematica pura ed applicata si annodano e si collegano fra loro per vie inaspettate; e le idee, che traggono origine da elementari problemi della pratica, sembra debbano maturarsi per una lunga elaborazione di pensiero, nelle regioni più alte della teoria, prima che possano discendere feconde nel campo di attività della vita.

— Federigo Enriques

Las aplicaciones lineales entre dos espacios vectoriales son representadas por matrices (en la presencia de un par de bases). En este capítulo se examinará otros objetos, las *formas bilineales* sobre un espacio vectorial, que generalizan el concepto de producto escalar real. A cada forma bilineal se le asocia una matriz que la representa con respecto a una base dada.

Se distinguen tres clases de tales formas: (a) las formas bilineales simétricas, (b) las formas bilineales alternadas o antisimétricas, y en el caso complejo (c) las formas *sesquilineales* hermíticas. Las matrices asociadas son respectivamente simétricas, antisimétricas, o hermíticas. Se requiere clasificar estas formas hasta isomorfismo, por una cierta relación de equivalencia entre sus matrices.

4.1. Formas bilineales y sus matrices

Definición 4.1. Sea V un espacio vectorial finitodimensional sobre un cuerpo $\mathbb{F}$ cualquiera. Se dice que una aplicación $f\colon V \times V \to \mathbb{F}$ es una **forma bilineal** si las aplicaciones parciales $\boldsymbol{x} \mapsto f(\boldsymbol{x}, \boldsymbol{y})$; $\boldsymbol{y} \mapsto f(\boldsymbol{x}, \boldsymbol{y})$ son formas lineales sobre $\mathbb{F}$. En otras palabras, se debe cumplir

$$\begin{aligned} f(\boldsymbol{x}+\boldsymbol{z}, \boldsymbol{y}) &= f(\boldsymbol{x}, \boldsymbol{y}) + f(\boldsymbol{z}, \boldsymbol{y}), \\ f(\boldsymbol{x}, \boldsymbol{y}+\boldsymbol{w}) &= f(\boldsymbol{x}, \boldsymbol{y}) + f(\boldsymbol{x}, \boldsymbol{w}), \\ f(c\boldsymbol{x}, \boldsymbol{y}) &= f(\boldsymbol{x}, c\boldsymbol{y}) = c\, f(\boldsymbol{x}, \boldsymbol{y}), \end{aligned} \tag{4.1}$$

para todo $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}, \boldsymbol{w} \in V$; $c \in \mathbb{F}$. En consecuencia,

$$f\biggl(\sum_{i=1}^n c_i \boldsymbol{x}_i, \sum_{j=1}^n d_j \boldsymbol{x}_j\biggr) = \sum_{i=1}^n \sum_{j=1}^n c_i d_j\, f(\boldsymbol{x}_i, \boldsymbol{x}_j). \qquad \Diamond$$

Ejemplo 4.2. El producto punto de vectores $f(\boldsymbol{c}, \boldsymbol{d}) := \underline{\boldsymbol{c}^{\mathsf{t}}\boldsymbol{d}} = c_1 d_1 + \cdots + c_n d_n$ es una forma bilineal sobre $\mathbb{F}^n$. Más generalmente, si $A \in M_n(\mathbb{F})$ es una matriz cuadrada cualquiera, la siguiente receta define una forma bilineal sobre $\mathbb{F}^n$.

$$\boxed{f(\boldsymbol{c}, \boldsymbol{d}) := \boldsymbol{c}^{\mathsf{t}} A \boldsymbol{d}.} \qquad \Diamond$$

Definición 4.3. En vista de las fórmulas (4.1), la función f está completamente determinada por los valores $f(\boldsymbol{x}_i, \boldsymbol{x}_j)$ con los vectores $\boldsymbol{x}_i, \boldsymbol{x}_j$ en una *base* $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ del espacio vectorial V. Denótese estos **coeficientes** de f por

$$\boxed{a_{ij} := f(\boldsymbol{x}_i, \boldsymbol{x}_j),} \quad \text{para} \quad i, j = 1, \dots, n. \tag{4.2}$$

La bilinealidad de f implica que

$$f\biggl(\sum_{i=1}^{n} c_i \boldsymbol{x}_i, \sum_{j=1}^{n} d_j \boldsymbol{x}_j\biggr) = \sum_{i,j=1}^{n} c_i\, a_{ij}\, d_j.$$

Los escalares a_{ij} son entradas de una matriz cuadrada $[f]_{\mathcal{B}}^{\mathcal{B}} := A \in M_n(\mathbb{F})$: esta es la **matriz de la forma bilineal** f con respecto a la base $\mathcal{B}$. ◊

La bilinealidad de f expresada en las fórmulas (4.1) puede ser reformulada de la siguiente manera. En primer lugar, para cada vector fijo $\boldsymbol{y} \in V$, la aplicación $\boldsymbol{x} \mapsto f(\boldsymbol{x}, \boldsymbol{y})$ es una forma lineal sobre V, o sea, es un elemento $f_{\boldsymbol{y}}$ del espacio dual V^*. Después, por la linealidad en la segunda variable, si $\boldsymbol{y} = d_1\boldsymbol{x}_1 + \cdots + d_n\boldsymbol{x}_n$ entonces $f_{\boldsymbol{y}}(\boldsymbol{x}) = d_1 f_{\boldsymbol{x}_1}(\boldsymbol{x}) + \cdots + d_n f_{\boldsymbol{x}_n}(\boldsymbol{x})$ para todo $\boldsymbol{x} \in V$. Esto implica que la función $F \colon V \to V^* : \boldsymbol{y} \mapsto f_{\boldsymbol{y}}$ es una aplicación lineal, y por ende $F \in \mathcal{L}(V, V^*)$. (Nótese que $a_{ij} = f_{\boldsymbol{x}_j}(\boldsymbol{x}_i)$ por (4.2).)

Si $\mathcal{B}^* := \{f_1, \dots, f_n\}$ es la base (1.5) de V^* dual a la base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, entonces la matriz (1.9) de F, con respecto a estas bases de V y V^*, se obtiene de

$$F(\boldsymbol{x}_j) := f_{\boldsymbol{x}_j} = \sum_{i=1}^{n} f_{\boldsymbol{x}_j}(\boldsymbol{x}_i)\, f_i = \sum_{i=1}^{n} a_{ij} f_i.$$

Esto dice que la matriz de F es $[F]_{\mathcal{B}}^{\mathcal{B}^*} = A = [f]_{\mathcal{B}}^{\mathcal{B}}$. En otras palabras, la matriz de la forma bilineal f con respecto a la base $\mathcal{B}$ coincide con la matriz de la aplicación lineal $F \in \mathcal{L}(V, V^*)$, con respecto a las bases $\mathcal{B}$ y $\mathcal{B}^*$, en la terminología de la Definición 1.28.

La moraleja de esta historia es que las formas bilineales sobre V son miembros de un *espacio vectorial*, isomorfo al espacio vectorial $\mathcal{L}(V, V^*)$ mediante la correspondencia $f \mapsto F$ (que también es lineal!). Su dimensión es $(\dim V)(\dim V^*) = (\dim V)^2 = n^2 = \dim M_n(\mathbb{F})$.

► La diferencia sutil entre la asignación de matrices $A \in M_n(\mathbb{F})$ a operadores lineales en $\mathcal{L}(V, V)$ y formas bilineales en $\mathcal{L}(V, V^*)$ no es una mera complicación de notación. Tiene un efecto profundo sobre la reasignación de matrices bajo un cambio de la base $\mathcal{B}$ de V. En el caso de operadores lineales, un cambio de base mediante la matriz invertible $P = [1_V]_{\mathcal{B}'}^{\mathcal{B}}$, modifica la matriz A en una *matriz semejante*: $A \mapsto P^{-1}AP$. En el caso de formas bilineales, el ajuste de matrices viene dado por la Proposición siguiente.

Proposición 4.4. *Sean* $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ *y* $\mathcal{B}' = \{\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_n\}$ *dos bases de* V*; y sea* $P \in M_n(\mathbb{F})$ *la matriz de cambio de base* (1.12)*, dada por* $\boldsymbol{x}'_s =: \sum_{j=1}^n p_{js}\boldsymbol{x}_j$*. Si* $f\colon V \times V \to \mathbb{F}$ *es una forma bilineal, las matrices respectivas* $A = [f]_{\mathcal{B}}^{\mathcal{B}}$ *y* $B = [f]_{\mathcal{B}'}^{\mathcal{B}'}$ *de* f *cumplen la relación:*

$$\boxed{B = P^{\mathrm{t}}AP.} \tag{4.3}$$

Demostración. Las entradas de A están dadas por (4.2); análogamente, las entradas de B son $b_{rs} := f(\boldsymbol{x}'_r, \boldsymbol{x}'_s)$. Al usar la bilinealidad de f, se calcula:

$$b_{rs} = f\biggl(\sum_{i=1}^n p_{ir}\boldsymbol{x}_i, \sum_{j=1}^n p_{js}\boldsymbol{x}_j\biggr) = \sum_{i,j=1}^n p_{ir}\, f(\boldsymbol{x}_i, \boldsymbol{x}_j)\, p_{js} = \sum_{i,j=1}^n p_{ir}a_{ij}p_{js},$$

y se reconoce el lado derecho como la entrada (r, s) de la matriz $P^{\mathrm{t}}AP$. □

Definición 4.5. Dos matrices cuadradas $A, B \in M_n(\mathbb{F})$ se llaman **matrices congruentes**, escrito $A \doteq B$, si hay una matriz invertible $P \in M_n(\mathbb{F})$ tal que $\underline{B = P^{\mathrm{t}}AP}$. ◊

La igualdad $(P^{-1})^{\mathrm{t}} = (P^{\mathrm{t}})^{-1}$ implica que la congruencia de matrices es una relación de equivalencia.[1] En efecto, esta relación es reflexiva, porque $A = 1_n^{\mathrm{t}}A1_n$; es antisimétrica, porque $B = P^{\mathrm{t}}AP$ implica $A = (P^{-1})^{\mathrm{t}}BP^{-1}$; y es transitiva, porque $B = P^{\mathrm{t}}AP$ y $C = Q^{\mathrm{t}}BQ$ implican $C = (PQ)^{\mathrm{t}}A(PQ)$. La Proposición 4.4 dice que dos matrices que representan la misma forma bilineal con respecto a dos bases distintas son congruentes, mediante la matriz P de cambio de base.

Nótese que dos matrices congruentes *tienen el mismo rango*. Es sabido que el rango es invariante bajo cambios $A \mapsto QAP$ con Q, P invertibles – véase la discusión después de la Definición 1.31. Al tomar $Q = P^{\mathrm{t}}$, se ve que $\underline{r(A) = r(B)}$ cuando $A \doteq B$. En consecuencia, el rango de la matriz de una forma bilineal no depende de la base $\mathcal{B}$ de V: se puede definir el **rango de la forma bilineal** f como el rango $r(A)$ de cualquiera de sus matrices.

Definición 4.6. Una forma bilineal $f\colon V \times V \to \mathbb{F}$ es **no degenerada** si $f(\boldsymbol{x}, \boldsymbol{y}) = 0$ para todo $\boldsymbol{x} \in V$ implica $\boldsymbol{y} = \boldsymbol{0}$. ◊

Lema 4.7. *Para una forma bilineal* $f\colon V \times V \to \mathbb{F}$*, las siguientes condiciones son equivalentes:*

(a) f *es una forma bilineal no degenerada;*

(b) $f_{\boldsymbol{y}} = 0$ *en* V^* *implica que* $\boldsymbol{y} = \boldsymbol{0}$ *en* V*;*

(c) *la aplicación lineal asociada* $F \in \mathcal{L}(V, V^*)$ *es inyectiva;*

[1] Algunos autores escriben $\underline{A^{-\mathrm{t}}}$ para denotar la matriz $(A^{-1})^{\mathrm{t}} = (A^{\mathrm{t}})^{-1}$, por un abuso de las leyes de exponentes. Fíjese que $(AB)^{-\mathrm{t}} = A^{-\mathrm{t}}B^{-\mathrm{t}}$ si A, B son invertibles. $A^{-\mathrm{t}}$ se llama la **matriz contragrediente** de A.

(d) *la aplicación lineal asociada* $F \in \mathcal{L}(V, V^*)$ *es sobreyectiva;*

(e) *la aplicación lineal asociada* $F \in \mathcal{L}(V, V^*)$ *es biyectiva;*

(f) *si* $\mathcal{B}$ *es una base de* V*, la matriz* $A = [f]^{\mathcal{B}}_{\mathcal{B}}$ *es invertible.*

Demostración. Ad (a) $\iff$ (b): es inmediato, porque $f_{\boldsymbol{y}}(\boldsymbol{x}) := f(\boldsymbol{x}, \boldsymbol{y})$.

Ad (b) $\iff$ (c): es también inmediato, de la definición $F(\boldsymbol{y}) := f_{\boldsymbol{y}}$.

Ad (c) $\iff$ (d): Como $\dim V = \dim V^* = n$ es finita, el teorema de rango y nulidad dice que $r(F) = n$ si y solo si $n(F) = 0$; en otras palabras, F es sobreyectiva si y solo si F es inyectiva.

Ad (d) $\iff$ (e): se sigue por la misma razón.

Ad (e) $\iff$ (f): viene de la relación $A = [f]^{\mathcal{B}}_{\mathcal{B}} = [F]^{\mathcal{B}^*}_{\mathcal{B}}$. □

⟦ La Definición 4.6 puede parecer asimétrica. Al cambiar los papeles de los vectores $\boldsymbol{x}, \boldsymbol{y}$, se podría leer: "$f$ es no degenerada si $f(\boldsymbol{x}, \boldsymbol{y}) = 0$ para todo $\boldsymbol{y} \in V$ implica $\boldsymbol{x} = \boldsymbol{0}$". Para examinar esta cuestión, defínase $\underline{g(\boldsymbol{x}, \boldsymbol{y}) := f(\boldsymbol{y}, \boldsymbol{x})}$. De la fórmula (4.2) se ve enseguida que la matriz de g es la *transpuesta* de la matriz de f: $[g]^{\mathcal{B}}_{\mathcal{B}} = [a_{ji}] = A^{\mathrm{t}}$. Pero A^{t} es invertible si y solo si A es invertible, con $(A^{\mathrm{t}})^{-1} = (A^{-1})^{\mathrm{t}}$. El inciso (f) del Lema 4.7 entonces muestra que esta definición alternativa de no degeneración es equivalente a la original. ⟧

La sobreyectividad $F(V) = V^*$, que significa $\{\, f_{\boldsymbol{y}} : \boldsymbol{y} \in V \,\} = V^*$, dice que una forma bilineal f es no degenerada si y solo si cualquier elemento del espacio dual V^* está dado por $\underline{\boldsymbol{x} \mapsto f(\boldsymbol{x}, \boldsymbol{y})}$ para algún $\boldsymbol{y} \in V$.

4.2. Formas bilineales simétricas

Definición 4.8. Una forma bilineal $d : V \times V \to \mathbb{F}$ es **simétrica** si $\underline{d(\boldsymbol{x}, \boldsymbol{y}) = d(\boldsymbol{y}, \boldsymbol{x})}$ para todo $\boldsymbol{x}, \boldsymbol{y} \in V$.

Una forma bilineal $s : V \times V \to \mathbb{F}$ es **alternada** o **antisimétrica** si $\underline{s(\boldsymbol{x}, \boldsymbol{y}) = -s(\boldsymbol{y}, \boldsymbol{x})}$ para todo $\boldsymbol{x}, \boldsymbol{y} \in V$. ◊

Es evidente de (4.2) que una forma bilineal d es simétrica si y solo si su matriz $A = [d]^{\mathcal{B}}_{\mathcal{B}}$ es simétrica: $\underline{A^{\mathrm{t}} = A}$. Del mismo modo, una forma bilineal s es alternada si y solo si su matriz $B = [s]^{\mathcal{B}}_{\mathcal{B}}$ es antisimétrica: $\underline{B^{\mathrm{t}} = -B}$.

La relación de congruencia (4.3) mantiene estas condiciones:

$$
\begin{aligned}
(P^{\mathrm{t}}AP)^{\mathrm{t}} &= P^{\mathrm{t}}A^{\mathrm{t}}P = P^{\mathrm{t}}AP, \\
(P^{\mathrm{t}}BP)^{\mathrm{t}} &= P^{\mathrm{t}}B^{\mathrm{t}}P = -P^{\mathrm{t}}BP.
\end{aligned}
$$

Existen cuerpos $\mathbb{F}$ en los cuales $-1 = +1$, o equivalentemente, $1 + 1 = 0$; en cuyo caso la distinción entre formas simétricas y alternadas carece de sentido. El ejemplo más conocido es el cuerpo de dos elementos $\mathbb{F}_2 = \{\overline{0}, \overline{1}\}$ que es la base de la aritmética binaria.[2] Para dejar este caso de lado, en esta sección se asumirá que $1 + 1 \neq 0$ en $\mathbb{F}$.

Lema 4.9. *Si $1+1 \neq 0$ en $\mathbb{F}$, cada forma bilineal $f : V \times V \to \mathbb{F}$ es la suma, de manera única, de una forma bilineal simétrica y una forma bilineal alternada.*

Demostración. Se puede definir una forma simétrica d y una forma alternada s por

$$\begin{aligned} d(\boldsymbol{x},\boldsymbol{y}) &:= \tfrac{1}{2}\big(f(\boldsymbol{x},\boldsymbol{y}) + f(\boldsymbol{y},\boldsymbol{x})\big), \\ s(\boldsymbol{x},\boldsymbol{y}) &:= \tfrac{1}{2}\big(f(\boldsymbol{x},\boldsymbol{y}) - f(\boldsymbol{y},\boldsymbol{x})\big). \end{aligned} \tag{4.4}$$

Es evidente que $f(\boldsymbol{x},\boldsymbol{y}) = d(\boldsymbol{x},\boldsymbol{y}) + s(\boldsymbol{x},\boldsymbol{y})$.

Para la unicidad de d y s, basta observar que si $f = d' + s'$ es otra suma de este tipo, entonces $d - d' = s' - s$; esta es una igualdad entre una forma simétrica y otra alternada, lo cual solo es posible si ambas formas son nulas: $d - d' = s' - s \equiv 0$; luego, $d' = d$ y $s' = s$. □

Definición 4.10. Sea $d : V \times V \to \mathbb{F}$ una forma bilineal simétrica. A cada subespacio $M \leqslant V$ le corresponde un **subespacio ortogonal** *con respecto a d*:

$$\begin{aligned} M^\perp &= \{\, \boldsymbol{x} \in V : d(\boldsymbol{x},\boldsymbol{y}) = 0 \ \text{ para todo } \boldsymbol{y} \in M \,\} \\ &= \{\, \boldsymbol{y} \in V : d(\boldsymbol{x},\boldsymbol{y}) = 0 \ \text{ para todo } \boldsymbol{x} \in M \,\}. \end{aligned} \tag{4.5}$$

Debe estar claro que si $N \leqslant M \leqslant V$, entonces $M^\perp \leqslant N^\perp$; y que $\{\mathbf{0}\}^\perp = V$.

El subespacio $V^\perp = \{\, \boldsymbol{y} \in V : d_{\boldsymbol{y}} = 0 \,\} = \ker D$ es el **núcleo** de la forma bilineal d. ⟦ Aquí $D \in \mathcal{L}(V, V^*)$ es la aplicación lineal asociada con la forma d. ⟧ Fíjese que d es no degenerada si y solo si $\ker D = \{0\}$ (por el Lema 4.7), si y solo si $V^\perp = \{\mathbf{0}\}$. ◊

Proposición 4.11. *Si d es una forma bilineal simétrica* no degenerada *sobre V, y si M es un subespacio de V tal que $M \cap M^\perp = \{\mathbf{0}\}$, entonces $M \oplus M^\perp = V$.*

Demostración. Al considerar los valores $d(\boldsymbol{x},\boldsymbol{y})$ con $\boldsymbol{x} \in M$ solamente, $\boldsymbol{y} \in V$ cualquiera, se ve que $\boldsymbol{x} \mapsto d(\boldsymbol{x},\boldsymbol{y})$ es un elemento $d'_{\boldsymbol{y}}$ de M^* y que $\boldsymbol{y} \mapsto d'_{\boldsymbol{y}}$ es una aplicación lineal $D_M \in \mathcal{L}(V, M^*)$. De (4.5) se ve que $M^\perp = \{\, \boldsymbol{y} \in V : d'_{\boldsymbol{y}} = 0 \,\} = \ker D_M$.

Por otro lado, si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es una base de M y $\{g_1, \dots, g_m\}$ es la base dual de M^*, se puede completar la primera para obtener una base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, cuya base dual

[2] En cualquier cuerpo $\mathbb{F}$, sucede una de dos cosas: (a) una suma finita de copias del elemento identidad 1 es igual a 0; o bien (b) ninguna suma finita de copias de 1 es 0. En el caso (a), sea p el menor entero $\geqslant 2$ tal que $1 + 1 + \cdots + 1$ (p veces) es 0. Resulta que p es *primo*; por ejemplo, si $1 + 1 + 1 + 1 + 1 + 1 = (1 + 1)(1 + 1 + 1) = 0$, entonces $1 + 1 = 0$ o bien $1 + 1 + 1 = 0$, poniendo de manifiesto que p no puede ser un número compuesto; este p es la **característica** del cuerpo $\mathbb{F}$. En el caso (b), dícese que $\mathbb{F}$ es *de característica* 0; en cuyo caso, $\mathbb{F}$ incluye una copia de $\mathbb{N}$ y por ende una copia de $\mathbb{Q}$. Los cuerpos finitos, entonces, son de característica p para algún p primo.

de V^* es $\mathcal{B}^* = \{f_1, \dots, f_n\}$. Para $j = 1, \dots, m$, la fórmula $f_j(\boldsymbol{x}_k) = [\![j = k]\!]$ implica que $f_j(\boldsymbol{x}) = g_j(\boldsymbol{x})$ para todo $\boldsymbol{x} \in M$.

Del Lema 4.7 se obtiene $f_j = d_{\boldsymbol{y}_j}$ para algún $\boldsymbol{y}_j \in V$ y por ende $g_j = d'_{\boldsymbol{y}_j} = D_M(\boldsymbol{y}_j)$. Se deduce que la imagen de D_M es todo M^*. Del teorema de rango y nulidad se concluye que

$$\dim V = n(D_M) + r(D_M) = \dim M^\perp + \dim M^* = \dim M^\perp + \dim M.$$

La condición $M \cap M^\perp = \{\boldsymbol{0}\}$ ahora implica $\dim V = \dim(M \oplus M^\perp)$, así que $M \oplus M^\perp = V$. □

► Los cálculos explícitos con formas bilineales simétricas $d(\boldsymbol{x}, \boldsymbol{y})$ se simplifican considerablemente al considerar el caso $\boldsymbol{x} = \boldsymbol{y}$. Cómo se verá en seguida, se puede recuperar la forma bilineal d a partir de tales valores.

Definición 4.12. Sea $d \colon V \times V \to \mathbb{F}$ una forma bilineal simétrica. La **forma cuadrática** asociada a d es la función $q \colon V \to \mathbb{F}$ dada por $\underline{q(\boldsymbol{x}) := d(\boldsymbol{x}, \boldsymbol{x})}$. ◊

Ejemplo 4.13. Una forma cuadrática sobre $\mathbb{F}^n$ es una función cuadrática de n variables:

$$q(\boldsymbol{x}) = \sum_{i,j=1}^{n} a_{ij}\, x_i x_j,$$

donde $A \in M_n(\mathbb{F})$ es una matriz de coeficientes. Obviamente es posible combinar los términos

$$a_{ij}\, x_i x_j + a_{ji}\, x_j x_i = (a_{ij} + a_{ji}) x_i x_j.$$

Entonces la función q no cambia si se reemplaza cada a_{ij} por $\frac{1}{2}(a_{ij} + a_{ji})$; por eso, *se puede suponer que A es una matriz simétrica*, $A^t = A$. Hecho eso, se ve que $q(\boldsymbol{x}) = d(\boldsymbol{x}, \boldsymbol{x})$ donde $d(\boldsymbol{x}, \boldsymbol{y}) := \sum_{i,j=1}^{n} a_{ij} x_i y_j$ es una forma bilineal simétrica. ◊

Ejemplo 4.14. Una **superficie cuádrica** en $\mathbb{F}^3$, centrada en el origen, tiene una ecuación de la forma $\underline{q(x, y, z) = 1}$, donde

$$\boxed{q(x, y, z) := ax^2 + 2hxy + by^2 + 2gxz + 2fyz + cz^2} = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}. \qquad ◊$$

La Definición 4.12 dice cómo obtener una forma cuadrática q a partir de una forma bilineal simétrica d. Inversamente, es posible recuperar d a partir de q por el proceso de *polarización* – el mismo procedimiento[3] que permitió reconstruir un producto escalar real desde una norma cuadrada en el Lema 3.11.

[3]Con una fórmula levemente distinta: en el contexto actual, no se dispone de una ley del paralelogramo.

Proposición 4.15. *Una forma bilineal simétrica d es determinada por su forma cuadrática asociada, mediante*

$$\boxed{d(\boldsymbol{x}, \boldsymbol{y}) = \tfrac{1}{2}\big(q(\boldsymbol{x} + \boldsymbol{y}) - q(\boldsymbol{x}) - q(\boldsymbol{y})\big).} \tag{4.6}$$

Demostración. Por un cálculo directo, se obtiene

$$\begin{aligned} q(\boldsymbol{x} + \boldsymbol{y}) - q(\boldsymbol{x}) - q(\boldsymbol{y}) &= d(\boldsymbol{x} + \boldsymbol{y}, \boldsymbol{x} + \boldsymbol{y}) - d(\boldsymbol{x}, \boldsymbol{x}) - d(\boldsymbol{y}, \boldsymbol{y}) \\ &= d(\boldsymbol{x}, \boldsymbol{y}) + d(\boldsymbol{y}, \boldsymbol{x}) = 2\, d(\boldsymbol{x}, \boldsymbol{y}). \end{aligned}$$ □

► La *clasificación* de las formas cuadráticas, o lo que es o mismo, la clasificación de las formas bilineales simétricas, procede por convertir una matriz simétrica A en otra matriz diagonal, *congruente* con A. En contraste con la diagonalización por semejanza de una matriz real simétrica, que requiere averiguar los autovalores de la matriz, la diagonalización por congruencia es un algoritmo inductivo más sencillo.

Proposición 4.16. *Sea $d : V \times V \to \mathbb{F}$ una forma bilineal simétrica, de rango $\underline{k}$. Entonces hay una base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-k}\}$ de V para la cual la matriz de $\underline{d}$ es diagonal:*

$$[d]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} b_1 & 0 & \dots & 0 & & & \\ 0 & b_2 & \dots & 0 & & & \\ \vdots & \vdots & \ddots & \vdots & & & \\ 0 & 0 & \dots & b_k & & & \\ & & & & 0 & \dots & 0 \\ & & & & \vdots & \ddots & \vdots \\ & & & & 0 & \dots & 0 \end{bmatrix}.$$

Demostración. Si $A = 0$ y $k = 0$, por ser $d(\boldsymbol{x}, \boldsymbol{y}) \equiv 0$, tómese una base $\{\boldsymbol{z}_1, \dots, \boldsymbol{z}_n\}$ cualquiera. Si $d(\boldsymbol{x}, \boldsymbol{y}) \not\equiv 0$, entonces por (4.6) hay un vector $\boldsymbol{x}_1 \in V$ con $q(\boldsymbol{x}_1) \neq 0$; colóquese $\underline{b_1 := q(\boldsymbol{x}_1)}$.

La construcción de la base $\mathcal{B}$ procede por inducción. Supóngase, entonces, que ya se ha encontrado unos r vectores linealmente independientes $\boldsymbol{x}_1, \dots, \boldsymbol{x}_r \in V$ tales que $\underline{d(\boldsymbol{x}_i, \boldsymbol{x}_i) = b_i}$ con $b_i \neq 0$; y $d(\boldsymbol{x}_i, \boldsymbol{x}_j) = 0$ para $i \neq j$ en $\{1, \dots, r\}$. En tal caso, la restricción de d al subespacio $M_r := \operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_r \rangle$ posee una matriz invertible y por ende esta restricción es una forma bilineal no degenerada. En consecuencia, vale $M_r \cap M_r^{\perp} = \{\boldsymbol{0}\}$.

Si d es no degenerada (es decir, si $k = n$), la Proposición 4.11 permite deducir que $V = M_r \oplus M_r^{\perp}$. Resulta que esta descomposición sigue válida aun cuando $k < n$. Para $\boldsymbol{x} \in V$, colóquese

$$\boldsymbol{y} := \boldsymbol{x} - \frac{d(\boldsymbol{x}, \boldsymbol{x}_1)}{b_1}\, \boldsymbol{x}_1 - \frac{d(\boldsymbol{x}, \boldsymbol{x}_2)}{b_2}\, \boldsymbol{x}_2 - \cdots - \frac{d(\boldsymbol{x}, \boldsymbol{x}_r)}{b_r}\, \boldsymbol{x}_r\,.$$

Entonces, para $j = 1, \dots, r$, vale

$$\begin{aligned} d(\boldsymbol{y}, \boldsymbol{x}_j) = d(\boldsymbol{x}, \boldsymbol{x}_j) - \sum_{i=1}^{r} \frac{d(\boldsymbol{x}, \boldsymbol{x}_i)}{b_i}\, d(\boldsymbol{x}_i, \boldsymbol{x}_j) &= d(\boldsymbol{x}, \boldsymbol{x}_j) - \frac{d(\boldsymbol{x}, \boldsymbol{x}_j)}{b_j}\, d(\boldsymbol{x}_j, \boldsymbol{x}_j) \\ &= d(\boldsymbol{x}, \boldsymbol{x}_j) - d(\boldsymbol{x}, \boldsymbol{x}_j) = 0, \end{aligned}$$

así que $\boldsymbol{y} \in M_r^\perp$. Por tanto, $\boldsymbol{x}$ es una suma de dos partes:

$$\boldsymbol{x} = \sum_{i=1}^{r} \underline{b_i^{-1} d(\boldsymbol{x}, \boldsymbol{x}_i)}\, \boldsymbol{x}_i + \boldsymbol{y} \in M_r \oplus M_r^\perp. \tag{4.7}$$

Como $\boldsymbol{x} \in V$ es arbitrario, se obtiene $V = M_r \oplus M_r^\perp$ (suma directa vectorial).

La matriz de d con respecto a esta descomposición de V tiene evidentemente el aspecto

$$\begin{bmatrix} B_r & 0 \\ 0 & C_{n-r} \end{bmatrix}, \quad \text{donde} \quad B_r = \operatorname{diag}[b_1, \dots, b_r] \quad \text{y} \quad C_{n-r} \in M_{n-r}(\mathbb{F}).$$

Si $r < k$, de modo que $C_{n-r} \neq 0$, la restricción de d al subespacio $M_r^\perp$ no es nula. Entonces existe un vector $\underline{\boldsymbol{x}_{r+1} \in M_r^\perp}$ tal que $\underline{b_{r+1} := q(\boldsymbol{x}_{r+1})} \neq 0$. Ahora $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{r+1}\}$ es la base de un subespacio $M_{r+1} \leqslant V$ (con $M_r \leqslant M_{r+1}$, desde luego); y la restricción de d a M_{r+1} tiene matriz diagonal $B_{r+1} := \operatorname{diag}[b_1, \dots, b_{r+1}]$ con entradas diagonales no ceros.

Este proceso se repite hasta llegar a B_k, en cuyo caso $C_{n-k} = 0$ porque cada matriz de d tiene rango k. Al elegir una base cualquiera $\{z_1, \dots, z_{n-k}\}$ para $M_k^\perp$, se obtiene la base deseada $\mathcal{B} := \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k, z_1, \dots, z_{n-k}\}$ de V. □

Corolario 4.17. *Una forma cuadrática sobre V de rango k puede escribirse como una combinación lineal de cuadrados:*

$$\boxed{q(\boldsymbol{x}) = b_1 f_1(\boldsymbol{x})^2 + b_2 f_2(\boldsymbol{x})^2 + \cdots + b_k f_k(\boldsymbol{x})^2.} \tag{4.8}$$

Aquí $f_1, \dots, f_k \in V^*$ son formas lineales *sobre V; y $b_1, \dots, b_k \in \mathbb{F}$. Además, $f_1, \dots, f_k$ son linealmente independientes en V^*.*

Demostración. En la Proposición 4.16, se ha demostrado que

$$q(\boldsymbol{x}) = b_1 \xi_1^2 + \cdots + b_k \xi_k^2,$$

donde los ξ_j son los coeficientes de $\boldsymbol{x}$ en la base $\mathcal{B}$ de V; esto es, $\boldsymbol{x} = \xi_1 \boldsymbol{x}_1 + \cdots + \xi_k \boldsymbol{x}_k$.

Sea $\mathcal{B}^* = \{f_1, \dots, f_n\}$ la *base dual* de V, así que $\underline{\xi_i = f_i(\boldsymbol{x})}$ para todo $\boldsymbol{x} \in V$. De ahí resulta la fórmula (4.8). Fíjese que $f_i(\boldsymbol{x})$ es el coeficiente subrayado en la expresión (4.7). □

Corolario 4.18. *Cualquier polinomio cuadrático homogéneo en* $\mathbb{F}[t_1, \dots, t_n]$ *puede expresarse como una combinación lineal de* cuadrados de polinomios de primer grado*, sin términos de primer grado ni constantes.*

Demostración. Un polinomio cuadrático homogéneo es una expresión del tipo

$$q(t_1, \dots, t_n) := \sum_{i,j=1}^{n} a_{ij}\, t_i t_j \quad \text{con cada} \quad a_{ij} \in \mathbb{F}.$$

Las sustituciones $t_j \mapsto x_j$ expresan la **evaluación** de q en un vector $\boldsymbol{x} = (x_1, \dots, x_n) \in \mathbb{F}^n$ cualquiera. De este modo se obtiene una forma cuadrática q sobre $\mathbb{F}^n$.

La fórmula (4.8) expresa q en términos de varias formas *lineales* $f_1, \dots, f_k \in (\mathbb{F}^n)^*$. Cada f_i es explícitamente una suma $f_i(\boldsymbol{x}) =: c_{i1}\, x_1 + \cdots + c_{in}\, x_n$.

Defínanse unos polinomios de primer grado (sin términos constantes) en n variables, también denotados f_i, por $\underline{f_i(t_1, \dots, t_n)} := c_{i1}\, \underline{t_1} + \cdots + c_{in}\, \underline{t_n}$. Entonces

$$q(t_1, \dots, t_n) = b_1\, f_1(t_1, \dots, t_n)^2 + \cdots + b_k\, f_k(t_1, \dots, t_n)^2. \qquad \square$$

Resulta que estos polinomios de primer grado f_i pueden elegirse de tal modo que f_i depende solamente de las variables $t_i, \dots, t_n$. Esto sigue de la demostración algorítmica de la Proposición 4.16: la forma lineal f_{r+1} del Corolario 4.17 es el primer elemento de la base dual de la base $\{\boldsymbol{x}_{r+1}, \dots\}$ del subespacio $M_r^{\perp}$, así que no depende de las coordenadas iniciales $x_1, \dots, x_r$ del subespacio M_r.

Ejemplo 4.19. Considérese la siguiente forma cuadrática sobre $\mathbb{Q}$:

$$q(\boldsymbol{x}) := x_1^2 - 4x_1x_2 + 2x_1x_3 - 2x_1x_4 + 3x_2^2 - 6x_2x_3 + 8x_2x_4 + 2x_3x_4 + 2x_4^2.$$

La forma bilineal asociada (con las variables en "orden alfabética") es:

$$\begin{aligned} d(\boldsymbol{x}, \boldsymbol{y}) := x_1y_1 - 2x_1y_2 + x_1y_3 - x_1y_4 - 2x_2y_1 + 3x_2y_2 - 3x_2y_3 + 4x_2y_4 \\ + x_3y_1 - 3x_3y_2 + x_3y_4 - x_4y_1 + 4x_4y_2 + x_4y_3 + 2x_4y_4. \end{aligned}$$

Con $\boldsymbol{e}_1 = (1, 0, 0, 0)^{\mathrm{t}}$, se ve que $\underline{b_1 := q(\boldsymbol{e}_1) = 1}$. Defínase

$$f_1(\boldsymbol{x}) := \underline{b_1^{-1} d(\boldsymbol{x}, \boldsymbol{e}_1)} = x_1 - 2x_2 + x_3 - x_4.$$

Así se define una forma cuadrática residual:

$$\begin{aligned} q(\boldsymbol{x}) - b_1\, f_1(\boldsymbol{x})^2 &= q(\boldsymbol{x}) - (x_1 - 2x_2 + x_3 - x_4)^2 \\ &= -x_2^2 - 2x_2x_3 + 4x_2x_4 - x_3^2 + 4x_3x_4 + x_4^2 =: \underline{q'(\boldsymbol{x})}. \end{aligned}$$

Esta nueva forma cuadrática q' no depende de la coordenada x_1. Hasta ahora se ha identificado $M_1 = \mathrm{lin}\langle \boldsymbol{e}_1 \rangle$ y $M_1^\perp = \mathrm{lin}\langle \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4 \rangle$.

Con $\boldsymbol{e}_2 = (0, 1, 0, 0)^\mathrm{t}$, se ve que $\underline{b_2 := q'(\boldsymbol{e}_2) = -1}$. Defínase

$$f_2(\boldsymbol{x}) := \underline{b_2^{-1} d(\boldsymbol{x}, \boldsymbol{e}_2)} = x_2 + x_3 - 2x_4.$$

El segundo paso del algoritmo es:

$$q(\boldsymbol{x}) - b_1\, f_1(\boldsymbol{x})^2 - b_2\, f_2(\boldsymbol{x})^2 = q'(\boldsymbol{x}) - (x_2 + x_3 - 2x_4)^2 = 5x_4^2 =: \underline{q''(\boldsymbol{x})},$$

con $M_2 = \mathrm{lin}\langle \boldsymbol{e}_1, \boldsymbol{e}_2 \rangle$, $M_2^\perp = \mathrm{lin}\langle \boldsymbol{e}_3, \boldsymbol{e}_4 \rangle$. Ahora se toma $\boldsymbol{x}_3 := \boldsymbol{e}_4 = (0, 0, 0, 1)^\mathrm{t}$, observando que $b_3 = q''(\boldsymbol{e}_4) = 5$. Se llega así al resultado final:

$$q(\boldsymbol{x}) = (x_1 - 2x_2 + x_3 - x_4)^2 - (x_2 + x_3 - 2x_4)^2 + 5x_4^2. \qquad \Diamond$$

Es posible identificar un juego de formas lineales $f_1, \dots, f_k$ que definen la descomposición (4.8) de $q(\boldsymbol{x})$, mediante un proceso algorítmico conocido como la **reducción de Lagrange**.

Proposición 4.20. *La forma cuadrática q sobre $\mathbb{F}^n$ dada por $q(\boldsymbol{x}) = \boldsymbol{x}^\mathrm{t} A \boldsymbol{x} \equiv \sum_{i,j=1}^n a_{ij} x_i x_j$ con una matriz simétrica $A = A^\mathrm{t} \in M_n(\mathbb{F})$ puede expresarse como una combinación lineal* (4.8) *de cuadradas de formas lineales, como sigue.*

(a) *Si $a_{kk} \neq 0$ para algún k, se debe reordenar las variables para que $a_{11} \neq 0$. Entonces*

$$q(\boldsymbol{x}) = \frac{1}{a_{11}} \biggl(\sum_{j=1}^n a_{1j} x_j \biggr)^2 + q_1(\boldsymbol{x}), \tag{4.9a}$$

donde $q_1(\boldsymbol{x})$ depende solamente de $(x_2, \dots, x_n)$.

(b) *Si todo $a_{kk} = 0$, se debe reordenar variables para que $a_{12} \neq 0$. El tal caso, se obtiene*

$$q(\boldsymbol{x}) = \frac{1}{2a_{12}} \biggl(\sum_{j=1}^n (a_{1j} + a_{2j}) x_j \biggr)^2 - \frac{1}{2a_{12}} \biggl(\sum_{j=1}^n (a_{1j} - a_{2j}) x_j \biggr)^2 + q_2(\boldsymbol{x}), \tag{4.9b}$$

donde $q_2(\boldsymbol{x})$ depende solamente de $(x_3, \dots, x_n)$.

(c) *Al repetir el paso apropiado (a) o (b) con la nueva forma cuadrática $q_1(\boldsymbol{x})$ o $q_2(\boldsymbol{x})$, y así sucesivamente, hasta que la forma cuadrática residual es nula, se obtiene así la expresión* (4.8) *de $q(\boldsymbol{x})$.*

Demostración. Un cálculo directo verifica que $q_1(\boldsymbol{x})$ no depende de x_1; y otro cálculo directo comprueba que $q_2(\boldsymbol{x})$ no depende de x_1 ni x_2.

En el paso (a), si es factible, se construye $b_1 := a_{11}^{-1}$ y una forma lineal $f_1(\boldsymbol{x})$. De lo contrario, si q no la forma cuadrática nula, se puede ejecutar el paso (b), que produce $b_1 := (2a_{12})^{-1}$ y $b_2 := -(2a_{12})^{-1}$, junto con las formas lineales $f_1(\boldsymbol{x})$ y $f_2(\boldsymbol{x})$.

Los otros términos de (4.8) se obtienen al repetir estos pasos hasta agotar las x_i. □

Escolio 4.21. *Las formas lineales $f_1, \dots, f_k$ obtenidos por los pasos* (4.9a) *y* (4.9b) *son linealmente independientes.* ⊟

► La forma diagonal de la Proposición 4.16 no es única, porque hay cierta flexibilidad en la determinación de los coeficientes b_j. Por ejemplo, si se puede factorizar un coeficiente dado b_j como $\underline{b_j = a_j^2 c_j}$ con $a_j \neq 0$, entonces las sustituciones $b_j \mapsto c_j$ y $\boldsymbol{x}_j \mapsto a_j^{-1}\boldsymbol{x}_j$ dan lugar a

$$d(a_i^{-1}\boldsymbol{x}_i, a_j^{-1}\boldsymbol{x}_j) = a_i^{-1}a_j^{-1}\, d(\boldsymbol{x}_i, \boldsymbol{x}_j) = a_i^{-1}a_j^{-1} b_j\, [\![i = j]\!] = c_j\, [\![i = j]\!].$$

Por tanto, para la matriz de d en la nueva base $\mathcal{C} = \{a_1^{-1}\boldsymbol{x}_1, \dots, a_k^{-1}\boldsymbol{x}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-k}\}$ se obtiene $[d]_{\mathcal{C}}^{\mathcal{C}} = \operatorname{diag}[c_1, \dots, c_k, 0, \dots, 0]$. Es posible y es deseable, entonces, "normalizar" los coeficientes diagonales no ceros al dividirlos por cuadrados convenientes.

Se debe distinguir dos casos importantes. Cuando $\mathbb{F} = \mathbb{C}$, cualquier número complejo no cero posee una raíz cuadrada compleja.[4] En contraste, cuando $\mathbb{F} = \mathbb{R}$, solamente los números positivos (excluyendo 0) poseen raíces cuadradas reales. Estas consideraciones bastan para demostrar el siguiente teorema de Sylvester,[5] conocido como la **Ley de inercia**.

Teorema 4.22 (Ley de Inercia)**.** *Sea V un espacio vectorial de dimensión n sobre $\mathbb{F} = \mathbb{R}$ o $\mathbb{C}$ y sea $d\colon V \times V \to \mathbb{F}$ una forma bilineal simétrica de rango k.*

(a) *Si $\mathbb{F} = \mathbb{C}$, entonces hay una base $\mathcal{C}$ de V para la cual la matriz de $\underline{d}$ es*

$$[d]_{\mathcal{C}}^{\mathcal{C}} = \operatorname{diag}[1, \dots, 1, 0, \dots, 0] = 1_k \oplus 0_{n-k}, \tag{4.10a}$$

con $\underline{k}$ entradas diagonales iguales a 1.

(b) *Si $\mathbb{F} = \mathbb{R}$, entonces hay una base $\mathcal{C}$ de V para la cual la matriz de $\underline{d}$ es*

$$[d]_{\mathcal{C}}^{\mathcal{C}} = \operatorname{diag}[1, \dots, 1, -1, \dots, -1, 0, \dots, 0] = 1_p \oplus -1_q \oplus 0_{n-p-q}, \tag{4.10b}$$

con $\underline{p}$ entradas diagonales iguales a 1 *y $\underline{q}$ entradas diagonales iguales a* (-1)*, donde $p + q = k$.*

[4]De hecho, tiene *dos* raíces cuadradas, pues $(-\alpha)^2 = \alpha^2$. Por ejemplo, $(-1+i)^2 = (1-i)^2 = -2i$.

[5]*James Joseph Sylvester* (1814–1897), junto con su compatriota *Arthur Cayley* (1821–1895), se consideran padres fundadores del álgebra abstracta. Sylvester inventó buena parte de su terminología: la palabra *matriz*, el *discriminante* de un polinomio, la *función tociente* de Euler, y la *ley de inercia* para formas cuadráticas, fueron vocablos introducidos por él.

Demostración. Sea $\mathcal{B} := \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-k}\}$ la base de V obtenida en la demostración de la Proposición 4.16, para la cual la matriz de d es $[d]_{\mathcal{B}}^{\mathcal{B}} = \operatorname{diag}[b_1, \dots, b_k, 0, \dots, 0]$, donde cada $b_j \neq 0$.

Ad (a): Si $\mathbb{F} = \mathbb{C}$, sea a_j una (de dos) raíz cuadrada de b_j, para $j = 1, \dots, k$. Defínase $\boldsymbol{y}_j := a_j^{-1}\boldsymbol{x}_j$; la matriz de d para la base $\mathcal{C} := \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-k}\}$ es $[d]_{\mathcal{C}}^{\mathcal{C}} = 1_k \oplus 0_{n-k}$.

Ad (b): Si $\mathbb{F} = \mathbb{R}$, en la lista $(b_1, \dots, b_k)$ hay unas p entradas positivas y unas q entradas negativas, con $p + q = k$, desde luego. Por una permutación de los vectores $\boldsymbol{x}_j$, si fuera necesario, se puede suponer que $b_j > 0$ para $j = 1, \dots, p$ y que $b_j < 0$ para $j = p+1, \dots, k$.

Ahora colóquese $a_j := \sqrt{b_j}$ para $j = 1, \dots, p$ y $a_j := \sqrt{-b_j}$ para $j = p+1, \dots, k$; en seguida, defínase $\boldsymbol{y}_j := a_j^{-1}\boldsymbol{x}_j$; entonces la matriz de d respecto de la base $\mathcal{C} := \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-k}\}$ es $[d]_{\mathcal{C}}^{\mathcal{C}} = 1_p \oplus -1_q \oplus 0_{n-p-q}$. □

Definición 4.23. Sea $A = A^{\mathrm{t}} \in M_n(\mathbb{R})$ *una matriz simétrica real*. La forma bilineal simétrica $d_A(\boldsymbol{x}, \boldsymbol{y}) := \boldsymbol{x}^{\mathrm{t}} A \boldsymbol{y}$ sobre $\mathbb{R}^n$ determina una matriz diagonal (4.10b) con p entradas diagonales $+1$ y $q = r(A) - p$ entradas diagonales -1. La diferencia[6] $\underline{s(A) := p - q} \in \mathbb{Z}$ se llama la **signatura** de la matriz A. ◊

► La importancia de la signatura, junto con el rango, en el caso real, es que estos dos números caracterizan las clases de equivalencia bajo la relación de congruencia. Eso es lo que afirma el resultado siguiente.

Proposición 4.24. *Dos matrices simétricas reales $A, B \in M_n(\mathbb{R})$ son congruentes si y solo si tienen el mismo rango y la misma signatura: $r(A) = r(B)$ y $s(A) = s(B)$.*

Demostración. Por el Teorema 4.22, fórmula (4.10b), hay enteros $p, p', q, q' \in \mathbb{N}$ tales que $A \simeq 1_p \oplus -1_q \oplus 0_{n-p-q}$ mientras $B \simeq 1_{p'} \oplus -1_{q'} \oplus 0_{n-p'-q'}$.

Ad ($\Leftarrow$): Si $r(A) = r(B) = k$ y si $s(A) = s(B) = s$, entonces con $p + q = p' + q' = k$ y además $p - q = p' - q' = s$, de donde $p = \frac{1}{2}(k+s) = p'$ y $q = \frac{1}{2}(k-s) = q'$. (Fíjese, de paso, que los enteros k y s tienen la misma paridad.) Se deduce que $A \simeq B$.

Ad ($\Rightarrow$): Si A y B son congruentes, ya se sabe que A y B tienen el mismo rango. Denótese $k := r(A) = r(B)$ en ese caso. Basta entonces mostrar que

$$(1_p \oplus -1_q \oplus 0_{n-k}) \simeq (1_{p'} \oplus -1_{q'} \oplus 0_{n-k}).$$

Si R es una matriz invertible $k \times k$ tal que $(1_{p'} \oplus -1_{q'}) = R^{\mathrm{t}}(1_p \oplus -1_q)R$, entonces la matriz invertible $P := R \oplus 1_{n-k}$ satisface $(1_{p'} \oplus -1_{q'} \oplus 0_{n-k}) = P^{\mathrm{t}}(1_p \oplus -1_q \oplus 0_{n-k})P$. Por eso, no se pierde generalidad en suponer que $k = n$ (o sea, que A y B son invertibles).

[6] Algunos autores llaman *signatura* al par ordenado (p, q); a la diferencia $p - q$ la llaman el **índice** de una forma bilineal simétrica con matriz A.

Basta entonces verificar que las matrices diagonales $(1_p \oplus 1_q)$ y $(1_{p'} \oplus -1_{q'})$, donde $p + q = p' + q' = n$, son congruentes si y solo si $p = p'$.

Considérese la forma bilineal simétrica sobre $\mathbb{R}^n$ dada por

$$d(\boldsymbol{x}, \boldsymbol{y}) := x_1 y_1 + \cdots + x_p y_p - x_{p+1} y_{p+1} - \cdots - x_n y_n. \tag{4.11}$$

La matriz de d en la base estándar $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de $\mathbb{R}^n$ es entonces $[d]^{\mathcal{E}}_{\mathcal{E}} = (1_p \oplus -1_q)$. Sea $\mathcal{U} := \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ otra base de $\mathbb{R}^n$ tal que $[d]^{\mathcal{U}}_{\mathcal{U}} = 1_{p'} \oplus -1_{q'}$. Explícitamente, con $\boldsymbol{x} = x'_1 \boldsymbol{u}_1 + \cdots + x'_n \boldsymbol{u}_n$, $\boldsymbol{y} = y'_1 \boldsymbol{u}_1 + \cdots + y'_n \boldsymbol{u}_n$, supóngase que se verifica

$$d(\boldsymbol{x}, \boldsymbol{y}) = x'_1 y'_1 + \cdots + x'_{p'} y'_{p'} - x'_{p'+1} y'_{p'+1} - \cdots - x'_n y'_n.$$

Resulta que los subespacios $M := \operatorname{lin}\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_p \rangle$ y $N' := \operatorname{lin}\langle \boldsymbol{u}_{p'+1}, \dots, \boldsymbol{u}_n \rangle$ son suplementarios. En efecto, es obvio que

$$d(\boldsymbol{y}, \boldsymbol{y}) > 0 \text{ si } \boldsymbol{y} \in M,\ \boldsymbol{y} \neq \boldsymbol{0}, \quad \text{mientras} \quad d(\boldsymbol{z}, \boldsymbol{z}) < 0 \text{ si } \boldsymbol{z} \in N',\ \boldsymbol{z} \neq \boldsymbol{0}.$$

De ahí se ve que $M \cap N' = \{0\}$ y luego $M + N' = M \oplus N'$ (es decir, la suma vectorial de M y N es una suma directa). Al contar dimensiones, se ve que

$$p + q' = \dim(M \oplus N') \leqslant n = p + q, \quad \text{así que} \quad q' \leqslant q.$$

Del mismo modo, los subespacios $M' := \operatorname{lin}\langle \boldsymbol{u}_1, \dots, \boldsymbol{u}_{p'} \rangle$ y $N := \operatorname{lin}\langle \boldsymbol{e}_{p+1}, \dots, \boldsymbol{e}_n \rangle$ cumplen $M' \cap N = \{0\}$. Esto implica que

$$p' + q = \dim(M' \oplus N) \leqslant n = p + q \quad \text{y por ende} \quad p' \leqslant p.$$

Pero $p' + q' = n = p + q$, de donde $p' = p$ y $q' = q$ necesariamente. □

Definición 4.25. Si $d \colon V \times V \to \mathbb{R}$ es una forma bilineal simétrica sobre un espacio vectorial real, la **signatura** de d es la signatura $s := p - q \in \mathbb{Z}$ de cualquiera de sus matrices $[d]^{\mathcal{B}}_{\mathcal{B}}$. También se dice que s es la signatura de la forma cuadrática real q asociada con d. ◊

Definición 4.26. Una forma bilineal simétrica real d sobre V es **positiva** si $d(\boldsymbol{x}, \boldsymbol{x}) \geqslant 0$ para todo $\boldsymbol{x} \in V$. Una forma cuadrática real q es positiva si $q(\boldsymbol{x}) \geqslant 0$ para todo $\boldsymbol{x} \in V$.

Si estas desigualdades son estrictas para todo $\boldsymbol{x} \neq \boldsymbol{0}$, se dice que d [respectivamente, q] es **definida positiva**. ◊

Si $\mathcal{C} = \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_p, \boldsymbol{y}_{p+1}, \dots, \boldsymbol{y}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-k}\}$ es una base de V para la cual d tiene la matriz (4.10b), entonces $q(\boldsymbol{y}_{p+1}) = d(\boldsymbol{y}_{p+1}, \boldsymbol{y}_{p+1}) < 0$ si $q \neq 0$ y en consecuencia d no es positiva. De ahí se ve que d es positiva *si y solo si* $q = 0$, *si y solo si* su rango y signatura coinciden ($s = k$). Además, q y d son definidas positivas si d es positiva y no degenerada (el caso $s = k = n$).

► Obsérvese que *un producto escalar real* no es más que una forma bilineal simétrica real que es definida positiva. La teoría de espacios $\mathbb{R}$-vectoriales euclidianos admite una generalización que consiste en reemplazar el producto escalar por otra forma bilineal simétrica *indefinida*. Por ejemplo, se podría reemplazar el producto punto en $\mathbb{R}^n$ por la forma (4.11).

4.3. Formas hermíticas

Un producto escalar real sobre un espacio $\mathbb{R}$-vectorial es en particular una forma bilineal (que es también definida positiva). En contraste con eso, un producto escalar *complejo* sobre un espacio $\mathbb{C}$-vectorial *no es bilineal*: es lineal en una variable pero solo semilineal (antilineal) en la otra. Para poder incorporar los productos escalares complejos en la temática de este capítulo, es necesario reconsiderar el punto de partida. Lo que se requiere es sustituir la bilinealidad de una función de dos vectores por "$\frac{3}{2}$-linealidad". Con la ayuda del prefijo latino *sesqui-* que significa tres medios, es posible abordar el estudio de formas sesquilineales sobre un espacio vectorial complejo, donde otra vez resulta posible clasificar algunas de tales formas por su rango y signatura.

En esta sección, V denotará un espacio vectorial finitodimensional sobre $\mathbb{C}$.

Definición 4.27. Una aplicación $h\colon V \times V \to \mathbb{C}$ es una **forma sesquilineal** si la aplicación parcial $\boldsymbol{w} \mapsto h(\boldsymbol{z}, \boldsymbol{w})$ es una forma *lineal* sobre V y la aplicación parcial $\boldsymbol{z} \mapsto h(\boldsymbol{z}, \boldsymbol{w})$ es *semilineal*; en otras palabras,

$$\begin{aligned} h(\boldsymbol{z} + \boldsymbol{z}', \boldsymbol{w}) &= h(\boldsymbol{z}, \boldsymbol{w}) + h(\boldsymbol{z}', \boldsymbol{w}), \qquad & h(\boldsymbol{z}, \alpha\boldsymbol{w}) &= \alpha\, h(\boldsymbol{z}, \boldsymbol{w}), \\ h(\boldsymbol{z}, \boldsymbol{w} + \boldsymbol{w}') &= h(\boldsymbol{z}, \boldsymbol{w}) + h(\boldsymbol{z}, \boldsymbol{w}'), \qquad & h(\alpha\boldsymbol{z}, \boldsymbol{w}) &= \bar{\alpha}\, h(\boldsymbol{z}, \boldsymbol{w}), \end{aligned}$$

para todo $\boldsymbol{z}, \boldsymbol{z}', \boldsymbol{w}, \boldsymbol{w}' \in V$, $\alpha \in \mathbb{C}$.

Una forma sesquilineal h se llama **hermítica** si

$$h(\boldsymbol{z}, \boldsymbol{w}) = \overline{h(\boldsymbol{w}, \boldsymbol{z})} \quad \text{para todo} \quad \boldsymbol{z}, \boldsymbol{w} \in V. \qquad \Diamond$$

Esta definición es obviamente análogo a una forma bilineal simétrica.

Con respecto a una base $\mathcal{B} = \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_n\}$ de V, la fórmula $a_{ij} := h(\boldsymbol{z}_i, \boldsymbol{z}_j)$ determina la matriz $A = [h]^{\mathcal{B}}_{\mathcal{B}}$ de una forma sesquilineal. Es evidente que una forma sesquilineal es hermítica si y solo si su matriz es hermítica: $\underline{A^* = A}$.

Dos matrices de una forma sesquilineal referido a bases diferentes están ligadas por una relación parecida, pero no idéntica, a la congruencia de matrices (4.3). Una leve modificación a la prueba de la Proposición 4.4 demuestra lo siguiente.

Escolio 4.28. *Sean $\mathcal{B} = \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_n\}$ y $\mathcal{B}' = \{\boldsymbol{z}'_1, \dots, \boldsymbol{z}'_n\}$ dos bases de V; y sea $P \in M_n(\mathbb{C})$ la matriz de cambio de base, dada por $\boldsymbol{z}'_s =: \sum_{j=1}^n p_{js}\boldsymbol{z}_j$. Si $h\colon V \times V \to \mathbb{C}$ es una forma sesquilineal, las matrices respectivas $A = [f]^{\mathcal{B}}_{\mathcal{B}}$ y $B = [h]^{\mathcal{B}'}_{\mathcal{B}'}$ de h cumplen la relación:*

$$\boxed{B = P^*AP.} \tag{4.12}$$

Definición 4.29. Dos matrices cuadradas $A, B \in M_n(\mathbb{C})$ se llaman **matrices $*$-congruentes** (o *matrices conjuntivas*) si hay una matriz invertible $P \in M_n(\mathbb{C})$ tal que $\underline{B = P^*AP}$. $\Diamond$

La $*$-congruencia es una relación de equivalencia: por ejemplo, si $B = P^*AP$, entonces $A = (P^{-1})^*BP^{-1}$. Dos matrices $*$-congruentes tienen el mismo rango: este es el caso de $A \mapsto QAP$ con $Q = P^*$. Por lo tanto, el rango de la matriz $[h]^{\mathcal{B}}_{\mathcal{B}}$ de una forma sesquilineal no depende de la base $\mathcal{B}$ elegida, y se llama el **rango de la forma sesquilineal** h.

Dada una forma hermítica $h\colon V \times V \to \mathbb{C}$, el concepto de **subespacio ortogonal** $M^\perp$ (con respecto a h) de un subespacio $M \leqslant V$ se define por el mismo protocolo (4.5) usado para formas bilineales simétricas:

$$\begin{aligned} M^\perp &= \{\, \boldsymbol{z} \in V : h(\boldsymbol{z}, \boldsymbol{w}) = 0 \text{ para todo } \boldsymbol{w} \in M \,\} \\ &= \{\, \boldsymbol{w} \in V : h(\boldsymbol{z}, \boldsymbol{w}) = 0 \text{ para todo } \boldsymbol{z} \in M \,\}. \end{aligned}$$

Ejemplo 4.30. Si V es un espacio de Hilbert, su producto escalar

$$h(\boldsymbol{z}, \boldsymbol{w}) := \langle \boldsymbol{z}, \boldsymbol{w} \rangle$$

es una forma sesquilineal y hermítica, que además es *definida positiva*, porque $h(\boldsymbol{z}, \boldsymbol{z}) > 0$ para $\boldsymbol{z} \neq \mathbf{0}$ en V. Esto dice que una forma hermítica es una generalización de la noción de producto escalar, donde se puede omitir el requisito de positividad. ◊

Ejemplo 4.31. Si $p, q \in \mathbb{N}$ con $p + q = k \leqslant n$, la forma sesquilineal siguiente sobre $\mathbb{C}^n$ es hermítica:

$$\underline{h(\boldsymbol{z}, \boldsymbol{w})} := \bar{z}_1 w_1 + \cdots + \bar{z}_p w_p - \bar{z}_{p+1} w_{p+1} - \cdots - \bar{z}_{p+q} w_{p+q}.$$

Con respecto a la base estándar $\mathcal{E}$, es evidente que

$$[h]^{\mathcal{E}}_{\mathcal{E}} = \operatorname{diag}[1, \dots, 1, -1, \dots, -1, 0, \dots, 0] = \begin{bmatrix} 1_p & 0 & 0 \\ 0 & -1_q & 0 \\ 0 & 0 & 0_{n-p-q} \end{bmatrix}. \tag{4.13}$$

Esta matriz de h muestra que la forma hermítica h tiene *rango* $k = p + q$ y sugiere que h tiene *signatura* $s := p - q$. Efectivamente, hay una *ley de inercia* para formas hermíticas, directamente análoga a la expresión (4.10b) para formas bilineales simétricas reales. ◊

Proposición 4.32. *Si* $\dim V = n$ *y si* $h\colon V \times V \to \mathbb{C}$ *es una forma hermítica de rango* k*, entonces hay una base* $\mathcal{C}$ *de* V *tal que* $[h]^{\mathcal{C}}_{\mathcal{C}} = 1_p \oplus -1_q \oplus 0_{n-k}$*, donde* $p + q = k$*.*

Demostración. Tómese una base cualquiera $\mathcal{B} = \{z_1, \dots, z_n\}$ del espacio $\mathbb{C}$-vectorial V y sea $A = [h]^{\mathcal{B}}_{\mathcal{B}}$ la matriz de h con respecto a esta base. Como h es hermítica, se sigue que $A^* = A$ en $M_n(\mathbb{C})$. Por el Lema 3.50 – justo antes del teorema espectral – hay una matriz *unitaria* $U \in M_n(\mathbb{C})$ tal que $U^{-1}AU = U^*AU$ es diagonal; y sus entradas diagonales son los autovalores (reales) de A.

Si $\mathcal{B}' = \{z'_1, \dots, z'_n\}$ es la base dada por $z'_s := \sum_{j=1}^n u_{js} z_j$, la matriz de h en esta base es diagonal: $[h]^{\mathcal{B}'}_{\mathcal{B}'} = U^*AU = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$ donde $\lambda_1, \dots, \lambda_n$ son los n autovalores de A.

Es posible elegir U de tal manera que los autovectores z'_s de A estén en un orden predeterminado. ⟦ Una permutación de los vectores en una base dada se efectúa por una matriz P cuyas columnas son las de la base estándar en otro orden; este P es real y ortogonal, y por ende unitaria; es cuestión de cambiar U por UP para obtener $\mathcal{B}'$ a partir de $\mathcal{B}$. ⟧ Entonces se puede asumir, sin perder generalidad, que los primeros autovalores son *positivos*: $\lambda_i > 0$ para $i = 1, \dots, p$; seguidos por unos autovalores de A *negativos*: $\lambda_j < 0$ para $j = p+1, \dots, p+q$; y por último los autovalores nulos: $\lambda_r = 0$ para $r = p+q+1, \dots, n$. (Se debe notar que el rango de la matriz diagonal U^*AU es $p+q = r(A) = k$.)

Dicho de otro modo: hay números positivos $a_1, \dots, a_{p+q}$ tales que

$$\lambda_1 = +a_1^2, \dots\ \lambda_p = +a_p^2; \quad \lambda_{p+1} = -a_{p+1}^2, \dots\ \lambda_{p+q} = -a_{p+q}^2; \quad \lambda_{p+q+1} = \cdots = \lambda_n = 0.$$

Sea $\mathcal{C} := \{a_1^{-1}z'_1, \dots, a_p^{-1}z'_p; a_{p+1}^{-1}z'_{p+1}, \dots, a_{p+q}^{-1}z'_{p+q}; z'_{p+q+1}, \dots, z'_n\}$. La matriz de h en esta base entonces es $[h]_{\mathcal{C}}^{\mathcal{C}} = 1_p \oplus -1_q \oplus 0_{n-k}$, como en (4.13). □

Se puede comprobar, al repetir la demostración de la Proposición 4.24, que el rango $\underline{r(A) = p+q}$ y la **signatura** $\underline{s(A) := p-q}$, obtenidos de la matriz diagonal (4.13) no dependen de la base $\mathbb{C}$ elegida.

Escolio 4.33. *Dos matrices hermíticas $A, B \in M_n(\mathbb{C})$ son $*$-congruentes si y solo si tienen el mismo rango y la misma signatura: $r(A) = r(B)$ y $s(A) = s(B)$.* ⊟

En el transcurso de la demostración de la Proposición 4.32, se obtuvo una nueva caracterización de la signatura de una forma o matriz hermítica que es importante declarar.

Corolario 4.34. *Sea $A = A^* \in M_n(\mathbb{C})$ una matriz hermítica, $*$-congruente con la matriz diagonal $(1_p \oplus -1_q \oplus 0_{n-p-q})$. Entonces $\underline{p}$ es* el número total de autovalores positivos *de A; $\underline{q} = r(A) - p$ es* el número total de autovalores negativos *de A; además, $n(A) = n - p - q$ es el número total de autovalores ceros de A. La signatura de A es $s(A) := p - q \in \mathbb{Z}$.* ⊟

► Es evidente que las leyes de inercia para formas simétricas reales, la Proposición 4.32(b); y para formas hermíticas, la Proposición 4.32, conducen a resultados similares; pero sus demostraciones difieren porque en el segundo caso fue posible aprovechar el teorema espectral.

Cabe preguntar si en el caso de formas simétricas reales, se puede obtener la signatura por el mismo conteo de signos de los autovalores de una matriz $A = A^t$. Resulta *que sí*: pero requiere una prueba diferente. Sucede que se puede diagonalizar una matriz simétrica real A (cuyos autovalores son reales) por un *cambio de base ortogonal* de $\mathbb{R}^n$; en otras palabras, hay una matriz *ortogonal* Q tal que $Q^{-1}AQ = Q^tAQ$ es diagonal. Para ese resultado, consúltese la sección IX.13 del libro de Gantmacher. En seguida, el resto de la demostración de la Proposición 4.32 es aplicable al caso simétrico real.

4.4. Formas bilineales alternadas

Las formas bilineales *alternadas* tienen una estructura más sencilla que las formas bilineales simétricas, pero no son menos importantes. En esta sección $\mathbb{F}$ denota un cuerpo cualquiera en el cual $1 + 1 \neq 0$. El ejemplo primordial de una forma bilineal alternada es la aplicación $s_0 \colon \mathbb{F}^2 \times \mathbb{F}^2 \to \mathbb{F}$ dada por

$$s_0(\boldsymbol{x}, \boldsymbol{y}) := x_1 y_2 - x_2 y_1.$$

Su matriz con respecto a la base estándar $\{\boldsymbol{e}_1, \boldsymbol{e}_2\}$ de $\mathbb{F}^2$ es la siguiente matriz antisimétrica:

$$J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{4.14}$$

⟦ Esta es la misma matriz (2.5) del Ejemplo 2.13. ⟧ En adelante se verá que cualquier matriz antisimétrica es congruente con una suma directa de varias copias de esta J_2.

Definición 4.35. Una forma bilineal alternada $s \colon V \times V \to \mathbb{F}$ que es *no degenerada* se llama una **forma bilineal simpléctica** sobre V. ◊

Definición 4.36. Si $s \colon V \times V \to \mathbb{F}$ es una forma bilineal alternada, y si M es un subespacio de V, su **complemento simpléctico** con respecto a $\underline{s}$ se define por analogía con (4.5):

$$\begin{aligned} M^\perp &= \{\, \boldsymbol{x} \in V : s(\boldsymbol{x}, \boldsymbol{y}) = 0 \text{ para todo } \boldsymbol{y} \in M \,\} \\ &= \{\, \boldsymbol{y} \in V : s(\boldsymbol{x}, \boldsymbol{y}) = 0 \text{ para todo } \boldsymbol{x} \in M \,\}. \end{aligned} \tag{4.15}$$

Un subespacio $M \leqslant V$ tal que $\underline{M \subseteq M^\perp}$ se llama un **subespacio isotrópico** de V. ◊

A primera vista, la relación $M \subseteq M^\perp$ puede parecer extraño. Es una primera indicación de que los complementos simplécticos presentan un fuerte contraste con los complementos ortogonales determinados por formas simétricas o hermíticas (o por productos escalares). Con respecto a una forma alternada s, cualquier vector $\boldsymbol{x} \in V$ cumple $s(\boldsymbol{x}, \boldsymbol{x}) = 0$ por antisimetría. En consecuencia, si $N := \mathbb{F}\boldsymbol{x} = \{\, c\boldsymbol{x} : c \in \mathbb{F} \,\}$ es el subespacio unidimensional generado por $\boldsymbol{x}$, entonces $N \subseteq N^\perp$.

Proposición 4.37. *Sea $s \colon V \times V \to \mathbb{F}$ una forma bilineal alternada, de rango $\underline{k}$. Entonces k es un número entero par, $\underline{k = 2m}$; y hay una base $\mathcal{B} = \{\boldsymbol{x}_1, \boldsymbol{y}_1, \dots, \boldsymbol{x}_m, \boldsymbol{y}_m, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-2m}\}$ en la cual la matriz de $\underline{s}$ es:*

$$[s]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} J_{2m} & 0 \\ 0 & 0_{n-2m} \end{bmatrix}, \quad \text{con} \quad \underline{J_{2m}} := \begin{bmatrix} J_2 & 0 & \dots & 0 \\ 0 & J_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & J_2 \end{bmatrix} = \underbrace{J_2 \oplus \cdots \oplus J_2}_{m \text{ veces}}. \tag{4.16}$$

Demostración. Si $s \equiv 0$ es la forma bilineal idénticamente nula, se puede usar una base arbitraria $\mathcal{B} = \{\boldsymbol{z}_1, \dots, \boldsymbol{z}_n\}$ de V, porque $[s]_{\mathcal{B}}^{\mathcal{B}} = 0$.

En cambio, si s no es nula, hay dos vectores $\boldsymbol{x}_1, \boldsymbol{y}'_1 \in V$ con $s(\boldsymbol{x}_1, \boldsymbol{y}'_1) =: c_1 \neq 0$. Fíjese que $\boldsymbol{x}_1, \boldsymbol{y}'_1$ son linealmente independientes: si fuera $\boldsymbol{y}'_1 = a\,\boldsymbol{x}_1$ entonces sería $s(\boldsymbol{x}_1, \boldsymbol{y}'_1) = a\,s(\boldsymbol{x}_1, \boldsymbol{x}_1) = 0$, y si fuera $\boldsymbol{x}_1 = b\,\boldsymbol{y}'_1$ entonces sería $s(\boldsymbol{x}_1, \boldsymbol{y}'_1) = b\,s(\boldsymbol{y}'_1, \boldsymbol{x}_1) = 0$ por antisimetría.

Defínase $\boldsymbol{y}_1 := c_1^{-1}\boldsymbol{y}'_1$, de modo que $s(\boldsymbol{x}_1, \boldsymbol{y}_1) = 1$ y $s(\boldsymbol{y}_1, \boldsymbol{x}_1) = -1$. En el subespacio 2-dimensional $\operatorname{lin}\langle \boldsymbol{x}_1, \boldsymbol{y}_1\rangle \simeq \mathbb{F}^2$, con la base $\{\boldsymbol{x}_1, \boldsymbol{y}_1\}$, la matriz de s es la J_2 de (4.14).

Nótese que s tiene la matriz (4.16) con respecto a la base $\mathcal{B}$ del enunciado si y solo si

$$\boxed{s(\boldsymbol{x}_j, \boldsymbol{y}_j) = 1, \quad s(\boldsymbol{y}_j, \boldsymbol{x}_j) = -1,} \tag{4.17}$$

y además $s(\boldsymbol{z}, \boldsymbol{w}) = 0$ para cualquier otro par de vectores $\boldsymbol{z}, \boldsymbol{w}$ de la base $\mathcal{B}$. Lo que falta, entonces, es identificar un juego de n vectores que cumple esos requisitos.

Hasta ahora solo se dispone de $\{\boldsymbol{x}_1, \boldsymbol{y}_1\}$. Para obtener más vectores $\boldsymbol{x}_j, \boldsymbol{y}_j$, es necesario argumentar por inducción. Supóngase, entonces, que ya se ha elegido unos $2r$ vectores linealmente independientes $\{\boldsymbol{x}_1, \boldsymbol{y}_1, \dots, \boldsymbol{x}_r, \boldsymbol{y}_r\}$ en V tales que

$$\begin{aligned} s(\boldsymbol{x}_j, \boldsymbol{y}_j) = -s(\boldsymbol{y}_j, \boldsymbol{x}_j) = 1 \quad & \text{para} \quad j = 1, 2, \dots, r; \\ s(\boldsymbol{x}_i, \boldsymbol{y}_j) = s(\boldsymbol{y}_j, \boldsymbol{x}_i) = 0 \quad & \text{para} \quad i, j \in \{1, 2, \dots, r\} \quad \text{con} \quad i \neq j; \\ s(\boldsymbol{x}_i, \boldsymbol{x}_j) = s(\boldsymbol{y}_i, \boldsymbol{y}_j) = 0, \quad & \text{para todo} \quad i, j \in \{1, 2, \dots, r\}. \end{aligned}$$

Entonces la restricción de s al subespacio $M_{2r} := \operatorname{lin}\langle \boldsymbol{x}_1, \boldsymbol{y}_1, \dots, \boldsymbol{x}_r, \boldsymbol{y}_r\rangle$ es no degenerada. El cálculo siguiente muestra que $M_{2r} \cap M_{2r}^{\perp} = \{\boldsymbol{0}\}$:

$$\begin{aligned} s(a_1\boldsymbol{x}_1 + \dots + a_r\boldsymbol{x}_r + b_1\boldsymbol{y}_1 + \dots + b_r\boldsymbol{y}_r, \boldsymbol{y}_j) = 0 &\implies a_j = 0, \\ s(a_1\boldsymbol{x}_1 + \dots + a_r\boldsymbol{x}_r + b_1\boldsymbol{y}_1 + \dots + b_r\boldsymbol{y}_r, \boldsymbol{x}_j) = 0 &\implies b_j = 0. \end{aligned}$$

Si s es no degenerada (o sea, si $k = n$), entonces $\underline{V = M_{2r} \oplus M_{2r}^{\perp}}$ por la Proposición 4.11. (La demostración de esa Proposición sigue válida sin cambio alguno para formas bilineales *alternadas* en vez de simétricas). Ahora bien, resulta que esta relación es válida aun para $k < n$. Para verlo, tómese $\boldsymbol{x} \in V$ y colóquese

$$\boldsymbol{z} := \boldsymbol{x} - \sum_{i=1}^{r} s(\boldsymbol{x}, \boldsymbol{y}_i)\,\boldsymbol{x}_i + \sum_{i=1}^{r} s(\boldsymbol{x}, \boldsymbol{x}_i)\,\boldsymbol{y}_i.$$

Entonces, para $j = 1, \dots, r$, las relaciones (4.17) implican que

$$\begin{aligned} s(\boldsymbol{z}, \boldsymbol{x}_j) &= s(\boldsymbol{x}, \boldsymbol{x}_j) + s(\boldsymbol{x}, \boldsymbol{x}_j)s(\boldsymbol{y}_j, \boldsymbol{x}_j) = 0, \\ s(\boldsymbol{z}, \boldsymbol{y}_j) &= s(\boldsymbol{x}, \boldsymbol{y}_j) - s(\boldsymbol{x}, \boldsymbol{y}_j)s(\boldsymbol{x}_j, \boldsymbol{y}_j) = 0, \end{aligned}$$

así que $\boldsymbol{z} \in M_{2r}^{\perp}$. Por lo tanto,

$$\boldsymbol{x} = \left(\sum_{i=1}^{r} s(\boldsymbol{x}, \boldsymbol{y}_i)\boldsymbol{x}_i - \sum_{i=1}^{r} s(\boldsymbol{x}, \boldsymbol{x}_i)\boldsymbol{y}_i \right) + \boldsymbol{z} \in M_{2r} \oplus M_{2r}^{\perp}.$$

Como $\boldsymbol{x} \in V$ era arbitrario, esto dice que $V = M_{2r} \oplus M_{2r}^{\perp}$.

En particular, se obtiene $\dim M_{2r}^{\perp} = n - 2r$. Tomando (4.17) en cuenta, la matriz de s con respecto a la descomposición $V = M_{2r} \oplus M_{2r}^{\perp}$ es una suma directa de bloques:

$$\begin{bmatrix} J_{2r} & 0 \\ 0 & C_{n-2r} \end{bmatrix}, \quad \text{donde} \quad C_{n-2r} \in M_{n-2r}(\mathbb{F}).$$

Si $2r < k$, entonces $C_{n-2r} \neq 0$ y la restricción de s al subespacio $M_{2r}^{\perp}$ es una forma bilineal alternada *no nula*. En este caso, se puede repetir el argumento inicial (que produjo los vectores $\boldsymbol{x}_1, \boldsymbol{y}_1$) para obtener dos vectores independientes $\boldsymbol{x}_{r+1}, \boldsymbol{y}_{r+1} \in M_{2r}^{\perp}$ tales que

$$s(\boldsymbol{x}_{r+1}, \boldsymbol{y}_{r+1}) = 1, \quad s(\boldsymbol{y}_{r+1}, \boldsymbol{x}_{r+1}) = -1.$$

Hecho eso, $\{\boldsymbol{x}_1, \boldsymbol{y}_1, \dots, \boldsymbol{x}_r, \boldsymbol{y}_r, \boldsymbol{x}_{r+1}, \boldsymbol{y}_{r+1}\}$ es una base de un subespacio $M_{2r+2} \leqslant V$; y la restricción de s a este subespacio tiene matriz J_{2r+2}.

Este proceso inductivo se puede repetir hasta llegar a J_{2m}, donde $\underline{2m = k}$. (Fíjese que si k fuera impar, con $k = 2m + 1$, sería imposible elegir dos vectores independientes en $M_{2m}^{\perp}$ en los cuales s no se anula.) Como el bloque J_{2m} ya tiene $2m = k$ columnas linealmente independientes, el bloque remanente es nulo: $C_{n-2m} = 0$. Ahora es cuestión de elegir una base cualquiera $\{\boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-2m}\}$ para $M_{2m}^{\perp}$ para obtener así la base deseada de V. □

A veces es más cómodo permutar los vectores de la base $\mathcal{B}$ obtenida en la Proposición anterior para cambiarla a $\mathcal{C} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m, \boldsymbol{y}_1, \dots, \boldsymbol{y}_m, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-2m}\}$ y la cual la matriz de s en esta base tiene el formato:

$$[s]_{\mathcal{C}}^{\mathcal{C}} = \begin{bmatrix} 0 & 1_m & 0 \\ -1_m & 0 & 0 \\ 0 & 0 & 0_{n-2m} \end{bmatrix}.$$

Corolario 4.38. *El rango de una matriz* antisimétrica $\underline{A = -A^{\mathrm{t}}} \in M_n(\mathbb{F})$ *es* par. *Dos matrices antisimétricas reales $A, B \in M_n(\mathbb{F})$ son congruentes si y solo si tienen* el mismo rango.

Demostración. Defínase una forma bilineal s_A sobre $\mathbb{F}^n$ por $s_A(\boldsymbol{x}, \boldsymbol{y}) := \boldsymbol{x}^{\mathrm{t}} A \boldsymbol{y}$. Si $A^{\mathrm{t}} = -A$, entonces s_A es una forma bilineal alternada. La Proposición 4.37 anterior dice que hay una matriz de cambio de base $P = [1_n]_{\mathcal{E}}^{\mathcal{B}}$ tal que $P^{\mathrm{t}} A P = J_{2m} \oplus 0_{n-2m}$, para algún m con $2m \leqslant n$. El rango de A es $r(A) = r(P^{\mathrm{t}} A P) = 2m$.

Si $A^{\mathrm{t}} = -A$, $B^{\mathrm{t}} = -B$ y $r(A) = r(B) = 2m$, entonces la misma Proposición, aplicada a las formas alternadas s_A y s_B, muestra que $A \simeq J_{2m} \oplus 0_{n-2m} \simeq B$. □

Corolario 4.39. *Existe una forma bilineal* simpléctica $s \colon V \times V \to \mathbb{F}$ *solo si* $\dim V$ *es* par*: la condición $n = 2m$ es necesaria.* ⊟

► La "forma canónica" (4.16) de una matriz antisimétrica tiene una consecuencia importante, que se manifiesta al examinar la relación de congruencia. Sea $A = -A^{\mathrm{t}} \in M_n(\mathbb{F})$ una matriz antisimétrica *invertible*. Entonces $s_A(\boldsymbol{x}, \boldsymbol{y}) := \boldsymbol{x}^{\mathrm{t}} A \boldsymbol{y}$ es una forma simpléctica (por ser no degenerada) sobre $\mathbb{F}^n$, y el rango de A es $n = 2m$. Entonces $A \doteq J_{2m}$ por la Proposición 4.37. Concretamente, hay una matriz invertible R tal que $A = R^{\mathrm{t}} J_{2m} R$. Fíjese que $\det J_{2m} = (\det J_2)^m = 1$. Como $\det R^{\mathrm{t}} = \det R$, esto dice que

$$\det A = (\det R)^2. \tag{4.18}$$

Si la matriz antisimétrica es singular, con $\det A = 0$, la fórmula (4.18) sigue válida; por ejemplo, se puede tomar para R alguna matrix singular con $\det R = 0$.

Esta fórmula (4.18) asegura que hay una matriz R tal que el cuadrado de $\det R$ es $\det A$. Algo que no es evidente, pero resulta ser cierto, es que $\underline{\det R}$ es *un polinomio con coeficientes enteros en las entradas* de A. ⟦ Esto significa, por ejemplo, que si las entradas de A son números enteros, entonces $\det R$ es un número entero. ⟧ Sucede que $\det R$ es calculable por la evaluación en las entradas de A de un polinomio "universal", muy análogo al polinomio (1.19) que define el determinante de A por la fórmula de Leibniz.

► La definición de $\underline{\det R}$ requiere otro inciso de la teoría de polinomios. Para $i, j = 1, \dots, n$ con $i < j$, tómese una *incógnita* t_{ij}, y sea $\boldsymbol{t} := (t_{12}, t_{13}, \dots, t_{n-1,n})$. Los cocientes de polinomios en estas incógnitas, con coeficientes racionales y denominador no nulo,[7] conforman un cuerpo $\mathbb{Q}(\boldsymbol{t})$. ⟦ En el cuerpo $\mathbb{Q}(\boldsymbol{t})$, está claro que $1 + 1 \neq 0$: la teoría ya vista es aplicable al caso $\mathbb{F} = \mathbb{Q}(\boldsymbol{t})$. ⟧ Considérese la siguiente matriz antisimétrica T con entradas en ese cuerpo:

$$T := \begin{bmatrix} 0 & t_{12} & t_{13} & \dots & t_{1n} \\ -t_{12} & 0 & t_{23} & \dots & t_{2n} \\ -t_{13} & -t_{23} & 0 & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -t_{1n} & -t_{2n} & -t_{3n} & \dots & 0 \end{bmatrix} \in M_n(\mathbb{Q}(\boldsymbol{t})).$$

Se deduce que $\underline{\det T = (\det S)^2}$ para cierta matriz S con entradas en $\mathbb{Q}(\boldsymbol{t})$.

De la fórmula (1.19) se sabe que $\det S$ es un polinomio en las entradas de S, así que $\det S = q(\boldsymbol{t})/r(\boldsymbol{t})$ donde q, r son polinomios en $\mathbb{Z}[\boldsymbol{t}]$ sin factor común. También por (1.19), resulta que $\det T =: s(\boldsymbol{t})$ es otro polinomio en las incógnitas $\boldsymbol{t}$. La relación $\det T = (\det S)^2$ entonces implica que

$$s(\boldsymbol{t}) r(\boldsymbol{t})^2 = q(\boldsymbol{t})^2. \tag{4.19}$$

Los tres términos en esta igualdad: $q(\boldsymbol{t})$, $r(\boldsymbol{t})$, $s(\boldsymbol{t})$ son elementos de $\mathbb{Z}[\boldsymbol{t}]$, un anillo de polinomios con coeficientes enteros en varias variables. Este anillo admite *factorización*

[7] Al multiplicar el numerador y denominador de un tal cociente por un número entero apropiado, se puede suponer que tanto el numerador como el denominador tienen coeficientes en $\mathbb{Z}$: ellos pertenecen al anillo de polinomios $\mathbb{Z}[\boldsymbol{t}]$.

única en productos de polinomios irreducibles.[8] Entonces, por (4.19), cada factor irreducible de $r(\boldsymbol{t})$ es también un factor de $q(\boldsymbol{t})$. Como $q(\boldsymbol{t})$ y $r(\boldsymbol{t})$ no tienen factor común, esto dice que $r(\boldsymbol{t}) \equiv \pm 1$. En consecuencia, $\underline{\det T = q(\boldsymbol{t})^2}$ *es un cuadrado perfecto* en $\mathbb{Z}[\boldsymbol{t}]$.

Definición 4.40. Sea $A = -A^{\mathsf{t}} \in M_n(\mathbb{F})$ una matriz *antisimétrica*, donde $\underline{n = 2m}$ es par. La evaluación de polinomios $t_{ij} \mapsto a_{ij}$ lleva $\mathbb{Z}[\boldsymbol{t}]$ en $\mathbb{F}$ y lleva la matriz T en A. El polinomio en las entradas a_{ij} que es la imagen de $\underline{\det S} \in \mathbb{Z}[\boldsymbol{t}]$ es el **pfaffiano** de A:

$$\underline{\operatorname{Pf} A} := q(a_{12}, a_{13}, \dots, a_{n-1,n}).$$

Para resolver la ambigüedad de signo en $q(\boldsymbol{t})$, se requiere adicionalmente imponer la condición de que $\underline{\operatorname{Pf}(J_{2m}) = +1}$. ◊

La evaluación de polinomios preserva productos; en consecuencia,[9]

$$\boxed{\det A = (\operatorname{Pf} A)^2.} \tag{4.20}$$

Ejemplo 4.41. En el caso $n = 2$, con $A = \begin{bmatrix} 0 & a_{12} \\ -a_{12} & 0 \end{bmatrix}$, lo cual implica $\det A = a_{12}^2$, así que $\operatorname{Pf} A = \pm a_{12}$. El signo queda determinado por la condición $\operatorname{Pf} J_2 = +1$. Luego, $\underline{\operatorname{Pf} A = a_{12}}$. ◊

Ejemplo 4.42. En el caso $n = 4$, el pfaffiano se calcula sin dificultad:

$$A = \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{bmatrix} \implies \boxed{\operatorname{Pf} A = a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23}.} \tag{4.21}$$

Nótese que $\det A$ es un polinomio de grado 4 en las entradas de A, y que $\operatorname{Pf} A$ es un polinomio de grado 2, como la fórmula (4.20) hace esperar. ◊

Si $A = -A^{\mathsf{t}} \in M_n(\mathbb{F})$ es una matriz antisimétrica cuyo lado $\underline{n = 2m+1}$ es *impar*, la Proposición 4.37 muestra que $A \simeq J_{2r} \oplus 0_{2m-2r+1}$ para algún $r \leqslant m$, y en particular que A *no es invertible*, pues $\det A = 0$. En este caso conviene definir $\underline{\operatorname{Pf} A := 0}$ también.

[8] Es un teorema de álgebra que los polinomios en varias variables $\mathbb{Z}[t_{12}, \dots, t_{n-1,n}]$ heredan la factorización única del anillo de coeficientes $\mathbb{Z}$. Véase, por ejemplo, el Teorema 2.25 del libro: Nathan Jacobson, *Basic Algebra*, Dover Books, Mineola, NY, 1985. Nótese que en $\mathbb{Z}$, la factorización es única hasta un signo: por ejemplo, $6 = 3 \cdot 2 = (-3)(-2)$; lo mismo sucede con $\mathbb{Z}[\boldsymbol{t}]$.

[9] El polinomio $\operatorname{Pf} A$ fue bautizado **pfaffiano** por Arthur Cayley, en honor al matemático alemán *Johann Friedrich Pfaff*. Cayley demostró la fórmula (4.20) en 1849 mientras estudiaba ciertos trabajos de Pfaff sobre sistemas de ecuaciones diferenciales ordinarias. Como nota lingüística, los consonantes dobles 'pf' y 'ff' se pueden aproximar en español por 'f' solo: *fafiano*.

Ejemplo 4.43. La *fórmula general* para el pfaffiano de una matriz antisimétrica $A = -A^{\mathrm{t}}$ en $M_{2m}(\mathbb{F})$ es la siguiente:

$$\operatorname{Pf} A := \frac{1}{2^m\, m!} \sum_{\sigma} (-1)^{\sigma}\, a_{\sigma(1)\sigma(2)}\, a_{\sigma(3)\sigma(4)}\, \dots\, a_{\sigma(2m-1)\sigma(2m)}. \tag{4.22}$$

Aquí σ recorre las $(2m)!$ permutaciones de $(1, 2, \dots, 2m)$; debido a la antisimetría de A, la sumatoria tiene muchos productos repetidos, y el factor $1/(2^m\, m!)$ sirve para eliminar redundancias en esta sumatoria. La comprobación de (4.22) aparecerá en los Ejercicios. ◊

Proposición 4.44. *Si $A = -A^{\mathrm{t}}$ es una matriz antisimétrica en $M_n(\mathbb{F})$ y si $R \in M_n(\mathbb{F})$ es una matriz cualquiera, entonces*

$$\boxed{\operatorname{Pf}(R^{\mathrm{t}} A R) = (\det R)\,(\operatorname{Pf} A).} \tag{4.23}$$

Demostración. Si n es impar, los dos lados de la ecuación son iguales a 0. Supóngase, entonces, que n es par, $n = 2m$.

Obsérvese que $(R^{\mathrm{t}} A R)^{\mathrm{t}} = R^{\mathrm{t}} A^{\mathrm{t}} R = -R^{\mathrm{t}} A R$: la matriz $R^{\mathrm{t}} A R$ es también antisimétrica y posee un pfaffiano. Ahora

$$\det(R^{\mathrm{t}} A R) = (\det R^{\mathrm{t}})(\det A)(\det R) = (\det R)^2 \det A.$$

De (4.20) se obtiene de inmediato:

$$\operatorname{Pf}(R^{\mathrm{t}} A R) = \pm(\det R)\,(\operatorname{Pf} A).$$

Solo falta comprobar que el signo al lado derecho es $+1$ y no -1.

Los dos lados de la última igualdad se obtienen por la evaluación de una identidad polinomial, con T en lugar de A y una matriz análoga en lugar de R – sus entradas son nuevas incógnitas r_{ij}. Esto significa que el signo al lado derecho es el mismo, cualesquiera que sean A y R. Al tomar $R = 1_n$, y al recordar que $\det 1_n = +1$, se ve que el signo es positivo: la fórmula (4.23) queda comprobada. □

► Para cerrar este capítulo sobre formas bilineales, es oportuno retomar el caso de escalares reales, $\mathbb{F} = \mathbb{R}$, para abordar una situación en la cual las formas bilineales simétricas y alternadas juegan papeles complementarios.

Considérese un espacio de Hilbert W, de dimensión m sobre $\mathbb{C}$, con un producto escalar *complejo* $\langle -, - \rangle$. A veces conviene considerar W como un espacio vectorial *real*, de dimensión $2m$ sobre $\mathbb{R}$, haciendo caso omiso de las multiplicaciones escalares $\boldsymbol{x} \mapsto \pm i\boldsymbol{x}$ donde $i = \sqrt{-1}$. El producto escalar entonces define dos formas $\mathbb{R}$-bilineales, d y s, así:

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle =: \underline{d(\boldsymbol{x}, \boldsymbol{y})} + i\, \underline{s(\boldsymbol{x}, \boldsymbol{y})}. \tag{4.24}$$

La parte real $\underline{d}$ es bilineal (sobre $\mathbb{R}$) y simétrica, pero la parte imaginaria $\underline{s}$ es alternante:

$$d(\boldsymbol{y},\boldsymbol{x}) + i\,s(\boldsymbol{y},\boldsymbol{x}) = \langle \boldsymbol{y},\boldsymbol{x}\rangle = \overline{\langle \boldsymbol{x},\boldsymbol{y}\rangle} = d(\boldsymbol{x},\boldsymbol{y}) - i\,s(\boldsymbol{x},\boldsymbol{y}).$$

La positividad definida del producto escalar implica que d y s son no degeneradas.

El interés de la fórmula (4.24) reside en *el problema inverso*, que empieza con *un espacio vectorial real* V de dimensión par $n = 2m$, dotado con un producto escalar real $d\colon V\times V \to \mathbb{R}$ o bien una forma bilineal simpléctica $s\colon V\times V \to \mathbb{R}$, y se quiere transformar V en *un espacio de Hilbert* de dimensión m sobre $\mathbb{C}$.

Se requiere una manera de prescribir la multiplicación escalar por $\pm i$ sobre V. Además, con d dada hace falta definir s; o bien, si s es dada, hace falta definir d. En los dos casos, las recetas que satisfacen estos requisitos son parecidas, pero con sutiles diferencias.

Definición 4.45. Sea V un espacio vectorial real de dimensión par $n = 2m$; y sea $d\colon V\times V \to \mathbb{R}$ una forma bilineal simétrica *definida positiva* sobre V. Una **estructura compleja ortogonal** sobre (V, d) es un operador $\mathbb{R}$-lineal $J \in \mathcal{L}(V)$ tal que

$$\underline{J^2 = -1_V} \quad \text{y} \quad \underline{d(J\boldsymbol{x}, J\boldsymbol{y}) = d(\boldsymbol{x},\boldsymbol{y})} \quad \text{para todo } \boldsymbol{x},\boldsymbol{y} \in V. \tag{4.25}$$

Por la ley de inercia, Teorema 4.22, hay una base $\mathcal{E} = \{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$ de V tal que

$$d(\boldsymbol{x},\boldsymbol{y}) = x_1y_1 + \cdots + x_ny_n, \quad \text{cuando} \quad \begin{cases} \boldsymbol{x} = x_1\,\boldsymbol{e}_1 + \cdots + x_n\,\boldsymbol{e}_n, \\ \boldsymbol{y} = y_1\,\boldsymbol{e}_1 + \cdots + y_n\,\boldsymbol{e}_n. \end{cases}$$

Al identificar $\boldsymbol{x} \in V$ con el vector columna $(x_1,\dots,x_n)^{\mathrm{t}} \in \mathbb{R}^n$, se ve que $d(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{t}}\boldsymbol{y}$. ◊

El operador buscado $J \in \mathcal{L}(V)$ tendrá una matriz $[J]^{\mathcal{E}}_{\mathcal{E}} \in M_n(\mathbb{R})$; con un leve abuso de notación, se puede denotar esta matriz con la misma letra J. Las propiedades que debe cumplir la matriz $J \in M_n(\mathbb{R})$ son estas dos:

$$\underline{J^2 = -1_n} \quad \text{y} \quad \underline{J^{\mathrm{t}}J = 1_n}\,.$$

La segunda relación viene del cálculo siguiente:

$$\boldsymbol{x}^{\mathrm{t}}J^{\mathrm{t}}J\boldsymbol{y} = (J\boldsymbol{x})^{\mathrm{t}}J\boldsymbol{y} = d(J\boldsymbol{x},J\boldsymbol{y}) = d(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^{\mathrm{t}}\boldsymbol{y} \quad \text{para todo} \quad \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n.$$

Esto dice que J es *una matriz ortogonal*; que además cumple $\underline{J^{\mathrm{t}} = -J} = J^{-1}$.

Ejemplo 4.46. Si $V = \mathbb{R}^{2m}$, se puede tomar $J := J_{2m}$ como en (4.16), la suma directa de m copias de la matriz J_2. Entonces

$$J_{2m}^2 = J_2^2 \oplus \cdots \oplus J_2^2 = (-1_2) \oplus \cdots \oplus (-1_2) = -1_{2m}$$

y J_{2m} es antisimétrica, $J_{2m}^{\mathrm{t}}J_{2m} = -J_{2m}^2 = 1_{2m}$, luego J_{2m} es ortogonal. ◊

Lema 4.47. *Sea (V, d) un espacio vectorial real de dimensión $2m$ con un producto escalar real d, y sea $J \in \mathcal{L}(V)$ una estructura compleja ortogonal. Entonces $s(\boldsymbol{x}, \boldsymbol{y}) := d(J\boldsymbol{x}, \boldsymbol{y})$ define una forma bilineal simpléctica $s\colon V \times V \to \mathbb{R}$ tal que $s(J\boldsymbol{x}, J\boldsymbol{y}) = s(\boldsymbol{x}, \boldsymbol{y})$ para $\boldsymbol{x}, \boldsymbol{y} \in V$.*

Demostración. Es evidente que s es $\mathbb{R}$-bilineal. Si $\boldsymbol{x}, \boldsymbol{y} \in V$, entonces

$$s(\boldsymbol{y}, \boldsymbol{x}) = d(J\boldsymbol{y}, \boldsymbol{x}) = d(J^2\boldsymbol{y}, J\boldsymbol{x}) = -d(\boldsymbol{y}, J\boldsymbol{x}) = -d(J\boldsymbol{x}, \boldsymbol{y}) = -s(\boldsymbol{x}, \boldsymbol{y}), \tag{4.26}$$

usando las propiedades (4.25) y la simetría de d. Por lo tanto, s es alternada.

Si $s(\boldsymbol{x}, \boldsymbol{y}) = 0$ para todo $\boldsymbol{y} \in V$, entonces $d(J\boldsymbol{x}, \boldsymbol{y}) = 0$ para todo $\boldsymbol{y}$; en particular, $d(J\boldsymbol{x}, J\boldsymbol{x}) = 0$. Luego $J\boldsymbol{x} = \boldsymbol{0}$ por ser d definida positiva; eso implica $\boldsymbol{x} = \boldsymbol{0}$ pues J es invertible con $J^{-1} = -J$. Se ha verificado que s es no degenerada.

Además, $s(J\boldsymbol{x}, J\boldsymbol{y}) = d(J^2\boldsymbol{x}, J\boldsymbol{y}) = s(\boldsymbol{x}, \boldsymbol{y})$, en vista de (4.26) con $\boldsymbol{x} \leftrightarrow \boldsymbol{y}$. □

De manera similar, se puede empezar con un par (V, s) donde s es una forma bilineal simpléctica, introduciendo un operador J para luego definir d por una variante del Lema 4.47. En contraste con el caso anterior, la positividad definida de d no emerge de la forma alternada s, sino que se debe recargarla en la definición de J.

Definición 4.48. Sea V un espacio vectorial real de dimensión par $n = 2m$; y sea $s\colon V \times V \to \mathbb{R}$ una forma bilineal simpléctica sobre V. Una **estructura compleja positiva** sobre (V, s) es un operador $\mathbb{R}$-lineal $J \in \mathcal{L}(V)$ tal que

$$J^2 = -1_V; \quad s(J\boldsymbol{x}, J\boldsymbol{y}) = s(\boldsymbol{x}, \boldsymbol{y}) \quad \text{para } \boldsymbol{x}, \boldsymbol{y} \in V; \text{ y} \quad s(\boldsymbol{x}, J\boldsymbol{x}) > 0 \quad \text{para } \boldsymbol{x} \neq \boldsymbol{0}. \tag{4.27}$$

Por la Proposición 4.37, hay una base $\mathcal{E}$ de $V \simeq \mathbb{R}^{2m}$ tal que $[s]^{\mathcal{E}}_{\mathcal{E}} = J_{2m}$. ◊

Lema 4.49. *Sea (V, s) un espacio vectorial real de dimensión $2m$ con una forma bilineal simpléctica s, y sea $J \in \mathcal{L}(V)$ una estructura compleja positiva. Entonces $d(\boldsymbol{x}, \boldsymbol{y}) := s(\boldsymbol{x}, J\boldsymbol{y})$ define un producto escalar real $d\colon V \times V \to \mathbb{R}$ tal que $d(J\boldsymbol{x}, J\boldsymbol{y}) = d(\boldsymbol{x}, \boldsymbol{y})$ para $\boldsymbol{x}, \boldsymbol{y} \in V$.*

Demostración. Es evidente que d es $\mathbb{R}$-bilineal. Si $\boldsymbol{x}, \boldsymbol{y} \in V$, entonces

$$d(\boldsymbol{y}, \boldsymbol{x}) = s(\boldsymbol{y}, J\boldsymbol{x}) = s(J\boldsymbol{y}, J^2\boldsymbol{x}) = -s(J\boldsymbol{y}, \boldsymbol{x}) = s(\boldsymbol{x}, J\boldsymbol{y}) = d(\boldsymbol{x}, \boldsymbol{y}), \tag{4.28}$$

usando las propiedades (4.27) y la antisimetría de s. Por lo tanto, d es simétrica.

Fíjese que $d(\boldsymbol{x}, \boldsymbol{x}) = s(\boldsymbol{x}, J\boldsymbol{x}) \geqslant 0$, con igualdad si y solo si $\boldsymbol{x} = \boldsymbol{0}$, debido a (4.27). Se ha verificado que d es definida positiva.

Además, $d(J\boldsymbol{x}, J\boldsymbol{y}) = s(J\boldsymbol{x}, J^2\boldsymbol{y}) = s(\boldsymbol{x}, J\boldsymbol{y}) = d(\boldsymbol{x}, \boldsymbol{y})$, también por (4.27). □

► Hasta ahora, se ha comprobado una reciprocidad entre las formas bilineales d y s, basada en los requisitos (4.25) o (4.27), según el caso, sobre el operador J. El último paso es el uso de J, cuyo cuadrado es -1, para introducir una multiplicación escalar compleja.

Proposición 4.50. *Sea V un espacio vectorial real de dimensión $2m$, dotado con una de estas estructuras:*

(a) *un producto escalar real $\underline{d}$ y una estructura compleja ortogonal J; o bien*

(b) *una forma bilineal simpléctica $\underline{s}$ y una estructura compleja positiva J.*

Defínase una multiplicación escalar compleja sobre V por:

$$\boxed{(a+ib)\boldsymbol{x} := a\,\boldsymbol{x} + b\,J\boldsymbol{x},} \quad \textit{para todo} \quad a, b \in \mathbb{R}. \tag{4.29a}$$

Entonces V es un espacio de Hilbert, de dimensión m sobre $\mathbb{C}$, bajo el siguiente producto escalar:

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} \rangle &:= d(\boldsymbol{x}, \boldsymbol{y}) + i\,s(\boldsymbol{x}, \boldsymbol{y}) \\ &= d(\boldsymbol{x}, \boldsymbol{y}) + i\,d(J\boldsymbol{x}, \boldsymbol{y}) \\ &= s(\boldsymbol{x}, J\boldsymbol{y}) + i\,s(\boldsymbol{x}, \boldsymbol{y}). \end{aligned} \tag{4.29b}$$

⟦ *La segunda fórmula es aplicable en el caso (a) y la tercera fórmula en el caso (b).* ⟧

Demostración. Es fácil comprobar que la operación (4.29a) hace de V un espacio vectorial complejo: solo es necesario observar lo siguiente para $a, b, p, q \in \mathbb{R}$, $\boldsymbol{x} \in V$:

$$\begin{aligned} (a+ib)(p+iq)\,\boldsymbol{x} &= (a+ib)(p\,\boldsymbol{x} + q\,J\boldsymbol{x}) = ap\,\boldsymbol{x} + (aq+bp)\,J\boldsymbol{x} + bq\,J^2\boldsymbol{x} \\ &= (ap-bq)\,\boldsymbol{x} + (aq+bp)\,J\boldsymbol{x} \\ &= ((ap-bq) + i(aq+bp))\,\boldsymbol{x}. \end{aligned}$$

Las Lemas 4.47 y 4.49 hacen ver que los casos (a) y (b) del enunciado son equivalentes: se dispone de las dos formas d y s con las propiedades enunciadas. Entonces la función $\underline{h := d + is}\colon V \times V \to \mathbb{C}$ es una *forma hermítica* definida positiva sobre V; esto dice que $h = \langle -, - \rangle$ es un producto escalar complejo sobre V, así que (V, h) es un espacio de Hilbert.

La dimensión de este espacio de Hilbert se determina al exhibir una base ortonormal. Si $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_r\}$ es una familia ortonormal de vectores en V (la cual existe al menos cuando $r = 1$), tómese el conjunto de $2r$ vectores no ceros $\mathcal{B}_{2r} = \{\boldsymbol{u}_1, J\boldsymbol{u}_1, \dots, \boldsymbol{u}_r, J\boldsymbol{u}_r\} \subset V$. Por ser

$$d(\boldsymbol{u}_i, \boldsymbol{u}_j) + i\,d(J\boldsymbol{u}_i, \boldsymbol{u}_j) = \langle \boldsymbol{u}_i, \boldsymbol{u}_j \rangle,$$

se ve que $d(\boldsymbol{u}_i, \boldsymbol{u}_j) = d(J\boldsymbol{u}_i, \boldsymbol{u}_j) = 0$ si $i \neq j$; que $d(J\boldsymbol{u}_i, \boldsymbol{u}_i) = s(\boldsymbol{u}_i, \boldsymbol{u}_i) = 0$; y que $d(\boldsymbol{u}_i, \boldsymbol{u}_i) = \langle \boldsymbol{u}_i, \boldsymbol{u}_i \rangle = 1$ para $i = 1, \dots, r$. Entonces $\mathcal{B}_{2r}$ es una *base ortonormal real* de un subespacio $M_{2r} \leqslant V$, con dimensión real $2r$.

Si $r < m$, existe $\boldsymbol{u}_{r+1} \in V$ tal que $d(\boldsymbol{u}_{r+1}, \boldsymbol{u}_{r+1}) = 1$ y $d(\boldsymbol{u}_{r+1}, \boldsymbol{u}_i) = d(\boldsymbol{u}_{r+1}, J\boldsymbol{u}_i) = 0$ para $i = 1, \dots, r$, por compleción de una base ortonormal real en V (usando el Corolario 3.19

en el caso real). Es evidente que $d(J\boldsymbol{u}_{r+1}, J\boldsymbol{u}_{r+1}) = d(\boldsymbol{u}_{r+1}, \boldsymbol{u}_{r+1}) = 1$ y que $d(J\boldsymbol{u}_{r+1}, \boldsymbol{u}_i) = d(J\boldsymbol{u}_{r+1}, J\boldsymbol{u}_i) = 0$ si $i \leqslant r$, así que $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_{r+1}\}$ es una familia ortonormal (compleja) en V.

Al llegar a $r = m$, se ha construido una base ortonormal real $\mathcal{B}_{2m} = \{\boldsymbol{u}_1, J\boldsymbol{u}_1, \dots, \boldsymbol{u}_m, J\boldsymbol{u}_m\}$ de (V, d) y al mismo tiempo una base ortonormal compleja $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_m\}$ del espacio de Hilbert (V, h). De paso, se ha verificado que $\dim_{\mathbb{C}} V = m$. □

► En esta sección se ha visto dos recetas que convierte un espacio vectorial real V de dimensión par $n = 2m$, con ciertas estructuras extras, en un espacio de Hilbert de dimensión compleja m. Ahora bien, hay otro artificio que produce un espacio de Hilbert de dimensión compleja $2m$ a partir de V, la llamada *complexificación* de V, que se discute brevemente a continuación.

Definición 4.51. Sea V un espacio vectorial real de dimensión n. La **complexificación** de V es el espacio vectorial complejo

$$\underline{V_{\mathbb{C}} = V \oplus iV} := \{\, \boldsymbol{x} + i\,\boldsymbol{y} : \boldsymbol{x}, \boldsymbol{y} \in V \,\},$$

con multiplicación escalar "ordinaria":

$$(a + ib)(\boldsymbol{x} + i\boldsymbol{y}) := (a\boldsymbol{x} - b\boldsymbol{y}) + i(b\boldsymbol{x} + a\boldsymbol{y}) \quad \text{para} \quad a, b \in \mathbb{R};\ \boldsymbol{x}, \boldsymbol{y} \in V.$$

Fíjese que $V_{\mathbb{C}} = V \oplus iV$ tiene dimensión n sobre $\mathbb{C}$.

Una forma $\mathbb{R}$-bilineal simétrica $d\colon V \times V \to \mathbb{R}$ se puede *ampliar* en una forma $\mathbb{C}$-bilineal simétrica $d\colon V_{\mathbb{C}} \times V_{\mathbb{C}} \to \mathbb{C}$ al declarar que

$$d(\boldsymbol{x} + i\boldsymbol{y}, \boldsymbol{x}' + i\boldsymbol{y}') := d(\boldsymbol{x}, \boldsymbol{x}') + i\, d(\boldsymbol{x}, \boldsymbol{y}') + i\, d(\boldsymbol{y}, \boldsymbol{x}') - d(\boldsymbol{y}, \boldsymbol{y}').$$

Si d es definida positiva[10] sobre V, es posible dotar $V_{\mathbb{C}}$ de un producto escalar complejo:

$$\langle\!\langle \boldsymbol{z}, \boldsymbol{w} \rangle\!\rangle := 2\, d(\bar{\boldsymbol{z}}, \boldsymbol{w}) \quad \text{para} \quad \boldsymbol{z}, \boldsymbol{w} \in V_{\mathbb{C}} \tag{4.30}$$

donde se ha escrito $\underline{\bar{\boldsymbol{z}} := \boldsymbol{x} - i\boldsymbol{y}}$ cuando $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y} \in V_{\mathbb{C}}$. ◊

Escolio 4.52. *Sea V un espacio $\mathbb{R}$-vectorial de dimensión $\underline{2m}$ con un producto escalar real $\underline{d}$ y una estructura compleja ortogonal J. Entonces el subespacio complejo de dimensión $\underline{m}$,*

$$W := \{\, \boldsymbol{x} - iJ\boldsymbol{x} : \boldsymbol{x} \in V \,\} \leqslant V_{\mathbb{C}}$$

es isomorfo a V como espacio de Hilbert. En efecto, el operador $P := \frac{1}{2}(1 - iJ) \in \mathcal{L}(V_{\mathbb{C}})$ es un proyector ortogonal con imagen W, y se cumple $\langle\!\langle P\boldsymbol{x}, P\boldsymbol{y} \rangle\!\rangle = \overline{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}$ para todo $\boldsymbol{x}, \boldsymbol{y} \in V$. ⊟

[10] Es importante observar que la forma $\mathbb{C}$-bilineal ampliada es *indefinida* sobre $V_{\mathbb{C}}$: por el Teorema 4.22(a), una forma bilineal simétrica compleja tiene rango pero no tiene signatura.

4.5. Ejercicios sobre formas bilineales

Ejercicio 4.1. Si $f\colon V \times V \to \mathbb{R}$ es una forma bilineal sobre un espacio $\mathbb{R}$-vectorial V, y si existen $\boldsymbol{x}, \boldsymbol{y} \in V$ tales que $f(\boldsymbol{x},\boldsymbol{x}) > 0$ y $f(\boldsymbol{y},\boldsymbol{y}) < 0$, comprobar que $\boldsymbol{x}$, $\boldsymbol{y}$ son linealmente independientes. En seguida, demostrar que hay un vector $\boldsymbol{z} \in V$ con $\boldsymbol{z} \neq \boldsymbol{0}$ tal que $f(\boldsymbol{z},\boldsymbol{z}) = 0$.

Ejercicio 4.2. Demostrar directamente que estas dos matrices *no* son congruentes en $M_2(\mathbb{R})$:

$$A = \begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}, \qquad B = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}.$$

Ejercicio 4.3. Sea V un espacio $\mathbb{F}$-vectorial y sean $f\colon V \times V \to \mathbb{F}$, $g\colon V \times V \to \mathbb{F}$ dos formas bilineales. Supóngase que la forma bilineal f es no degenerada. Demostrar que hay un único operador lineal $T \in \mathcal{L}(V)$ tal que

$$g(\boldsymbol{x},\boldsymbol{y}) = f(\boldsymbol{x},T\boldsymbol{y}) \quad \text{para todo} \quad \boldsymbol{x},\boldsymbol{y} \in V.$$

Mostrar que T es biyectivo si y solo si g también es no degenerada.

⟦ Indicación: tomar una base fija $\mathcal{B}$ de V y expresar la matriz de T en términos de las matrices de f y g. ⟧

Ejercicio 4.4. El **discriminante** de una forma bilineal simétrica d, con respecto a una base $\mathcal{B} = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ de V, es el determinante $\underline{D := \det[d(\boldsymbol{x}_i,\boldsymbol{x}_j)]}$. Verificar que la forma d es no degenerada si y solo si $D \neq 0$.

Ejercicio 4.5. Sea d una forma bilineal simétrica sobre V. Para cada subespacio $M \leqslant V$, denótese por $M^\perp$ el subespacio ortogonal a M con respecto a d, dado por la fórmula (4.5).

(a) Si N es otro subespacio de V, demostrar que $(M + N)^\perp = M^\perp \cap N^\perp$.

(b) Demostrar también que $(M \cap N)^\perp = M^\perp + N^\perp$ si d es no degenerada.

Ejercicio 4.6. Considerar la siguiente forma cuadrática sobre $\mathbb{R}^3$:

$$q(\boldsymbol{x}) = x_1^2 + 4x_1x_3 + 2x_2^2 - 2x_3^2\cdot,$$

(a) Hallar una matriz simétrica $A \in M_3(\mathbb{R})$ tal que $q(\boldsymbol{x}) = \boldsymbol{x}^{\mathrm{t}}A\boldsymbol{x}$.

(b) Encontrar una base ortonormal de $\mathbb{R}^3$ formado por tres autovectores de A.

(c) Encontrar una matriz *ortogonal* $Q \in M_3(\mathbb{R})$ tal que $B = Q^{\mathrm{t}}AQ$ es diagonal.[11]

[11] Este ejercicio ejemplifica el resultado (no demostrado) de que una matriz simétrica real es congruente y similar a una matriz diagonal.

Ejercicio 4.7. Demostrar que un forma bilineal $s\colon V \times V \to \mathbb{F}$ es alternante si y solo si $s(\boldsymbol{x}, \boldsymbol{x}) = 0$ para todo $\boldsymbol{x} \in V$.

Ejercicio 4.8. Aplicar la reducción de Lagrange para expresar las formas cuadráticas siguientes como una combinación de cuadrados de formas lineales:

(a) $q(x_1, x_2) = 4x_1x_2$,

(b) $q(x_1, x_2, x_3) = x_1^2 + 2x_1x_2 - 4x_1x_3 + 2x_2^2 - x_3^2$,

(c) $q(x_1, x_2, x_3) = -2x_1^2 + 6x_1x_2 + 10x_1x_3 + x_2^2 - 2x_2x_3 + 4x_3^2$,

(d) $q(x_1, x_2, x_3) = x_1x_2 + x_1x_3 + x_2x_3$,

(e) $q(x_1, x_2, x_3, x_4) = x_1^2 + 2x_1x_2 - 2x_1x_3 + 2x_1x_4 + 4x_2^2 + 4x_2x_3 + 2x_2x_4 + 4x_3^2 - 2x_3x_4 - x_4^2$,

(f) $q(x_1, x_2, x_3, x_4) = 4x_1^2 - 4x_1x_2 - 4x_1x_3 + 4x_1x_4 + x_2^2 + 4x_2x_3 - 4x_2x_4 + x_3^2 + x_4^2$.

Ejercicio 4.9. ¿Cuáles son el rango y la signatura de cada una de las formas cuadráticas del Ejercicio 4.8 anterior?

Ejercicio 4.10. Determinar el rango y la signatura de las siguientes formas cuadráticas sobre $\mathbb{R}^n$:

(a) $q(\boldsymbol{x}) = x_1x_2 + x_3x_4 + x_5x_6 + \cdots + x_{2m-1}x_{2m}$, si $n = 2m$;

(b) $\tilde{q}(\boldsymbol{x}) = \sum_{i<j} (x_i - x_j)^2 = (x_1 - x_2)^2 + (x_1 - x_3)^2 + (x_2 - x_3)^2 + \cdots + (x_{n-1} - x_n)^2$.

Ejercicio 4.11. Si $a, b \in \mathbb{R}$, defínase:

$$
\begin{aligned}
q(\boldsymbol{x}) &:= a \sum_{k=1}^{n} x_k^2 + b \sum_{i<j} x_i x_j \\
&= a(x_1^2 + \cdots + x_n^2) + b(x_1x_2 + x_1x_3 + x_2x_3 + x_1x_4 + \cdots + x_{n-1}x_n).
\end{aligned}
$$

Determinar el rango y la signatura de q. (Hay varios casos, según los valores de a y b.)

Ejercicio 4.12. Cada forma cuadrática no degenerada q sobre $\mathbb{R}^2$ define una **cónica** (centrada en el origen), la cual es la curva cuya ecuación es $q(\boldsymbol{x}) = 1$, o bien $d(\boldsymbol{x}, \boldsymbol{x}) = 1$. Si $\boldsymbol{y} \in \mathbb{R}^2$ es un punto dado, la **recta polar** de $\boldsymbol{y}$ con respecto a esta cónica es la recta con ecuación $d(\boldsymbol{y}, \boldsymbol{x}) = 1$.

Por ejemplo, si la cónica es la hipérbola $x_1^2 - 4x_1x_2 + 2x_2^2 = 1$, la recta polar del punto $\boldsymbol{y} = (2, 3)$ es la recta $-4x_1 + 2x_2 = 1$.

Verificar que $\boldsymbol{y} \in \mathbb{R}^2$ es un punto de la curva $q(\boldsymbol{x}) = 1$ si y solo si $\boldsymbol{y}$ es un punto de su propia recta polar. Concluir que esa recta polar es tangencial a la cónica en ese punto.[12]

[12] Una recta es tangente a una cónica si lo corta una sola vez.

Ejercicio 4.13. Demostrar que la forma cuadrática

$$q(x_1, x_2, x_3) := x_1^2 + 4x_1x_2 + 3x_2^2 + 2x_2x_3 + 6x_3^2$$

no es definida positiva. Dar un ejemplo de un vector $\boldsymbol{x} \in \mathbb{R}^3$ tal que $q(\boldsymbol{x}) < 0$.

Ejercicio 4.14. Sean $\alpha_1, \dots, \alpha_r$ los autovalores distintos de la matriz simétrica $A \in M_n(\mathbb{R})$, en orden decreciente: $\alpha_1 > \alpha_2 > \cdots > \alpha_r$. Demostrar que la forma cuadrática $\underline{q(\boldsymbol{x}) := \boldsymbol{x}^t A\boldsymbol{x}}$ cumple las desigualdades

$$\alpha_r\, \boldsymbol{x}^t\boldsymbol{x} \leqslant \boldsymbol{x}^t A\boldsymbol{x} \leqslant \alpha_1\, \boldsymbol{x}^t\boldsymbol{x}.$$

⟦ Indicación: se sabe que hay una matriz ortogonal Q tal que $Q^{-1}AQ$ es diagonal. ⟧

La esfera unitaria $\mathbb{S}^{n-1} \subset \mathbb{R}^n$ tiene ecuación $\underline{\boldsymbol{x}^t\boldsymbol{x} = 1}$. Comprobar que los valores máximo y mínimo de la función $q(\boldsymbol{x})$ sobre esta esfera son α_1 y α_r, respectivamente.

Ejercicio 4.15. Encontrar una matriz ortogonal $Q \in M_3(\mathbb{R})$ con $Q^{-1}AQ$ diagonal, si

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

En seguida, hallar los valores máximo y mínimo de la función cuadrática

$$q(x, y, z) := x^2 + xy + xz + y^2 + yz + z^2$$

sobre la esfera unitaria $x^2 + y^2 + z^2 = 1$.

Ejercicio 4.16. Esta matriz $A \in M_4(\mathbb{Q})$ es antisimétrica:

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

Encontrar una matriz invertible $R \in M_4(\mathbb{Q})$ tal que $R^t AR = J_4$.

Ejercicio 4.17. Sea V un espacio $\mathbb{F}$-vectorial con $\dim V = 2m$ y sea $s\colon V \times V \to \mathbb{F}$ una forma simpléctica.

(a) Si M es un subespacio de V y $M^\perp$ denota su *complemento simpléctico* con respecto a s, demostrar que $(M^\perp)^\perp = M$.

(b) Demostrar que hay un subespacio $L \leqslant V$ con $\dim L = m$ que es isotrópico con respecto a s. Concluir que $\underline{L = L^\perp}$ y que L es un subespacio isotrópico maximal.[13]

⟦ Indicación: usar la base $\mathcal{B}$ construida en la Proposición 4.37 (con $n = 2m$). ⟧

Ejercicio 4.18. (a) Sea $B + C \in M_n(\mathbb{R})$ una matriz invertible, donde $B^{\mathrm{t}} = B$ y $C^{\mathrm{t}} = -C$. Sea $P := (B + C)^{-1}(B - C)$. Comprobar las relaciones de congruencia:

$$P^{\mathrm{t}}(B + C)P = B + C, \qquad P^{\mathrm{t}}(B - C)P = B - C.$$

(b) Si $A = -A^{\mathrm{t}} \in M_n(\mathbb{R})$ es una matriz antisimétrica, comprobar que 1 y (-1) *no son autovalores* de A; concluir que las matrices $(1_n - A)$ y $(1_n + A)$ son invertibles.

(c) Si $A = -A^{\mathrm{t}} \in M_n(\mathbb{R})$, demostrar que la **transformada de Cayley** de A, dada por

$$\underline{Q := (1_n + A)^{-1}(1_n - A)},$$

es una matriz ortogonal.

Ejercicio 4.19. (a) Sean $p(t), q(t) \in \mathbb{F}[t]$ dos polinomios y sea $B \in M_n(\mathbb{F})$ una matriz cuadrada. Si $p(B)$ es invertible en $M_n(\mathbb{F})$, demostrar que $p(B)^{-1}q(B) = q(B)p(B)^{-1}$.

(b) Si $Q \in M_n(\mathbb{R})$ es una matriz ortogonal y si 1 y (-1) no son autovalores de Q, demostrar que su transformada de Cayley $A := (1_n + Q)^{-1}(1_n - Q)$ es una matriz antisimétrica.

Ejercicio 4.20. Considérese el espacio vectorial $\mathbb{R}^{2m}$ con su producto escalar real estándar $\underline{d(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^{\mathrm{t}}\boldsymbol{y}}$ y su forma simpléctica estándar $\underline{s(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^{\mathrm{t}}J_{2m}\boldsymbol{y}}$. Sean $A, B \in M_{2m}(\mathbb{R})$ dos matrices tales que $AJ_{2m} = J_{2m}A$ y $BJ_{2m} = -J_{2m}B$. Verificar las siguientes *reglas de transposición* para la forma alternante s:

$$s(\boldsymbol{x}, A^{\mathrm{t}}\boldsymbol{y}) = s(A\boldsymbol{x}, \boldsymbol{y}), \qquad s(\boldsymbol{x}, B^{\mathrm{t}}\boldsymbol{y}) = s(\boldsymbol{y}, B\boldsymbol{x}).$$

Ejercicio 4.21. Si $A \in M_m(\mathbb{R})$, demostrar que

$$\operatorname{Pf}\begin{bmatrix} 0 & A \\ -A^{\mathrm{t}} & 0 \end{bmatrix} = (-1)^{m(m-1)/2} \det A.$$

⟦ Indicación: expresar la matriz al lado izquierdo como un producto $R^{\mathrm{t}}JR$ para ciertas matrices apropiadas $R, J \in M_{2m}(\mathbb{R})$. ⟧

Ejercicio 4.22. Si $A = -A^{\mathrm{t}} \in M_{2m}(\mathbb{F})$ es una matriz antisimétrica tal que $a_{ij} = 0$ para $j > i + 1$, demostrar directamente de la definición de pfaffiano que

$$\operatorname{Pf} A = a_{12}\, a_{34}\, a_{56} \cdots a_{2m-1,2m}\,.$$

⟦ Indicación: calcular $\det A$ por expansiones en columnas y filas. ⟧

[13] Un subespacio $M \leqslant V$ es *isotrópico* si $s(\boldsymbol{x}, \boldsymbol{x}) = 0$ para $\boldsymbol{x} \in M$, o equivalentemente, si $M \subseteq M^\perp$. Un subespacio isotrópico maximal L se llama un **subespacio lagrangiano** de V.

Ejercicio 4.23. Hay una fórmula inductiva para calcular el pfaffiano de una matriz antisimétrica A por expansión *en filas y columnas* a la vez. Denótese por $A_{ij,ij}$ la submatriz $(n-2)\times(n-2)$ de A obtenida al borrar las filas i, j y también las columnas i, j de A; nótese que esta submatriz es antisimétrica. Con la notación $\underline{p_{ij} := \operatorname{Pf}(A_{ij,ij})}$, la regla de expansión es:

$$\underline{\operatorname{Pf} A = a_{12}p_{12} - a_{13}p_{13} + \cdots + (-1)^n a_{1n}p_{1n}} = \sum_{j=2}^{n} (-1)^j a_{1j}p_{1j}.$$

Más generalmente, $\operatorname{Pf} A = \sum_{j=1}^{n} (-1)^{i+j-1} a_{ij}p_{ij}$ para cualquier $i \in \{1, 2, \dots, n\}$.

(a) Usar esta fórmula para hallar la expresión explícita del pfaffiano de una matriz antisimétrica 6×6, en términos de sus entradas a_{ij} con $i < j$.

(b) Verificar, por inducción sobre m, que esta fórmula conduce a la expresión general (4.22) para el pfaffiano de una matriz antisimétrica $2m \times 2m$.

► En los ejercicios que siguen:

- ◇ V es un espacio vectorial real de dimensión par $n = 2m$;
- ◇ d es un producto escalar real sobre V;
- ◇ J es una estructura compleja ortogonal sobre (V, d);
- ◇ s es la forma simpléctica sobre V definida por $\underline{s(\boldsymbol{x}, \boldsymbol{y}) := d(J\boldsymbol{x}, \boldsymbol{y})}$; y
- ◇ $V_{\mathbb{C}}$ denota la complexificación de V.

Cada operador $\mathbb{R}$-lineal $T\colon V \to V$ se puede ampliar a un operador $\mathbb{C}$-lineal del mismo nombre $T\colon V_{\mathbb{C}} \to V_{\mathbb{C}}$, mediante la redefinición $T(\boldsymbol{x} + i\boldsymbol{y}) := T(\boldsymbol{x}) + iT(\boldsymbol{y})$ para $\boldsymbol{x}, \boldsymbol{y} \in V$.

Ejercicio 4.24. Si $R\colon V \to V$ es un *operador lineal ortogonal*: $\underline{d(R\boldsymbol{x}, R\boldsymbol{y}) = d(\boldsymbol{x}, \boldsymbol{y})}$ para todo $\boldsymbol{x}, \boldsymbol{y} \in V$, demostrar que R es invertible con R^{-1} también ortogonal, y que el operador lineal RJR^{-1} es otra estructura compleja ortogonal sobre V.

Ejercicio 4.25. Sean θ, ϕ dos ángulos cualesquiera. Demostrar que las siguientes matrices determinan estructuras complejas ortogonales sobre $V = \mathbb{R}^4$ (con su producto escalar usual):

$$J_{\theta\phi} = \begin{bmatrix} 0 & -\cos\theta & -\operatorname{sen}\theta\cos\phi & -\operatorname{sen}\theta\operatorname{sen}\phi \\ \cos\theta & 0 & -\operatorname{sen}\theta\operatorname{sen}\phi & \operatorname{sen}\theta\cos\phi \\ \operatorname{sen}\theta\cos\phi & \operatorname{sen}\theta\operatorname{sen}\phi & 0 & -\cos\theta \\ \operatorname{sen}\theta\operatorname{sen}\phi & -\operatorname{sen}\theta\cos\phi & \cos\theta & 0 \end{bmatrix}.$$

Ejercicio 4.26. Sobre $V = \mathbb{R}^4$ también, comprobar que estas matrices $J'_{\alpha\beta}$ determinan estructuras complejas ortogonales diferentes de las $J_{\theta\phi}$ del Ejercicio 4.25 anterior:[14]

$$J'_{\alpha\beta} = \begin{bmatrix} 0 & -\cos\alpha & -\operatorname{sen}\alpha\cos\beta & -\operatorname{sen}\alpha\operatorname{sen}\beta \\ \cos\alpha & 0 & \operatorname{sen}\alpha\operatorname{sen}\beta & -\operatorname{sen}\alpha\cos\beta \\ \operatorname{sen}\alpha\cos\beta & -\operatorname{sen}\alpha\operatorname{sen}\beta & 0 & \cos\alpha \\ \operatorname{sen}\alpha\operatorname{sen}\beta & \operatorname{sen}\alpha\cos\beta & -\cos\alpha & 0 \end{bmatrix}.$$

Ejercicio 4.27. Se puede considerar V como un espacio de Hilbert complejo V con el producto escalar (4.24) dado por $\langle \boldsymbol{x}, \boldsymbol{y} \rangle := d(\boldsymbol{x}, \boldsymbol{y}) + i\, s(\boldsymbol{x}, \boldsymbol{y})$.

Un operador $\mathbb{R}$-lineal $Q \colon V \to V$ se llama **$\mathbb{C}$-lineal** si $QJ = JQ$; en cambio, Q se llama **$\mathbb{C}$-semilineal** si $QJ = -JQ$.

(a) Sean $R := \frac{1}{2}(Q - JQJ)$ y $S := \frac{1}{2}(Q + JQJ)$. Fíjese que $Q = R + S$. Verificar que R es $\mathbb{C}$-lineal y que S es $\mathbb{C}$-semilineal.

(b) Si Q es un operador ortogonal sobre V, mostrar que R^{t} y S^{t} son las partes $\mathbb{C}$-lineal y $\mathbb{C}$-semilineal de Q^{-1}. Además, verificar las relaciones:

$$RR^{\mathrm{t}} + SS^{\mathrm{t}} = R^{\mathrm{t}}R + S^{\mathrm{t}}S = 1_V\,, \qquad RS^{\mathrm{t}} = -SR^{\mathrm{t}}, \quad R^{\mathrm{t}}S = -S^{\mathrm{t}}R.$$

⟦ Indicación: Estudiar la relación $QQ^{-1} = Q^{-1}Q = 1_V$. ⟧

Ejercicio 4.28. Demostrar el Escolio 4.52: el operador $P := \frac{1}{2}(1 - iJ)$ es un proyector ortogonal sobre el espacio de Hilbert $V_{\mathbb{C}}$ con su producto escalar $\langle\!\langle -, - \rangle\!\rangle$.

Ejercicio 4.29. Cada subespacio $W \leqslant V_{\mathbb{C}}$ tiene una "pareja" $\overline{W} := \{\, \boldsymbol{x} - i\boldsymbol{y} : \boldsymbol{x} + i\boldsymbol{y} \in W \,\}$, donde $\boldsymbol{x}, \boldsymbol{y} \in V$. Una **polarización** de $V_{\mathbb{C}}$ es un subespacio complejo $W \leqslant V_{\mathbb{C}}$ que es *isotrópico* para d [esto es, $d(\boldsymbol{w}, \boldsymbol{z}) = 0$ para todo $\boldsymbol{w}, \boldsymbol{z} \in W$] y cumple $W \cap \overline{W} = \{\boldsymbol{0}\}$ y $W \oplus \overline{W} = V_{\mathbb{C}}$.

(a) Mostrar que $W_0 := P(V) = \{\, \boldsymbol{x} - iJ\boldsymbol{x} : \boldsymbol{x} \in V \,\}$ es una polarización de $V_{\mathbb{C}}$.

(b) Demostrar que W_0 y $\overline{W_0}$ son los subespacios de autovectores para el operador ampliado $J \colon V_{\mathbb{C}} \to V_{\mathbb{C}}$ con los autovalores i, $-i$ respectivamente.

(c) Si $Q \colon V \to V$ es un operador lineal ortogonal y si W es una polarización de $V_{\mathbb{C}}$, demostrar que $Q(W) := \{\, Q\boldsymbol{x} + iQ\boldsymbol{y} : \boldsymbol{x} + i\boldsymbol{y} \in W \,\}$ es otra polarización de $V_{\mathbb{C}}$.

(d) Si W es una polarización de $V_{\mathbb{C}}$ y si $\boldsymbol{x}, \boldsymbol{y}_1, \boldsymbol{y}_2 \in V$ son vectores tales que $\boldsymbol{x} + i\boldsymbol{y}_1 \in W$, $\boldsymbol{x} + i\boldsymbol{y}_2 \in W$, mostrar que $\boldsymbol{y}_1 = \boldsymbol{y}_2$. Concluir que hay un operador $\mathbb{R}$-lineal $J_W \colon V \to V$ determinado por $J_W(\boldsymbol{x}) := -\boldsymbol{y}$ cuando $\boldsymbol{x} + i\boldsymbol{y} \in W$.

(e) Comprobar que J_W es una estructura compleja ortogonal sobre V.

[14]Estas $J_{\theta\phi}$ y $J'_{\alpha\beta}$ son *todas* las estructuras complejas ortogonales sobre $\mathbb{R}^4$. Geométricamente, forman dos copias disjuntas de la esfera $\mathbb{S}^2$; sobre ellas (θ, ϕ) y (α, β) denotan coordenadas esféricas.