MA–460: ALGEBRA LINEAL II

Joseph C. Várilly

Escuela de Matemática, Universidad de Costa Rica

II Ciclo Lectivo del 2007

Introducción

El álgebra lineal comprende el estudio de los espacios vectoriales y las aplicaciones lineales entre ellos. La estructura de un espacio vectorial finitodimensional es sencilla: todos los vectores son combinaciones de un número finito de vectores básicos. Dadas unas bases de dos espacios vectoriales, una aplicación lineal del primero al segundo se manifiesta con la matriz de sus coeficientes con respecto a estas dos bases. De este modo, se obtiene una estrecha relación entre las propiedades estructurales de las aplicaciones lineales y los algoritmos para manipular sus 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.

Inicialmente, 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 bien de una aplicación lineal), con el objetivo de transformar una matriz dada a una forma diagonal. Este meta no siempre puede realizarse; por ende, se examinará en detalle la estructura de una aplicación lineal cualquiera, para obtener las llamadas formas canónicas y normales de sus matrices.

Otro concepto fundamental es la de *ortogonalidad*. En presencia de un producto escalar (o producto interno) sobre un espacio vectorial real o complejo, las aplicaciones lineales se reparten en diversos clases: ortogonales o unitarios, simétricos o hermíticos, positivas. Estas 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 otras maneras. Las formas antisimétricas se caracterizan por su rango; las formas simétricas, por su rango y signatura. Una forma bilineal simétrica sobre un espacio vectorial sirve para construir una estructura multiplicativa llamada *álgebra de Clifford* que la caracteriza, en términos de matrices sobre escalares reales, complejos o cuaternionicos.

Estos apuntes van acompañados de diversos ejercicios, los cuales, además de ofrecer una práctica rutinaria acerca de los tópicos discutidos, sirven para amplificar y complementar esos temas. La evaluación del curso estará basado en estos ejercicios.

Programa de materias

1 Fundamentos del Algebra Lineal: 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 Algebras Exteriores y de Clifford

Producto tensorial de dos espacios vectoriales. Algebras tensorial, simétrica y exterior de un espacio vectorial. Integral de Berezin, pfaffianos y gaussianos. Algebra de Clifford de una forma cuadrática. Clasificación matricial de las álgebras de Clifford, el Octuple Sendero.

1 Fundamentos del Algebra Lineal

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 a éste. 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**,[1] 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: 0, el cero aditivo y 1, la unidad multiplicativa.

Tres cuerpos ya son bien conocidos: los *números racionales* $\mathbb{Q}$, *los números reales* $\mathbb{R}$ y los *números complejos* $\mathbb{C}$. En lo sucesivo, $\mathbb{Z}$ denotará los números enteros y $\mathbb{N} = \{0,1,2,3,\dots\}$ denotará los números enteros no negativos, a veces llamados "números naturales".[2] Obsérvese 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 por 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 $\mathbb{F}$ es un conjunto V, cuyos elementos se llaman **vectores**, en donde se definen 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 identidades

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

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 $\mathbb{F}^n$. Las sumas y los múltiplos escalares de n-tuplas se definen "entrada por entrada".

Ejemplo 1.2. La totalidad de *polinomios* con coeficientes en $\mathbb{F}$ se denota por $\mathbb{F}[t]$, en donde la letra t es una *indeterminada*. 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 la multiplicación de polinomios por escalares, $\mathbb{F}[t]$ es un espacio vectorial sobre $\mathbb{F}$.

[1] 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.

[2] Conviene incluir 0 como número natural, aunque este costumbre no tiene aceptación 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.*

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 vectorial sobre $\mathbb{F}$, con 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 $\boldsymbol{x} + \boldsymbol{y} \in W$, $c\boldsymbol{x} \in W$.

Úsase la notación $\underline{W \leq V}$ para significar que W es un subespacio de V.

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 $\mathbb{F}[t]$.

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

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

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 $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in X$.

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

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

para denotar este subespacio.

Si X es una parte, posiblemente infinita, de V, el subespacio $\operatorname{lin}\langle X \rangle$ 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 para V si y sólo 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

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

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 $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$, donde

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

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

El conjunto $\{1, t, t^2, \dots, t^n, \dots\}$ es una base para el espacio vectorial $\mathbb{F}[t]$ de todos los polinomios sobre $\mathbb{F}$.

Definición 1.8. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ una base para V y sean $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m \in V$. Si $m > n$, entonces $\boldsymbol{y}_1, \dots, \boldsymbol{y}_m$ son linealmente dependientes. En consecuencia, si $\{\boldsymbol{y}_1, \dots, \boldsymbol{y}_m\}$ es otra base para V, entonces $m = n$. Si V posee una base finita, el número n de sus elementos es la **dimensión** del espacio vectorial V; en símbolos, $n = \dim V$. (Se dice que V es *finitodimensional* en este caso.) En particular, es $\dim \mathbb{F}^n = n$.

Para construir una base de un espacio vectorial dado, es útil saber que siempre se puede *completar una base parcial*, es decir, es posible prolongar una base para un subespacio en una base para el espacio de marras, en vista de la siguiente Proposición.

Proposición 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.* $\boxminus$

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 de V y W, dotado de las siguientes operaciones de suma y multiplicación escalar:

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

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

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

En general, $\dim(U + W) \le \dim U + \dim W$, 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 $U \oplus W = V$. En general, cada subespacio de V posee muchos 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** $\overline{V/W}$ es el conjunto de los traslados $\boldsymbol{x} + W := \{\boldsymbol{x} + \boldsymbol{w} : \boldsymbol{w} \in W\}$ 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

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

La totalidad de aplicaciones lineales $T\colon V \to W$ se denota por $\mathcal{L}(V,W)$. Este es también un espacio vectorial sobre $\mathbb{F}$, bajo las operaciones:

$$\begin{aligned} T+S &: \boldsymbol{x}\mapsto T(\boldsymbol{x})+S(\boldsymbol{x}),\\ cT &: \boldsymbol{x}\mapsto c\,T(\boldsymbol{x}). \end{aligned}$$

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*[3] es la aplicación lineal $ST\in\mathcal{L}(V,Z)$ dado por $ST : \boldsymbol{x}\mapsto S(T(\boldsymbol{x}))$.

Para manejar las aplicaciones lineales de modo explícito, se usa la propiedad clave de que *una aplicación lineal es determinada por sus valores sobre los elementos de una base* de su dominio.

Proposición 1.14. *Sean V y W dos espacios vectoriales sobre $\mathbb{F}$, 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 un cuerpo $\mathbb{F}$. Una **forma lineal** sobre V es una aplicación lineal $f\colon V\to\mathbb{F}$. El **espacio dual** de V es el espacio vectorial $\underline{V^*}$ de todas las formas lineales, vale decir, $V^* := \underline{\mathcal{L}(V,\mathbb{F})}$. 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** $\{f_1,\dots,f_n\}$ 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 ahora un convenio de notación.[4] Si $R(x)$ es una relación lógica cualquiera que involucra un parámetro x, la notación $[\![R(x)]\!]$ denota la siguiente función booleana:

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

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

[4]Esta notación fue introducido en 1962 por Kenneth E. Iverson. 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.

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 $\mathbf{1}_A(x) := [\![x \in A]\!]$, 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 $\operatorname{signo}(t) := [\![t > 0]\!] - [\![t < 0]\!]$.

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

Si V es finitodimensional, es $\dim(V^*)^* = \dim V^* = \dim V$. 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^{**}; por conteo de dimensiones, este subespacio es todo V^{**}. En otras palabras, 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. (Estas propiedades de reciprocidad entre V y V^* justifican el empleo de la palabra "dual" para V^*.)

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

$$T^t(g) := g \circ T, \quad \text{para todo} \quad g \in W^*.$$

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

Si $T \in \mathcal{L}(V, W)$ y $S \in \mathcal{L}(W, Z)$, resulta que $(ST)^t = T^t S^t$, porque

$$(T^t S^t)(h) = T^t(S^t(h)) = T^t(h \circ S) = (h \circ S) \circ T = h \circ (ST) = (ST)^t(h), \quad \text{para} \quad h \in Z^*.$$

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

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

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

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

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

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

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

(b) T es sobreyectivo si y sólo si $T(V) = W$, si y sólo si $r(T) = \dim W$. $\boxminus$

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

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

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

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

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

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

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

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

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

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

(a) $r(T) = r(T^t)$,

(b) $r(T) + n(T) = \dim V$. ⊟

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 $A = [a_{ij}]$, donde se sobreentiende que $i = 1, 2, \ldots, m$ y $j = 1, 2, \ldots, n$.

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

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, $A^t = [a_{ji}]$. Una matriz cuadrada $A \in M_n(\mathbb{F})$ se llama **simétrica** si $A^t = A$.

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 $i \neq j$.

Las *columnas* de una matriz $A \in \mathbb{F}^{m\times n}$ pueden considerarse como *vectores en* $\mathbb{F}^m$. Una columna típica es

$$\boldsymbol{a}_j = \begin{bmatrix} a_{1j} \\ a_{2j} \\ \vdots \\ a_{mj} \end{bmatrix}.$$

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

Además, las filas de A pueden considerarse como vectores en el espacio dual $(\mathbb{F}^n)^* \simeq \mathbb{F}^n$. Una fila típica[5] es

$$\boldsymbol{a}^i = \begin{bmatrix} a_{i1} & a_{i2} & \dots & a_{in} \end{bmatrix},$$

de modo que $A = [\boldsymbol{a}^1, \boldsymbol{a}^2, \dots, \boldsymbol{a}^m]^t$. (Para distinguir los vectores de fila de los vectores de columna, conviene usar exponentes o "superíndices" para etiquetar aquellos.)

La fila $\boldsymbol{a}^i$ corresponde a la forma lineal

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

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

Definición 1.25. Si $A \in \mathbb{F}^{m\times n}$ y $B \in \mathbb{F}^{n\times r}$, el **producto de matrices** $\underline{AB}$ es la matriz $C = AB \in \mathbb{F}^{m\times r}$ cuya entrada (i, j) es el producto punto de la fila i de A con la columna j de B; es decir,

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

El producto matricial obedece las leyes algebraicas $A(BC) = (AB)C$; $(A + D)B = AB + DB$; $A(B + E) = AB + AE$; $(AB)^t = B^tA^t$. Sin embargo, este producto no es conmutativo, porque $AB \neq BA$ en general.

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 las primeras tres igualdades del párrafo anterior. En otras palabras, $M_n(\mathbb{F})$ es a la vez un *espacio vectorial* sobre $\mathbb{F}$ y un *anillo* (no conmutativo): se dice[6] que $M_n(\mathbb{F})$ es un **álgebra** sobre $\mathbb{F}$.

Esta álgebra tiene un elemento unidad, la **matriz identidad** $I_n \in M_n(\mathbb{F})$, que cumple $AI_n = I_nA = A$ para todo $A \in M_n(\mathbb{F})$. Concretamente, $I_n = [\delta_{ij}]$, donde $\delta_{ij} = [\![i = j]\!]$ es la "delta de Kronecker". Esta es una matriz diagonal:

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

En un contexto en donde el tamaño n es fijo, se suele abreviar $I := I_n$.

Definición 1.26. Un elemento $A \in M_n(\mathbb{F})$ es una **matriz inversible** o *matriz no singular* si hay otra matriz $C \in M_n(\mathbb{F})$ tal que $AC = CA = I$. Si A es inversible, su **matriz inversa** C es única y se denota $C =: A^{-1}$. Fíjese que $(AB)^{-1} = B^{-1}A^{-1}$ cuando A, B son matrices inversibles.

[5]Algunos autores franceses usan la notación $\mathbb{F}_n$ en vez de $(\mathbb{F}^n)^*$ para denotar el espacio dual de $\mathbb{F}^n$.

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

Definición 1.27. Sea $T\colon V \to W$ 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[7] es la matriz $A = [a_{ij}] \in \mathbb{F}^{m\times n}$ dada por

$$T(\boldsymbol{x}_j) =: \sum_{i=1}^{m} a_{ij}\boldsymbol{y}_i. \tag{1.6}$$

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

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

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

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

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

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

En otras palabras, $[T(\boldsymbol{x})]^{\mathcal{C}} = A\,[\boldsymbol{x}]^{\mathcal{B}}$, o bien

$$[T(\boldsymbol{x})]^{\mathcal{C}} = [T]^{\mathcal{C}}_{\mathcal{B}}\,[\boldsymbol{x}]^{\mathcal{B}}. \tag{1.7}$$

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, la Proposición 1.14 afirma que $T \mapsto A$ es inyectiva, y la aplicación inversa es $A \mapsto T_A$, donde

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

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,

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

[7] Fíjese bien en la forma "enrevesada" de combinar los índices al lado derecho de esta ecuación.

(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 $T^t \in \mathcal{L}(W^*,V^*)$ es la matriz transpuesta A^t:
$$[T]^{\mathcal{C}}_{\mathcal{B}} = A \implies [T^t]^{\mathcal{B}^*}_{\mathcal{C}^*} = A^t.$$

Proposición 1.28. *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 $T_A(\boldsymbol{x}) = A\boldsymbol{x} \in \mathbb{F}^m$. Escríbase $\boldsymbol{x} = \sum_{j=1}^n x_j \boldsymbol{e}_j$, al desarrollar el vector $\boldsymbol{x}$ en términos de la base estándar $\{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$ de $\mathbb{F}^n$. Es evidente que $A\boldsymbol{e}_j = \boldsymbol{a}_j$ para $j = 1,\dots,n$. Por la linealidad de T_A, se ve que
$$A\boldsymbol{x} = T_A(\boldsymbol{x}) = \sum_{j=1}^n x_j T_A(\boldsymbol{e}_j) = x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n. \tag{1.8}$$
De ahí es evidente que cada vector $A\boldsymbol{x}$ es una combinación lineal de las columnas $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n$, e inversamente, que cada combinación lineal de estas columnas es de la forma $A\boldsymbol{x}$ para algún $\boldsymbol{x} \in \mathbb{F}^n$. En otras palabras, es $T_A(\mathbb{F}^n) = \operatorname{lin}\langle \boldsymbol{a}_1,\dots,\boldsymbol{a}_n\rangle \leq \mathbb{F}^m$. □

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

Definición 1.29. 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, $r(A) := r(T_A)$. Del mismo modo, la **nulidad** de A se define como la nulidad de T_A, esto es, $n(A) := n(T_A)$.

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

► 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
$$[T]^{\mathcal{C}'}_{\mathcal{B}'} = [I]^{\mathcal{C}'}_{\mathcal{C}}\,[T]^{\mathcal{C}}_{\mathcal{B}}\,[I]^{\mathcal{B}}_{\mathcal{B}'}, \quad \text{o bien} \quad B = QAP,$$
donde $P = [I]^{\mathcal{B}}_{\mathcal{B}'}$ y $Q = [I]^{\mathcal{C}'}_{\mathcal{C}}$ son las *matrices de cambio de base* en V, W respectivamente. Concretamente,
$$\boldsymbol{x}'_s =: \sum_{j=1}^n p_{js}\boldsymbol{x}_j, \qquad \boldsymbol{y}_i =: \sum_{r=1}^m q_{ri}\boldsymbol{y}'_r. \tag{1.9}$$
Obsérvese que P, por ser la matriz de una aplicación identidad I respecto de ciertas bases, es una matriz cuadrada inversible. La matriz Q es inversible por la misma razón.

Definición 1.30. Dos **matrices** $A, B \in \mathbb{F}^{m\times n}$ se dicen **equivalentes** si hay un par de matrices inversibles $P \in M_n(\mathbb{F})$ y $Q \in M_m(\mathbb{F})$ tales que $\underline{B = QAP}$.

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 (1.9), entonces B es la matriz de T_A con respecto a las nuevas bases. En particular, es $\underline{r(A) = r(T) = r(B)}$.

Los cambios de base de mayor interés ocurren cuando $W = V$ y se toman $\mathcal{C} = \mathcal{B}$ 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 (1.9), o bien por la reciprocidad entre $P = [I]^{\mathcal{B}}_{\mathcal{B}'}$ y $Q = [I]^{\mathcal{B}'}_{\mathcal{B}}$, se ve que $Q = P^{-1}$ en $M_n(\mathbb{F})$.

Definición 1.31. Dos **matrices cuadradas** $A, B \in M_n(\mathbb{F})$ se dicen **semejantes** si hay una matriz inversible $P \in M_n(\mathbb{F})$ tal que $\underline{B = P^{-1}AP}$.

Si A y B representan una aplicación lineal $T \in \mathcal{L}(V, V)$ respecto de un par de bases para 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 $\mathcal{E} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ en $\mathbb{F}^n$ a la base $\mathcal{P} = \{\boldsymbol{p}_1, \dots, \boldsymbol{p}_n\}$ mediante (1.9), entonces $B = [T_A]^{\mathcal{P}}_{\mathcal{P}}$ es la matriz de T_A con respecto de la nueva base $\mathcal{P}$.

1.4 Ecuaciones lineales y eliminación gaussiana

Un sistema de ecuaciones lineales tiene la siguiente forma:

$$\begin{aligned} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n &= b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n &= b_2 \\ \cdots\cdots\cdots\cdots &= \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n &= b_m \end{aligned} \tag{1.10}$$

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 (por un factor $-a_{i1}/a_{11}$) de la primera ecuación. 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 seguir adelante 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.[8]

[8]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 planeta enano Pallas. Por supuesto, el método es mucho más antiguo. Aparece en el manuscrito chino *Jiuzhang Suanshu* (Nueve Capítulos sobre el Arte de Calcular), de autoría desconocido, de la época de la dinastía Han (∼150 a.C.). Su octavo capítulo, *Fang cheng* (arreglo cuadrilongo), se dedica al método de eliminación.

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

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

con $a_{11} \neq 0$, $a_{22}^{(2)} \neq 0$, …, $a_{nn}^{(n)} \neq 0$. 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* $a_{kk}^{(k)}$ se anulara, habrá lugar para otras posibilidades, como la inexistencia de soluciones o la existencia de más de una solución.

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 (1.10) puede escribirse como una sola ecuación matricial:

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

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 $\mathbb{F}^m$:

$$x_1\boldsymbol{a}_1 + x_2\boldsymbol{a}_2 + \cdots + x_n\boldsymbol{a}_n = \boldsymbol{b}, \tag{1.11}$$

en donde los vectores $\boldsymbol{a}_j \in \mathbb{F}^m$ son las columnas de la matriz A. Esta ecuación revela que un sistema de ecuaciones lineales puede expresar 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 (1.11) se ve que hay al menos una solución $\boldsymbol{x}$ al sistema si y sólo 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*.

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

Proposición 1.32. *Las operaciones de fila no cambian el rango de una matriz.* $\boxminus$

Proposición 1.33. *Cualquier operación de fila elemental sobre una matriz $B \in \mathbb{F}^{m\times n}$ se efectúa por premultiplicación*[9] *$B \mapsto AB$ de esa matriz por una matriz cuadrada $A \in \mathbb{F}^{m\times m}$.*

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

$$M_i(c) = \begin{bmatrix} 1 & 0 & \dots & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & & \vdots \\ 0 & 0 & \dots & c & \dots & 0 \\ \vdots & \vdots & & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 0 & \dots & 1 \end{bmatrix}. \tag{1.12a}$$

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

$$R_{ik}(c) = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & -c & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix} \quad \text{si} \quad k > i. \tag{1.12b}$$

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 con $B \mapsto P_{ik}B$, donde

$$P_{ik} = \begin{bmatrix} 1 & \dots & 0 & \dots & 0 & \dots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \dots & 0 & \dots & 1 & \dots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \dots & 1 & \dots & 0 & \dots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & \dots & 0 & \dots & 1 \end{bmatrix}. \tag{1.12c}$$

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

[9] Como el producto de matrices no es conmutativa, hay que distinguir entre los procesos de *premultiplicación* $B \mapsto AB$ y *posmultiplicación* $C \mapsto CA$ por una matriz A.

Todas estas matrices son inversibles: 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 matrices de tipo (1.12).

Por ejemplo, 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,

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

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

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

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,[10] es una **factorización** de la matriz A como $A = LV$, donde L es una matriz triangular inferior y V es una matriz triangular superior:

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

Obsérvese también que las entradas subdiagonales de la matriz 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, se escribe la matriz triangular inferior L sin necesidad de 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 $\underline{A = LV}$.

[10]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.

La matriz triangular L es "unipotente", es decir, todas sus entradas diagonales son iguales a 1. Además, las entradas diagonales $a_{kk}^{(k)}$ de V no son ceros. Por tanto, V puede factorizarse a su vez en un producto $V = DU$, 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 sólo si) la matriz original A es inversible, 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 inversibles, 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 consiste 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;[11] 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,[12] de todos los P_{kj} que ocurren durante el proceso. Se puede mostrar 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.[13]

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 sólo si) esto ocurre, el rango de la matriz A es menor que n y por ende A no es inversible.

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

Proposición 1.34. *Si $A \in \mathbb{F}^{m\times n}$, el conjunto de soluciones del sistema* homogéneo *de ecuaciones lineales $A\boldsymbol{x} = \boldsymbol{0}$ es el subespacio* $\ker T_A \leq \mathbb{F}^n$, *de dimensión $n(A)$. Hay una solución única ($\boldsymbol{x} = \boldsymbol{0}$, por supuesto) si y sólo si $n(A) = 0$, si y sólo si $\underline{r(A) = n}$.*

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

[12] 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 $P^{-1} = P_{23}P_{35}P_{68}$.

[13] Para un análisis detallado del método de eliminación gaussiana, véase: Gilbert W. Stewart, *Introduction to Matrix Computations*, Academic Press, New York, 1973.

El sistema inhomogéneo *de ecuaciones lineales* $Ax = b$*, con* $b \neq 0$*, posee una solución si y sólo si* $b \in T_A(\mathbb{F}^n)$*, si y sólo si el rango* $r([A \mid b])$ *de la matriz aumentada es igual a* $r(A)$*. En ese caso, su conjunto de soluciones* $\{ x : Ax = b \}$ *es el* subespacio afín $x_0 + \ker T_A$*, donde* x_0 *es alguna solución particular. Luego* $Ax = b$ *posee solución única si y sólo si* $r([A \mid b]) = r(A) = n$*.*

Definición 1.35. Se dice 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, $a_{j_1} = e_1$, $a_{j_2} = e_2, \dots, a_{j_k} = e_k$ para algún k;

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

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

Las columnas $a_{j_1}, \dots, a_{j_k}$ se llaman *columnas básicas*[14] 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 última fila consta de ceros. De hecho, es una consecuencia de la Definición 1.35 que las últimas $(m - k)$ filas son ceros, si $k < m$.

Proposición 1.36. *Cualquier matriz* $A \in \mathbb{F}^{m\times n}$ *puede transformarse en una única forma escalonada mediante operaciones de fila.* ⊟

Proposición 1.37. *Si* $A \in \mathbb{F}^{m\times n}$ *es una matriz en forma escalonada con* k *columnas básicas, entonces* $T_A(\mathbb{F}^n) = \operatorname{lin}\langle e_1, \dots, e_k \rangle$ *y por ende* $r(A) = k$*.* ⊟

Proposición 1.38. *Dos matrices* $A, B \in \mathbb{F}^{m\times n}$ *son equivalentes si y sólo si se puede transformar* A *en* B *por operaciones de fila* y de columna*, si y sólo si* $r(A) = r(B)$*.* ⊟

Proposición 1.39. *Si* $A \in \mathbb{F}^{m\times n}$ *es una matriz de rango* k*, hay matrices inversibles* $Q \in M_m(\mathbb{F})$*,* $P \in M_n(\mathbb{F})$ *tales que*

$$QAP = \begin{bmatrix} I_k & O \\ O & O \end{bmatrix},$$

en donde cada O *es un bloque rectangular de ceros.*

Demostración. La matriz A puede reducirse a su forma escalonada B, que contiene k columnas básicas, mediante operaciones de fila solamente. La Proposición 1.33 muestra que hay una matriz inversible Q tal que $QA = B$. En seguida, se puede *intercambiar columnas* para

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

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 puede hacerse como sigue: (i) transponer $B \mapsto B^t$; (ii) hacer la correspondiente operación de fila sobre B^t por una premultiplicación $B^t \mapsto MB^t$ donde M es una matriz de tipo (1.12); (iii) transponer de nuevo, $MB^t \mapsto BM^t$. En resumen, una operación de columna elemental se efectúa al *posmultiplicar* B por cierta matriz inversible. Por tanto, una sucesión de operaciones de columna transforma B en BP', donde P' es una matriz inversible. El resultado de mover las columnas básicas de B a la izquierda es entonces

$$QAP' = \begin{bmatrix} I_k & F \\ O & O \end{bmatrix}, \qquad (QAP')^t = \begin{bmatrix} I_k & O \\ F^t & O \end{bmatrix},$$

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

$$QAP = (QAP')(P'')^t = (P''(QAP')^t)^t = \begin{bmatrix} I_k & O \\ O & O \end{bmatrix}. \qquad \square$$

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 por

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

Este es un escalar que *determina* si la matriz es inversible o no: la matriz es inversible si su determinante no es cero. 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 inversible en $M_2(\mathbb{F})$ si y sólo si $\det A \neq 0$, en cuyo caso

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

El determinante de una matriz 3×3 puede definirse "por expansión según la primera fila":

$$\begin{aligned} \begin{vmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{vmatrix} &:= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{12} \begin{vmatrix} a_{21} & a_{23} \\ a_{31} & a_{33} \end{vmatrix} + a_{13} \begin{vmatrix} a_{21} & a_{22} \\ a_{31} & a_{32} \end{vmatrix} \\ &= a_{11}a_{22}a_{23} + a_{12}a_{23}a_{31} + a_{13}a_{21}a_{32} - a_{11}a_{23}a_{32} - a_{12}a_{21}a_{33} - a_{13}a_{22}a_{31}. \end{aligned} \tag{1.13a}$$

Para una matriz cuadrada $A =\in M_n(\mathbb{F})$, se puede definir el determinante por inducción, como sigue. Si $n = 1$, se define $\det[a_{11}] := a_{11} \in \mathbb{F}$. Supóngase que se dispone de una definición del determinante para matrices en $M_{n-1}(\mathbb{F})$. Si $A \in M_n(\mathbb{F})$, sea A_{ij} —con A mayúscula— la **submatriz** de A obtenida al borrar la fila i y la columna j de A. Cada A_{ij} es una matriz $(n-1)\times(n-1)$. Escríbase

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

Este escalar m_{ij} se llama el **menor** de la matriz A correspondiente a la entrada a_{ij}. Obsérvese que la ecuación (1.13a) puede abreviarse con esta notación:

$$\det A = a_{11}m_{11} - a_{12}m_{12} + a_{13}m_{13}. \tag{1.13b}$$

Cuando $n = 2$, es $m_{11} = a_{22}$ y $m_{12} = a_{21}$, así que $\det A = a_{11}a_{22} - a_{12}a_{21} = a_{11}m_{11} - a_{12}m_{12}$.

Definición 1.40. Si $A \in M_n(\mathbb{F})$; se define $\det A \in \mathbb{F}$ *por expansión según la primera fila*:

$$\det A := a_{11}m_{11} - a_{12}m_{12} + \cdots + (-1)^{1+n}a_{1n}m_{1n} = \sum_{j=1}^{n}(-1)^{1+j}a_{1j}m_{1j}, \tag{1.14a}$$

o bien *por expansión en según fila i*:

$$\det A := (-1)^{i+1}a_{i1}m_{i1} + (-1)^{i+2}a_{i2}m_{i2} + \cdots + (-1)^{i+n}a_{in}m_{in} = \sum_{j=1}^{n}(-1)^{i+j}a_{ij}m_{ij}, \tag{1.14b}$$

o bien *por expansión según la columna j*:

$$\det A := (-1)^{1+j}a_{1j}m_{1j} + (-1)^{2+j}a_{2j}m_{2j} + \cdots + (-1)^{n+j}a_{nj}m_{nj} = \sum_{i=1}^{n}(-1)^{i+j}a_{ij}m_{ij}. \tag{1.14c}$$

La siguiente Proposición muestra que todas estas definiciones son compatibles, pues conducen al mismo resultado.

Proposición 1.41. *Si $A \in M_n(\mathbb{F})$, entonces* $\det A$*, definido por cualquiera de las fórmulas en* (1.14)*, es igual a*

$$\det A = \sum_{\sigma\in S_n}(-1)^{\sigma}a_{1j_1}a_{2j_2}\dots a_{nj_n}, \tag{1.15}$$

donde la sumatoria[15] *recorre todas las $n!$ permutaciones $\sigma = (j_1,\dots,j_n)$ de $(1,2,\dots,n)$ y el signo $\underline{(-1)^{\sigma}} = \pm 1$ es $+1$ ó -1 según la permutación σ sea* par *o* impar.[16]

[15]La fórmula (1.15) 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 (1.15). 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 tiende a clarificar esta descripción verbal.

[16]Una *permutación* σ de $(1,2,\dots,n)$ 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. Según la fórmula (1.14a), es $\det A := \sum_{j_1=1}^{n}(-1)^{1+j_1}a_{1j_1}m_{1j_1}$, 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 (1.15), en donde cada producto contiene un factor tomado de cada fila y de columnas distintas. Hay n términos en (1.14a), $(n-1)$ 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 (1.15). 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 (1.14a) es efectivamente $+1$ si y sólo 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 (1.14b) ó (1.14c). (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 (1.15) por un múltiplo global de (± 1), que sería independiente de la matriz A. Un cálculo explícito muestra que todas las recetas en (1.14) dan $+1$ como el valor de $\det I_n$, así que el desarrollo (1.15) es correcto. □

Proposición 1.42. *Si* $A, B \in M_n(\mathbb{F})$, *entonces* $\det(AB) = \det A \det B$.

Demostración. Sea $C := AB$. Por (1.15), $\det C$ 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

$$\det(AB) = \sum (-1)^{\sigma} a_{1i_1}a_{2i_2}\dots a_{ni_n}\, b_{i_1j_1}b_{i_2j_2}\dots b_{i_nj_n}. \tag{1.16}$$

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 $\sum_\sigma (-1)^\sigma b_{i_1j_1}b_{i_2j_2}\dots b_{i_nj_n}$ se anula por cancelación de términos, así que aparecen en (1.16) 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,

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

Proposición 1.43. *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. En vista de la Proposiciones 1.33 y 1.42, es suficiente 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 (1.12).

Al expandir $\det A$ en cualquier fila i de A que tenga 1 en la diagonal y 0 en las demás entradas, la fórmula (1.14b) muestra que $\det A = (-1)^{i+i} m_{ii} = m_{ii}$. En consecuencia, se puede eliminar la fila i y también la columna i de A sin cambiar el determinante.

En las tres matrices definidas en (1.12), se puede eliminar así todas las filas y columnas excepto aquellas numeradas i y k —para $M_i(c)$, se puede tomar cualquier k con $k \neq i$. La expansión según filas (1.14b) 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 1.44. *Si $A \in M_n(\mathbb{F})$, entonces* $\det A^t = \det A$.

Demostración. La fórmula (1.15), aplicada a la transpuesta de A, da

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

Sea $\pi = (p_1, \dots, p_n) := \sigma^{-1}$ 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,

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

Proposición 1.45. *Si $A \in M_n(\mathbb{F})$ es una matriz triangular, entonces* $\det A = a_{11}a_{22}\dots a_{nn}$ *es el producto de los elementos diagonales de A. En particular,* $\det I_n = 1$.

Demostración. Supóngase que A es una matriz triangular inferior. La expansión (1.14a) 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

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

El mismo argumento es aplicable si A es triangular superior (o bien se puede apelar a la Proposición anterior). $\square$

Proposición 1.46. *Si $A \in M_n(\mathbb{F})$ y si dos filas de A son iguales; o bien si dos columnas de A son iguales; o bien si una fila o una columna de A es nula, entonces* $\det A = 0$. ⊟

Proposición 1.47. $\det A = 0$ *si y sólo si A es singular (es decir, no inversible).*

Demostración. Si A es inversible, tenemos $1 = \det I_n = (\det A)(\det A^{-1})$, 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 del segundo y tercer tipos. El resultado de este proceso es una matriz de la forma

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

donde U' es triangular superior y sus elementos diagonales son los pivotes $a_{11}, a_{22}^{(2)}, \dots, a_{kk}^{(k)}$; O es un bloque rectangular $(n-k)\times k$ de ceros; y X' es una matrix $k\times(n-k)$. La Proposición 1.43 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 inversible, el algoritmo de eliminación gaussiana se detiene en algún paso número k, con $k \leq n$, porque $a_{kk}^{(k)} = 0$. Entonces $\det A' = 0$ y por ende $\det A = 0$.

Por otro lado, si A es inversible, 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 inversible:

$$\det A = (-1)^r a_{11}a_{22}^{(2)} \dots a_{nn}^{(n)} \neq 0, \tag{1.17}$$

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, a veces vale la pena considerar los determinantes de sus *submatrices cuadradas*.

Proposición 1.48. *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. Sea $k = r(A)$. Entonces hay k columnas linealmente independientes en A (aquéllas que se convierten en columnas básicas al aplicar operaciones de fila para reducir A a su forma escalonada). Sea C la submatriz $m\times k$ de A que se obtiene al borrar las demás 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 demás filas de C. Entonces $B \in M_k(\mathbb{C})$ con $r(B) = k$, así que B con $\det B \neq 0$.

Si $k < \min\{m,n\}$, sea D una 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 una relación de dependencia lineal también (¿por qué?) así que D es singular y $\det D = 0$. □

► 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**.[17] En la práctica, es un método ineficiente para sistemas con más de tres variables; pero tiene importancia teórica. Por ejemplo, muestra que un sistema de ecuaciones con coeficientes enteros tiene soluciones racionales.

Definición 1.49. 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 remover la fila j y la columna i de A. Fíjese que $m_{ij} = \det((A^t)_{ij})$ es también un menor de la matriz transpuesta A^t.

La matriz $\operatorname{adj} A \in M_n(\mathbb{F})$, cuya entrada (i,j) es el cofactor de a_{ij}, es la **matriz adjugada**[18] de A. Para obtenerla, 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 el "tablero de ajedrez"; (iii) tomar la matriz transpuesta de ésta.

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

Proposición 1.50. *Si $A \in M_n(\mathbb{F})$, entonces*

$$A(\operatorname{adj} A) = (\operatorname{adj} A)A = (\det A)\, I_n. \tag{1.18}$$

Demostración. Obsérvese primero que la fórmula (1.14b) corresponde al producto de la fila i de A por la *columna* i de $(\operatorname{adj} A)$. Por otro lado, la fórmula (1.14c) representa el producto de la *fila* j de $(\operatorname{adj} A)$ por la columna j de A. En conjunto, estas dos fórmulas 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

$$(-1)^{i+1} a_{k1} m_{i1} + (-1)^{i+2} a_{k2} m_{i2} + \cdots + (-1)^{i+n} a_{kn} m_{in} = \sum_{j=1}^{n} (-1)^{i+j} a_{kj} m_{ij}. \tag{1.19}$$

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 la Proposición 1.46. Luego la expresión (1.19) 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 (1.18). □

[17] 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, en 1683.

[18] La matriz $\operatorname{adj} A$ a veces se llama la "matriz adjunta" de A. Sin embargo, cuando $\mathbb{F} = \mathbb{C}$, conviene reservar ese término para el conjugado hermítico A^*, que se verá en adelante. Algunos autores lo llaman el "adjunto clásico" de A; véase, por ejemplo: Kenneth Hoffman y Ray Kunze, *Algebra Lineal*, Prentice-Hall Internacional, Madrid, 1972. Aquí se adopta el convenio de usar la palabra *adjugada*, una mezcla inelegante de "adjunta" y "conjugada", con las disculpas apropiadas.

La fórmula (1.18) proporciona una fórmula para la matriz inversa de una matriz no singular A. Si $\det A \neq 0$, entonces

$$A^{-1} = \frac{1}{\det A}\,\operatorname{adj} A.$$

En el caso 2×2, esta relación es

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

Proposición 1.51 (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 sólo si $\det A \neq 0$, en cuyo caso*

$$x_j = \frac{\det B_j}{\det A}, \quad \textit{para} \quad j = 1,\dots,n. \tag{1.20}$$

Demostración. Ya se sabe que $A\boldsymbol{x} = \boldsymbol{b}$ tiene solución única sólo si $A\boldsymbol{x} = \boldsymbol{0}$ tiene solución única, sólo si $\ker T_A = \{\boldsymbol{0}\}$, sólo si $n(A) = 0$, sólo si $r(A) = n$, sólo si A es inversible, sólo 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, las coordenadas j de estos dos vectores de columna son

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

al usar la expansión (1.14c) en la columna j para evaluar $\det B_j$. Al dividir esta relación por $\det A$, se obtiene (1.20). $\square$

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) Demostrar que los tres polinomios $\frac{1}{2}t(t-1)$, $1-t^2$, $\frac{1}{2}t(t+1)$ son linealmente independientes en $\mathbb{Q}[t]$. 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: El polinomio $a_0 + a_1 t + a_2 t^2 + a_3 t^3$, si no es el polinomio nulo, no puede tener más de tres raíces distintas. ⟧

Ejercicio 1.3. Sea $V = C[-\pi,\pi]$ el espacio vectorial de funciones continuas $f\colon [-\pi,\pi] \to \mathbb{R}$. Para cada $n \in \mathbb{N}$, defínase $f_n \in V$ por $f_n(x) := \cos nx$. Demostrar que el conjunto $\{ f_n : n \in \mathbb{N} \}$ es linealmente independiente en V. [[Indicación: Evaluar la integral $\int_{-\pi}^{\pi} \cos mx \cos nx \, dx$.]]

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

Ejercicio 1.5. Un **monomio** en k variables $t_1,\dots,t_k$ es un producto $c\,t_1^{m_1} t_2^{m_2} \dots t_k^{m_k}$; su *grado* es la suma $m_1 + \cdots + m_k$ de los exponentes. Un **polinomio homogéneo** de grado m es una suma finita de monomios de grado m; ellos forman un espacio vectorial $P_m^{(k)}$. ¿Cuál es la dimensión de $P_m^{(k)}$?

Ejercicio 1.6. Sea V un espacio vectorial finitodimensional sobre $\mathbb{F}$ y sean M, N dos subespacios de V. Demostrar que su intersección $M \cap N$ y su suma $M+N$ son también subespacios de V. Verificar la fórmula

$$\dim(M+N) = \dim M + \dim N - \dim(M \cap N).$$

[[Indicación: Elíjase una base para $M \cap N$ y completarla de dos maneras para formar bases de M y de N. Verificar que la unión de estas dos bases es una base para $M+N$.]]

Ejercicio 1.7. Si L, M, N son tres subespacios de un espacio vectorial finitodimensional V, comprobar que $\dim(L+M+N)$ es igual a

$$\dim L + \dim M + \dim N - \dim(L \cap M) - \dim(L \cap N) - \dim(M \cap N) + \dim(L \cap M \cap N).$$

Ejercicio 1.8. (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

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

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 de $\mathbb{F}_n[t]^*$ consta de las evaluaciones $f_j\colon \mathbb{F}_n[t] \to \mathbb{F}$ definidas por $f_j(p(t)) := p(c_j)$.

Ejercicio 1.9. Encontrar los subespacios $\ker T$ y $T(\mathbb{R}^3)$ y las dimensiones $n(T)$ y $r(T)$, si $T \in \mathcal{L}(\mathbb{R}^3,\mathbb{R}^3)$ y la matriz de T 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 1.10. Si $T \in \mathcal{L}(V,W)$, $S \in \mathcal{L}(W,Z)$, demostrar las siguientes relaciones entre núcleos e imágenes:

$$\ker(ST) \supseteq \ker T, \qquad ST(V) \subseteq S(W),$$
$$\ker((ST)^t) \supseteq \ker(S^t), \qquad (ST)^t(Z^*) \subseteq T^t(W^*).$$

Concluir que $r(ST) \leq r(S)$ y que $r(ST) \leq r(T)$.

Ejercicio 1.11. (a) Si V, W son espacios vectoriales finitodimensionales con $\dim V > \dim W$, y si $T \in \mathcal{L}(V,W)$, demostrar que T no es inyectivo.

(b) Sea $C[a,b] := \{ f\colon [a,b] \to \mathbb{R} \text{ continua} \}$. Defínase $T\colon C[a,b] \to C[a,b]$ por

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

Demostrar que T es lineal e inyectiva, pero no sobreyectiva. Concluir que $C[a,b]$ es infinitodimensional sobre $\mathbb{R}$.

Ejercicio 1.12. Si $A \in M_n(\mathbb{F})$ es una matriz inversible, demostrar que $(A^t)^{-1} = (A^{-1})^t$. Concluir que A^{-1} es simétrica cuando A es simétrica.

Ejercicio 1.13. 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 1.14. La base estándar para $M_2(\mathbb{F})$ 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 $M = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$, defínase $L_M(A) := MA$, $R_M(A) := AM$ y $T(A) := A^t$. Demostrar que L_M, R_M 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 1.15. Sea A una matriz triangular superior con ceros en la diagonal:

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

así que $a_{ij} = 0$ para $i \geq j$. Demostrar que $A^n = O$. Concluir que $I_n + A$ es inversible, con

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

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

Ejercicio 1.16. Resolver el sistema de ecuaciones

$$\begin{aligned} x_1 + x_2 + x_3 &= 0 \\ 3x_1 + 3x_2 + 4x_3 &= 2 \\ x_1 + 2x_2 + x_3 &= -4 \end{aligned}$$

por el método de eliminación gaussiana con intercambio de filas. Escribir la factorización $A = PLV$ de la matriz de coeficientes.

Ejercicio 1.17. 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 1.18. Convertir cada uno de estas matrices en la forma escalonada equivalente:

$$A = \begin{bmatrix} 1 & 2 & 0 & 2 & 1 \\ -1 & -2 & 1 & 1 & 0 \\ 1 & 2 & -3 & -7 & -2 \end{bmatrix}, \qquad B = \begin{bmatrix} 2 & -1 & 3 & 1 & 1 \\ -1 & 0 & -2 & 1 & -3 \\ 1 & 2 & -1 & -4 & 3 \\ 3 & 2 & -2 & -3 & -1 \end{bmatrix}.$$

Ejercicio 1.19. Calcular $r(A)$, encontrar una base para el espacio de soluciones de $A\boldsymbol{x} = \boldsymbol{0}$ y describir el conjunto de soluciones de $A\boldsymbol{x} = \boldsymbol{b}$, donde

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

Ejercicio 1.20. Demostrar que cada matriz de rango k es una suma de k matrices de rango 1. [[Indicación: Usar la Proposición 1.39.]]

Ejercicio 1.21. Verificar la identidad

$$\begin{vmatrix} x & 1 & 0 & 0 & 0 \\ 4 & x & 2 & 0 & 0 \\ 0 & 3 & x & 3 & 0 \\ 0 & 0 & 2 & x & 4 \\ 0 & 0 & 0 & 1 & x \end{vmatrix} = x(x^2 - 4)(x^2 - 16).$$

Ejercicio 1.22. Verificar el *determinante de Vandermonde*[19]

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

⟦ Indicación: Usar eliminación gaussiana. ⟧

Ejercicio 1.23. 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 1.24. Si $A \in M_m(\mathbb{F})$, $B \in \mathbb{F}^{m\times n}$, $C \in \mathbb{F}^{n\times m}$ y $D \in M_n(\mathbb{F})$, demostrar que

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

si en ambos casos O representa un rectángulo de ceros.

Ejercicio 1.25. Sea $U_n \in M_n(\mathbb{F})$ la matriz cuadrada cuyas entradas son *todas* iguales a 1. Demostrar que

$$\det(U_n - I_n) = (-1)^{n-1}(n-1), \qquad \det(U_n + I_n) = n + 1.$$

Ejercicio 1.26. Obtener el rango de la matriz

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

por (a) cálculo de menores; (b) cambio a forma escalonada.

Ejercicio 1.27. Se puede definir el determinante de una aplicación lineal $T \in \mathcal{L}(V,V)$ por $\det T := \det A$, donde A la matriz de T respecto de alguna base de V. Comprobar que esta definición es consistente: esto es, si B es la matriz de T respecto de otra base de V, verificar que $\det B = \det A$.

[19]En 1770, Alexandre Vandermonde escribió un ensayo sobre la solución general de una ecuación polinomial de grado n. Sus ideas fueron generalizadas en el trabajo de *Joseph-Louis Lagrange*, “Réflexions sur la résolution algébrique des équations”, Mémoirs de l'Academie Royale des Sciences et Belles-Lettres de Berlin, 1771. Lagrange quiso expresar las soluciones como combinaciones de las raíces n-ésimas de 1 (llamados “resolventes de Lagrange”) y para despejar los coeficientes usó un determinante cuyas columnas son potencias sucesivas de la segunda columna. Posteriormente, la autoría de este determinante fue atribuido a Vandermonde.

Ejercicio 1.28. 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 1.29. (a) Si $A \in M_n(\mathbb{F})$ con $n \geq 2$, demostrar que $\det(\operatorname{adj} A) = (\det A)^{n-1}$.

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

[[Indicación: Considerar el producto de matrices $A(\operatorname{adj} A)\operatorname{adj}(\operatorname{adj} A)$.]]

Ejercicio 1.30. Si $A \in M_n(\mathbb{F})$, $\boldsymbol{b}, \boldsymbol{c} \in \mathbb{F}^n$ y $d \in \mathbb{F}$, sea $\begin{bmatrix} A & \boldsymbol{b} \\ \boldsymbol{c}^t & d \end{bmatrix}$ la matriz $(n+1)\times(n+1)$ formado al bordear A por la columna $\boldsymbol{b}$, la fila $\boldsymbol{c}^t$ y la entrada d. Demostrar que

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

Ejercicio 1.31. Resolver este sistema de ecuaciones por la regla de Cramer:

$$\begin{aligned} x_1 + 4x_2 - x_3 &= 1 \\ x_1 + x_2 + x_3 &= 0 \\ 2x_1 + 3x_3 &= 0\,. \end{aligned}$$

Ejercicio 1.32. (a) Sean (x,y) las coordenadas de un punto en $\mathbb{R}^2$. 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) 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 luego que pasa por los tres puntos dados.]]

2 Estructura de Aplicaciones Lineales

El álgebra lineal consiste mayormente en el estudio de las propiedades de aplicaciones lineales. En este capítulo se analizará la estructura de una aplicación lineal de un espacio vectorial V en sí mismo. Mucho depende de si V posee alguna estructura extra, como por ejemplo un producto escalar: el siguiente capítulo abordará ese caso. Por ahora, se considera la situación en donde V es finitodimensional, sin usar un concepto de ortogonalidad. A cada aplicación lineal se le asocia unos polinomios que sirven para revelar su estructura.

2.1 Autovalores y autovectores

Definición 2.1. Sea V un espacio vectorial 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 lineales sobre V se denotará por $\operatorname{End}_{\mathbb{F}}(V)$, o bien por $\underline{\operatorname{End}(V)}$ cuando el cuerpo $\mathbb{F}$ es implícito del contexto.[1]

El espacio vectorial $\operatorname{End}(V)$ es también un *anillo*, cuya operación multiplicativa es la *composición de operadores*. En efecto, si $R,S,T \in \operatorname{End}(V)$, entonces $R(ST) = (RS)T$ (asociatividad); la aplicación identidad $I\colon \boldsymbol{x} \mapsto \boldsymbol{x}$ cumple $IT = TI = T$; y las leyes distributivas $T(R+S) = TR+TS$ y $(R+S)T = RT+ST$ también se cumplen. Además, la composición de operadores es compatible con la multiplicación escalar: $c(ST) = (cS)T = S(cT)$ para $S,T \in \operatorname{End}(V)$ y $c \in \mathbb{F}$, ya que estas tres expresiones llevan $\boldsymbol{x} \in V$ en $cS(T(\boldsymbol{x})) \in V$. En otras palabras, $\operatorname{End}(V)$ es un **álgebra** sobre $\mathbb{F}$.

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

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

tenga una solución $\boldsymbol{x} \neq \boldsymbol{0}$. Un vector *no nulo*[2] $\boldsymbol{x} \in V$ que cumple (2.1) se llama un **autovector** asociado al autovalor λ.

Algunos autores dicen *valor propio* en vez de “autovalor” y *vector propio* en vez de “autovector”.[3]

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 (1.6)

[1] Una aplicación lineal de V en sí mismo recibe el nombre de *endomorfismo* de V. Hay que advertir que este término se vuelve ambiguo cuando el espacio vectorial V posee más estructura (un producto, por ejemplo), en cuyo caso se podría demandar que un endomorfismo de V en V preserva todas sus operaciones algebraicas. Para evitar esa clase de discusiones, se emplea el término *operador lineal* en vez de “endomorfismo” en este texto.

[2] La ecuación $T(\boldsymbol{x}) = \lambda\boldsymbol{x}$ tiene la solución trivial $\boldsymbol{x} = \boldsymbol{0}$ cualquiera que sea el coeficiente λ. Se descarta siempre la solución trivial: el vector $\boldsymbol{0}$ nunca puede ser autovector de un operador lineal.

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

es aplicable con la misma base en el dominio y el codominio de T, es decir,

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

Por la discusión después de la Definición 1.27, 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 $\operatorname{End}(V)$, $M_n(\mathbb{F})$ y $\operatorname{End}(\mathbb{F}^n)$. Además, 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* sobre $\mathbb{F}$.

Así las cosas, cada propiedad de aplicaciones lineales induce una propiedad paralela de matrices. Por ejemplo, las matrices pueden poseer 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

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

tenga una solución $\boldsymbol{x} \neq \mathbf{0}$ 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.

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

(a) λ es un autovalor de T;

(b) el operador lineal $(T - \lambda I)$ no es inversible en $\operatorname{End}(V)$*;*

(c) $\ker(T - \lambda I) \neq \{\mathbf{0}\}$.

Demostración. (a) $\iff$ (b): Un escalar λ es un autovalor de T si y sólo si hay $\boldsymbol{x} \in V$ con $\boldsymbol{x} \neq \mathbf{0}$ y $T(\boldsymbol{x}) = \lambda\boldsymbol{x}$, si y sólo si hay $\boldsymbol{x} \neq \mathbf{0}$ tal que $(T - \lambda I)(\boldsymbol{x}) = \mathbf{0}$, si y sólo si $(T - \lambda I)$ no es inyectivo, si y sólo si $(T - \lambda I)$ no es biyectivo. Esta última equivalencia se debe a que $n(T - \lambda I) + r(T - \lambda I) = \dim V$; por lo tanto, un operador lineal es inyectivo si y sólo si es sobreyectivo.[4]

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

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

(a) λ es un autovalor de A;

(b) la matriz $(A - \lambda I_n)$ no es inversible en $M_n(\mathbb{F})$;

(c) $\det(A - \lambda I_n) = 0$.

[4]Esta conclusión depende de la finitud de $\dim V$, para que las igualdades $n(T - \lambda I) = 0$ y $r(T - \lambda I) = \dim V$ sean equivalentes. Sobre espacios vectoriales de dimensión infinita, hay operadores lineales inyectivas pero no sobreyectivas. Véase el Ejercicio 1.11, por ejemplo.

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 1.47. □

Corolario 2.6. *Si $A \in M_n(\mathbb{F})$ es una matriz* triangular*, sus autovalores 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 $\det(A - \lambda I_n) = 0$ o bien, lo que es lo mismo, $\det(\lambda I_n - A) = 0$. Explícitamente,

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

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

En el caso de que A sea 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 se define por

$$p_A(t) := \det(t I_n - A) = (-1)^n \det(A - t I_n). \tag{2.3}$$

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 (1.15) muestra que

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

donde cada uno de los "otros términos" es un producto de (± 1) por n entradas de la matriz $t I_n - A$, de las cuales a lo sumo $(n-2)$ entradas pueden ser diagonales: esta parte forma un polinomio de grado no mayor que $(n-2)$. Entonces se ve que

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

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

[5] Algunos autores definen $p_A(t) := \det(A - t I_n)$. Bajo ese convenio, el primer coeficiente no nulo sería $(-1)^n$. No es mucha la diferencia; sin embargo, es más cómodo elegir el signo de manera que el polinomio característico sea mónico.

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 sólo si hay una matriz inversible P tal que $B = P^{-1}AP$. Ahora $\det P^{-1} = 1/(\det P)$ porque $(\det P^{-1})(\det P) = \det(P^{-1}P) = \det I_n = 1$. 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 $T \in \operatorname{End}(V)$, 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 $P = [I]^{\mathcal{B}}_{\mathcal{C}}$ es inversible, con inverso $P^{-1} = [I]^{\mathcal{C}}_{\mathcal{B}}$; por tanto,

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

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

Definición 2.11. Sea $T \in \operatorname{End}(V)$ un operador lineal sobre un espacio vectorial finitodimensional V. El **polinomio característico de** T es el polinomio $p_T(t) \in \mathbb{F}[t]$ definido por

$$p_T(t) := \det(tI - T) = (-1)^{\dim V} \det(T - tI). \tag{2.5}$$

Proposición 2.12. *Si $A \in M_n(\mathbb{F})$ es una matriz cuadrada, los autovalores de A 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 sólo si $p_A(\lambda) = 0$. $\square$

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

$$J = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{2.6}$$

Su polinomio característico es $\begin{vmatrix} t & -1 \\ 1 & t \end{vmatrix} = t^2 + 1$. Ahora, si $\mathbb{F} = \mathbb{R}$ ó $\mathbb{Q}$, 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 $\mathbb{F} = \mathbb{C}$, la factorización $t^2 + 1 = (t - i)(t + i)$ muestra que $\{i, -i\}$ podrían ser autovalores de J. Es fácil adivinar un par de autovectores en $\mathbb{C}^2$, para verificar que i y $-i$ son en efecto autovalores de J; por ejemplo,

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

[6]En el caso $\mathbb{F} = \mathbb{F}_p := \{\bar{0}, \bar{1}, \dots, \overline{p-1}\}$, 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 sólo si $p = 2$ o bien $p = 4m + 1$ para algún $m \in \mathbb{N}$.

En general, la búsqueda de autovectores es un asunto be 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

$$\begin{aligned} p_A(t) = \det(tI_3 - A) = \begin{vmatrix} t-1 & 0 & -2 \\ 0 & t+1 & 2 \\ -2 & 2 & t \end{vmatrix} &= (t-1)\begin{vmatrix} \lambda+1 & 2 \\ 2 & t \end{vmatrix} - 2\begin{vmatrix} 0 & t+1 \\ -2 & 2 \end{vmatrix} \\ &= (t-1)(t^2+t-4) - 2(2t+2) = t^3 - 9t \\ &= t(t-3)(t+3). \end{aligned}$$

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, hay que resolver (por eliminación gaussiana) los tres sistemas de ecuaciones de ecuaciones homogéneas $(\lambda I_3 - A)\boldsymbol{x} = \mathbf{0}$ para $\lambda = 0, 3, -3$ respectivamente. En cada caso, se cambia la matriz aumentada $[\lambda I_3 - A \mid \mathbf{0}]$ 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 el múltiplo x_3. En detalle:

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

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,

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

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 inversible: 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

$$P^{-1}AP = D. \tag{2.7}$$

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

► Para obtener una *fórmula* para el polinomio característico, conviene introducir un poco de notación para submatrices.

Notación. Considérese dos juegos de índices $I := \{i_1, \dots, i_k\} \subseteq \{1, \dots, m\}$ y $J := \{j_1, \dots, j_l\} \subseteq \{1, \dots, n\}$, numerados en orden creciente: $i_1 < i_2 < \cdots < i_k$ y $j_1 < j_2 < \cdots < j_l$. Si A es una matriz $m \times n$, denótese por $\underline{A_{IJ}}$ la *submatriz* $k \times l$ de A formado por las entradas a_{ij} con $i \in I$, $j \in J$.

Sean $\underline{I' := \{1, \dots, m\} \setminus I}$ y también $J' := \{1, \dots, n\} \setminus J$. Se dice 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 $I = \{i_1, \dots, i_k\} \subseteq \{1, \dots, n\}$, entonces $A_{II} \in M_k(\mathbb{F})$ se llama una *submatriz principal* de A. Su determinante $m_{II} := \det A_{II}$ se llama un **menor principal** de A. Para cada $k = 1, \dots, n$, hay $\binom{n}{k}$ menores principales de A obtenidos de submatrices $k \times k$.

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

$$p_A(t) = t^n - \tau_1(A)\,t^{n-1} + \tau_2(A)\,t^{n-2} - \cdots + (-1)^{n-1}\tau_{n-1}(A)\,t + (-1)^n\tau_n(A), \tag{2.8a}$$

donde $\tau_1(A) = \underline{\operatorname{tr} A} := a_{11} + \cdots + a_{nn}$ es la traza *de A; $\tau_n(A) = \det A$; y en general*

$$\tau_k(A) = \sum_{|I|=k} \det A_{II} \quad \textit{para} \quad k = 1, \dots, n \tag{2.8b}$$

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

Demostración. En el desarrollo de Leibniz del determinante

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

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

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

donde la suma recorre las permutaciones $\sigma \in S_n$ que dejan fijos los índices diagonales en I', es decir, $\sigma(i_r) = j_r$ para $r = 1, \dots, k$; $\sigma(l) = l$ para $l \notin I$. Entonces se puede escribir $\sigma = \tau\rho_{II'}$, donde $\rho_{II'}$ es la *permutación de baraje*[7] que lleva $(1, \dots, n)$ en $(I, I') = (i_1, \dots, i_k, i'_1, \dots, i'_{n-k})$ con $i_1 < \cdots < i_k$ y $i'_1 < \cdots < i'_{n-k}$; y τ es una permutación de $\{1, \dots, k\}$ que deja fijos $k+1, \dots, n$. Luego, el coeficiente de t^{n-k} es

$$\sum_{|I|=k} \sum_{\tau} (-1)^{\tau} (-a_{i_1 i_{\tau(1)}}) \dots (-a_{i_k i_{\tau(k)}}) = (-1)^k \sum_{|I|=k} \sum_{\tau} (-1)^{\tau} a_{i_1 i_{\tau(1)}} \dots a_{i_k i_{\tau(k)}} = (-1)^k \sum_{|I|=k} \det A_{II}.$$

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

[7] *Barajar* un naipe significa separar el naipe en dos partes y luego permutar las cartas de modo que se conserve el orden relativo dentro de cada parte. Las permutaciones de baraje forman un tema importante en la teoría combinatoria. Véase, por ejemplo: Richard Stanley, *Enumerative Combinatorics*, tomo 1, Cambridge University Press, 1997.

Corolario 2.17. *Las sumas de menores principales son invariantes bajo semejanza: si $A \in M_n(\mathbb{F})$ y si $P \in M_n(\mathbb{F})$ es inversible, entonces $\underline{\tau_k(A) = \tau_k(P^{-1}AP)}$ para $k = 1, \dots, n$.* ⊟

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

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

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.[8]

Proposición 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 las filas de A. Si $J = \{j_1, \dots, j_k\}$ es un juego de índices de k columnas cualesquiera, sea $s(I, J) := i_1 + \cdots + i_k + j_1 + \cdots + j_k$. Entonces*

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

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

La demostración de esta fórmula se deja como ejercicio. ⊟

2.2 El teorema de Cayley y Hamilton

Antes de abordar la propiedad más famosa del polinomio característico, es útil recordar ciertos propiedades elementales de los polinomios. Ya se sabe que $\mathbb{F}[t]$ es un álgebra conmutativa sobre el cuerpo $\mathbb{F}$. Esta álgebra es *entera*,[9] es decir, no posee "divisores de cero": si $f(t) \neq 0$

[8] *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.

[9] La terminología tiene una historia curiosa. Un anillo 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 $\mathbb{Z}$. A veces A se llama "dominio entero" o, menos correctamente, "dominio de integridad": Kronecker empleó este término para distinguirlo de un *cuerpo*, que él llamó "dominio de racionalidad".

y $g(t) \neq 0$, entonces $f(t)g(t) \neq 0$ también. Esto es evidente al recordar la ley de producto:

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

porque $a_n \neq 0$, $b_m \neq 0$ implican $a_n b_n \neq 0$. 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 $\underline{g(t) \setminus f(t)}$. En el caso contrario, donde $g(t)$ no divide $f(t)$, se puede ejecutar una *división con residuo*, según el lema siguiente.[10]

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

$$f(t) = q(t)g(t) + r(t), \quad \textit{con} \quad \begin{cases} \operatorname{gr} r(t) < \operatorname{gr} g(t), \\ \textit{o bien}\ \ r(t) = 0. \end{cases} \tag{2.10}$$

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

En cambio, si $m \leq n$, 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

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

y el resultado (2.10) sigue por la inducción sobre n.

Para la *unicidad* de $q(t)$ y $r(t)$, obsérvese que si $q(t)g(t) + r(t) = \tilde{q}(t)g(t) + \tilde{r}(t)$, entonces $(q(t) - \tilde{q}(t))g(t) = \tilde{r}(t) - r(t)$. Si esta ecuación no es $\underline{0 = 0}$, entonces al lado izquierdo el grado sería $\geq m$, mientras al lado derecho el grado sería $< m$, lo cual es imposible. Por tanto $\tilde{r}(t) = r(t)$ y $(q(t) - \tilde{q}(t))g(t) = 0$. Como $\mathbb{F}[t]$ es entero y $g(t) \neq 0$, se concluye que $q(t) - \tilde{q}(t) = 0$. □

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 (2.10). 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 $f(a) = (a-a)q(a) + r(a) = r(a) = r_0$. □

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

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

Definición 2.22. Si $f(t), g(t)$ son dos polinomios en $\mathbb{F}[t]$, su **máximo común divisor** $k(t) = \underline{\mathrm{mcd}(f(t), g(t))}$ 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 $k(t) = t^m + c_{m-1}t^{m-1} + \cdots + c_1 t + c_0$.

Es fácil ver que el máximo común divisor de dos polinomios es único, si existe. Su existencia puede comprobarse con el *algoritmo euclidiano*, en estricta analogía con el proceso de encontrar el máximo común divisor de dos números enteros. El Lema 2.19 produce una sucesión finita de divisiones con residuo:

$$\begin{aligned} f(t) = q_1(t)g(t) + r_1(t), \quad g(t) = q_2(t)r_1(t) + r_2(t), \quad r_1(t) = q_3(t)r_2(t) + r_3(t), \ \dots \\ r_{j-2}(t) = q_j(t)r_{j-1}(t) + r_j(t), \qquad r_{j-1}(t) = q_{j+1}(t)r_j(t) + 0, \end{aligned} \tag{2.11}$$

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.

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

$$a(t)\,f(t) + b(t)\,g(t) = \mathrm{mcd}(f(t), g(t)).$$

Demostración. Fíjese que $r_1(t) = f(t) - q_1(t)g(t)$, a partir de (2.11). Además,

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

Por sustitución repetida, se hallan $a_i(t), b_i(t) \in \mathbb{F}[t]$ tales que $r_i(t) = a_i(t)f(t) + b_i(t)g(t)$, para $i = 1, \dots, j$. Al dividir la j-ésima de estas ecuaciones por el coeficiente inicial de $r_j(t)$, se obtiene la relación deseada. □

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

$$a(t)\,f(t) + b(t)\,g(t) = 1.$$

► Un procedimiento muy útil, a veces 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 $\mathrm{End}(V)$.

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

$$f(A) := c_n A^n + \cdots + c_1 A + c_0 I_n \in M_n(\mathbb{F}). \tag{2.12a}$$

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

De igual manera, sea $T \in \operatorname{End}(V)$, donde V es un espacio vectorial sobre $\mathbb{F}$. Defínase

$$f(T) := c_n T^n + \cdots + c_1 T + c_0 I \in \operatorname{End}(V). \tag{2.12b}$$

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

Si $A = [T]_{\mathcal{B}}^{\mathcal{B}}$ es la matriz de T con respecto a 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 $\operatorname{End}(V)$ no son conmutativos. Tampoco son inyectivos, porque $M_n(\mathbb{F})$ y $\operatorname{End}(V)$ son finitodimensionales y $\mathbb{F}[t]$ es infinitodimensional. Entonces, dada una matriz A, debe de haber polinomios no nulos $f(t)$ tales que $f(A) = 0$. El siguiente teorema, debido a Hamilton[11] y a Cayley,[12] proporciona un polinomio específico con esta propiedad, el cual de hecho es el polinomio característico de A.

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

Demostración. La regla de Cramer demuestra que[13]

$$(tI_n - A)\operatorname{adj}(tI_n - A) = \det(tI_n - A)\, I_n = p_A(t)\, I_n. \tag{2.13}$$

Las entradas de la matriz $\operatorname{adj}(tI_n - A)$ son, salvo signo, menores $(n-1)\times(n-1)$ de la matriz $tI_n - A$. 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

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

[11] *William Rowan Hamilton* desarrolló la teoría de *cuaterniones*, que combinan escalares reales y vectores en un espacio vectorial $\mathbb{H} = \mathbb{R} \oplus \mathbb{R}^3$, dotado de un producto no conmutativo. Las aplicaciones lineales en $\operatorname{End}_{\mathbb{R}}(\mathbb{H})$ 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.

[12] *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 mostró los casos 2×2 y 3×3.

[13] La regla de Cramer es válido para matrices con entradas escalares. Para justificar (2.13), 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 $B \operatorname{adj} B = (\det B)\, I_n$ es una abreviatura para n^2 identidades polinomiales en las entradas de $B \in M_n(\mathbb{F})$, que sigue válido cuando el cuerpo de escalares $\mathbb{F}$ queda reemplazada por el álgebra $\mathbb{F}[t]$.

Al escribir $p_A(t) = t^n + c_{n-1}t^{n-1} + \cdots + c_1 t + c_0$, la ecuación (2.13) queda en la forma[14]

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

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

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

Al multiplicarlas por potencias sucesivas de A, se obtiene

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

Una suma telescópica de estas relaciones produce el resultado:

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

Corolario 2.27. *Sea $T \in \operatorname{End}(V)$ un operador lineal sobre el espacio vectorial V y sea $p_T(t) \in \mathbb{F}[t]$ su polinomio característico. Entonces $p_T(T) = 0$ en $\operatorname{End}(V)$.* ⊟

Proposición 2.28. *Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada. Entre todos los polinomios mónicos $f(t) \in \mathbb{F}[t]$ tales que $f(A) = O$, hay un único $q(t)$ de mínimo grado. Este $q(t)$ divide cualquier $f(t)$ tal que $f(A) = O$.*

Demostración. Sea $f(t)$ cualquier polinomio con $f(A) = O$ y sea $q(t)$ cualquier polinomio mónico tal que $q(A) = O$ en $M_n(\mathbb{F})$. Por divisibilidad con residuo (2.10), se puede escribir

$$f(t) = s(t)q(t) + r(t),$$

para un único par de polinomios $s(t)$, $r(t)$, donde $\operatorname{gr} r(t) < \operatorname{gr} q(t)$ si $r(t)$ no es nulo.

Además, $r(A) = f(A) - s(A)q(A) = O$. Cuando $m = \operatorname{gr} q(t)$ tiene su menor valor posible, se concluye que $r(t) = 0$. Por lo tanto, $f(t) = s(t)q(t)$, es decir, $q(t) \setminus f(t)$.

Si $\tilde{q}(t)$ es otro polinomio mónico de grado m con $\tilde{q}(A) = O$, el mismo argumento muestra que $\tilde{q}(t) = u(t)q(t)$ para algún polinomio $u(t)$. Por conteo de grados, se ve que $\operatorname{gr} u(t) = 0$, es decir, $u(t)$ es constante. Como $q(t)$ y $\tilde{q}(t)$ son mónicos, es $u(t) = 1$; luego $\tilde{q}(t) = q(t)$. □

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

[14] Ya se sabe por (2.8) que $c_k = (-1)^{n-k}\tau_{n-k}(A)$, pero esta demostración no requiere la fórmula explícita.

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

El teorema de Cayley y Hamilton muestra que $p_A(A) = O$. Por lo tanto, $\operatorname{gr} q_A(t) \leq n$. La Proposición anterior muestra que $q_A(t)$ divide $p_A(t)$. En particular, todas las raíces de $q_A(t)$ son autovalores de A. (La inversa también vale, como se verá más adelante, en el Corolario 2.40: todo autovalor de A es una raíz de su polinomio mínimo.)

De igual modo, si $T \in \operatorname{End}(V)$, 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

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

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

$$(A-3I_4)(A-2I_4) = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \neq O,$$

pero que $(A-3I_4)^2(A-2I_4) = O$, por cálculo directo. Se concluye que $q_A(t) = (t-3)^2(t-2)$. [[*Moraleja*: el polinomio mínimo no necesariamente tiene factores distintos.]]

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 (no siempre), se puede elegir esta base tal que la matriz $[T]^{\mathcal{C}}_{\mathcal{C}}$ sea una matriz diagonal.

La búsqueda del representante diagonal se reduce a un problema de clasificar las matrices cuadradas. En efecto, si $A = [T]^{\mathcal{B}}_{\mathcal{B}}$ es una matriz cualquiera que representa $T \in \operatorname{End}(V)$, se obtiene cualquier otro representante por cambio de base (de $\mathcal{B}$ a $\mathcal{C}$, concretamente). Si $P = [I]^{\mathcal{B}}_{\mathcal{C}}$ 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 problema matricial es el siguiente: *dada una matriz $A \in M_n(\mathbb{F})$, se busca una matriz inversible P tal que $P^{-1}AP$ sea una matriz diagonal.*

Este problema admite una solución, en primera instancia, si la matriz A posee *autovalores distintos*, en vista del siguiente resultado.

Proposición 2.32. *Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada. Si $\{\lambda_1, \dots, \lambda_k\}$ son* autovalores distintos *de A y si $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k\}$ son unos autovectores correspondientes, entonces los autovectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k$ 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 $\boldsymbol{x}_1 \neq \boldsymbol{0}$, pues $\boldsymbol{x}_1$ es un autovector.)

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

$$c_1\boldsymbol{x}_1 + c_2\boldsymbol{x}_2 + \cdots + c_k\boldsymbol{x}_k = \boldsymbol{0}, \tag{2.14}$$

con $c_1, \dots, c_k$ no todos cero. Renumerando la lista si fuera necesario, puede suponerse 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 (2.14) de esta relación, se obtiene

$$c_2(\lambda_2 - \lambda_1)\boldsymbol{x}_2 + c_3(\lambda_3 - \lambda_1)\boldsymbol{x}_3 + \cdots + c_k(\lambda_k - \lambda_1)\boldsymbol{x}_k = \boldsymbol{0}. \tag{2.15}$$

Los coeficientes en la ecuación (2.15) 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. $\square$

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

Demostración. Sea $\{\boldsymbol{p}_1, \dots, \boldsymbol{p}_n\}$ un juego de n autovectores que corresponde a los n autovalores distintos $\{\lambda_1, \dots, \lambda_n\}$ de A. Por la proposición anterior, $\mathcal{B} = \{\boldsymbol{p}_1, \dots, \boldsymbol{p}_n\}$ es una base de $\mathbb{F}^n$. Con respecto a la base estándar $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$, cada $\boldsymbol{p}_s$ puede desarrollarse así:

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

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

$$\begin{aligned} AP &= A\,[\boldsymbol{p}_1\ \boldsymbol{p}_2\ \dots\ \boldsymbol{p}_n] = [A\boldsymbol{p}_1\ A\boldsymbol{p}_2\ \dots\ A\boldsymbol{p}_n] \\ &= [\lambda_1\boldsymbol{p}_1\ \lambda_2\boldsymbol{p}_2\ \dots\ \lambda_n\boldsymbol{p}_n] = [\boldsymbol{p}_1\ \boldsymbol{p}_2\ \dots\ \boldsymbol{p}_n] \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} = PD, \end{aligned} \tag{2.16}$$

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 inversible 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. $\square$

Notación. La notación compacta

$$\underline{\operatorname{diag}[\lambda_1,\lambda_2,\dots,\lambda_n]} := \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \lambda_n \end{bmatrix} \tag{2.17}$$

denota la matriz diagonal con entradas diagonales $\lambda_1,\dots,\lambda_n$. 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 (1.6). 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 sólo si A tiene n autovectores linealmente independientes, si y sólo si $\mathbb{F}^n$ posee una base formado por autovectores de A.*

Demostración. Es evidente que la segunda condición es equivalente a la tercera, porque una base de $\mathbb{F}^n$ no es más que una colección de n vectores linealmente independientes.

Si $\{\boldsymbol{p}_1,\dots,\boldsymbol{p}_n\} \subset \mathbb{F}^n$ 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, por lo tanto su rango es n y la matriz P es *inversible*. Es $A\boldsymbol{p}_k = \lambda_k \boldsymbol{p}_k$ para $k = 1,\dots,n$, donde λ_k es el autovalor correspondiente al autovector $\boldsymbol{p}_k$. Si $D := \operatorname{diag}[\lambda_1,\dots,\lambda_n]$ es la matriz diagonal cuyas entradas diagonales son estos autovalores en el orden prescrito, entonces vale $AP = PD$, según (2.16). Se concluye que $P^{-1}AP = D$, es decir, A es diagonalizable con forma diagonal D.

Por otro lado, si A es diagonalizable, hay una matriz inversible P y una matriz diagonal $D := \operatorname{diag}[\lambda_1,\dots,\lambda_n]$ tal que $P^{-1}AP = D$. Por ende, es $AP = PD$; al comparar la k-ésima columna de ambos lados de esta igualdad matricial, se ve de (2.16) 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 inversible P tiene rango n, es decir, sus columnas son linealmente independientes y constituyen una base de $\mathbb{F}^n$. $\square$

La proposición anterior no requiere 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 $\{\nu_1,\dots,\nu_r\}$ los elementos *distintos* del juego de autovalores $(\lambda_1,\dots,\lambda_n)$. Si ν_1 ocurre k_1 veces, ν_2 ocurre k_2 veces,..., ν_r ocurre k_r veces, con $k_1 + k_2 + \cdots + k_r = n$, se puede permutar los λ_i para obtener

$$(\lambda_1,\dots,\lambda_n) = (\underbrace{\nu_1,\dots,\nu_1}_{k_1\text{ veces}},\underbrace{\nu_2,\dots,\nu_2}_{k_2\text{ veces}},\dots,\underbrace{\nu_r,\dots,\nu_r}_{k_r\text{ veces}}).$$

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

$$p_D(t) = (t-\nu_1)^{k_1}(t-\nu_2)^{k_2}\dots(t-\nu_r)^{k_r}.$$

En el caso diagonal, el teorema de Cayley y Hamilton tiene una comprobación directa: en el producto de matrices diagonales $(D-\nu_1 I_n)^{k_1}(D-\nu_2 I_n)^{k_2}\dots(D-\nu_r I_n)^{k_r}$ al menos uno de los factores tiene una entrada diagonal 0 en cada fila.

El *polinomio mínimo* de esta matriz D es

$$q_D(t) = (t-\nu_1)(t-\nu_2)\dots(t-\nu_r), \tag{2.18}$$

con r factores distintos de primer grado. El producto $(D-\nu_1 I_n)(D-\nu_2 I_n)\dots(D-\nu_r I_n)$ es la matriz O porque cada entrada diagonal es un producto de r escalares que incluyen un cero. Si se suprimiera uno de los factores $(t-\nu_i)$, el producto de matrices con $(D-\nu_i I_n)$ omitido posee entradas diagonales no nulas.

Sucede 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 comprobarlo, conviene examinar otros aspectos estructurales de las operadores lineales en general.

2.4 Descomposición primaria de un operador lineal

Definición 2.35. Sea V un espacio vectorial sobre $\mathbb{F}$ y sea $T \in \operatorname{End}(V)$ un operador lineal. se dice que un subespacio $W \leq V$ es un **subespacio invariante** para T si $\underline{T(W) \subseteq W}$.

Si W es un subespacio invariante para $T \in \operatorname{End}(V)$, con $\dim V = n$ y $\dim W = m \leq n$, 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 W. (Es cuestión de elegir una base de W y luego completarla en una base de V.) En el desarrollo (2.2) del operador T en esta base, la condición $T(W) \subseteq W$ implica que $a_{ij} = 0$ cuando $j \leq m$ pero $i > m$; en otras palabras, la matriz $[T]_{\mathcal{B}}^{\mathcal{B}}$ tiene la forma

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

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

Definición 2.36. Se dice que un subespacio invariante $W \leq V$ **reduce** el operador lineal $T \in \operatorname{End}(V)$ si hay otro subespacio invariante $U \leq V$ tal que $V = W \oplus U$. Si así fuera, existiría una base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V tal que $W = \operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_m \rangle$ y $U = \operatorname{lin}\langle \boldsymbol{x}_{m+1}, \dots, \boldsymbol{x}_n \rangle$, en cuyo caso la esquina X de la matriz $[T]_{\mathcal{B}}^{\mathcal{B}}$ es también un bloque de ceros:

$$[T]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} A & O \\ O & B \end{bmatrix}. \tag{2.19}$$

La matriz a la derecha de (2.19) se llama la **suma directa** de las matrices A y B.

Proposición 2.37. *Si el polinomio característico de un operador lineal $T \in \operatorname{End}(V)$ se factoriza en $p_T(t) = h(t)\,k(t)$ con $\operatorname{mcd}(h(t), k(t)) = 1$, entonces $V = W \oplus U$, donde $W = \ker h(T)$ y $U = \ker k(T)$ 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 $h(T)\,a(T) + k(T)\,b(T) = I$ en $\operatorname{End}(V)$. Ahora defínase $W := \operatorname{im} k(T)$ y $U := \operatorname{im} h(T)$. Ellos son subespacios invariantes para T, ya que $T(k(T)(\boldsymbol{x})) = k(T)(T(\boldsymbol{x}))$ y $T(h(T)(\boldsymbol{y})) = h(T)(T(\boldsymbol{y}))$. Además, para cada $\boldsymbol{x} \in V$ vale

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

así que $V = W + U$.

Para ver que esta suma de subespacios es directa, tómese $\boldsymbol{x} \in W \cap U$. 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 se obtiene $h(T)(\boldsymbol{x}) = p_T(T)(\boldsymbol{y}) = \boldsymbol{0}$ 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 $W \cap U = \{\boldsymbol{0}\}$ y luego $V = W \oplus U$.

Si $\boldsymbol{x} \in W$, hay $\boldsymbol{y} \in V$ tal que $\boldsymbol{x} = k(T)(\boldsymbol{y})$. Entonces $h(T)(\boldsymbol{x}) = h(T)k(T)(\boldsymbol{y}) = p_T(T)(\boldsymbol{y}) = \boldsymbol{0}$. Por tanto, es $W \subseteq \ker h(T)$. Al contar dimensiones, el teorema de rango y nulidad implica que

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

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

Proposición 2.38. *Sea $T \in \operatorname{End}(V)$ un operador lineal cuyo polinomio característico escinde en $\mathbb{F}[t]$.*[15] *Supóngase que $p_T(t) = h(t)k(t)$ con $\operatorname{mcd}(h(t), k(t)) = 1$. Entonces las restricciones de T a los subespacios $W = \ker h(T)$ y $U = \ker k(T)$ tienen polinomios característicos $h(t)$ y $k(t)$, respectivamente.*

Demostración. Obsérvese, por la demostración de la Proposición 2.37, que los subespacios W y U reducen T: es $T(W) \subseteq W$ y $T(U) \subseteq U$. Sean $T' \in \operatorname{End}(W)$ y $T'' \in \operatorname{End}(U)$ las restricciones de T a W y U, respectivamente. Sea $\mathcal{B}$ una base de W y $\mathcal{C}$ una base de U, así que $\mathcal{B} \uplus \mathcal{C}$ es una base de $W \oplus U = V$. Si $A = [T']^{\mathcal{B}}_{\mathcal{B}}$, $B = [T'']^{\mathcal{C}}_{\mathcal{C}}$, la matriz de T para la base $\mathcal{B} \uplus \mathcal{C}$ es

$$[T]^{\mathcal{B}\uplus\mathcal{C}}_{\mathcal{B}\uplus\mathcal{C}} = \begin{bmatrix} A & O \\ O & B \end{bmatrix}. \tag{2.20}$$

Entonces, si $r = \dim W$, $s = \dim U$, el polinomio característico de T es

$$p_T(t) = \det \begin{bmatrix} tI_r - A & O \\ O & tI_s - B \end{bmatrix} = \det(tI_r - A)\det(tI_s - B) = p_A(t)\,p_B(t).$$

Si λ es una raíz de $p_A(t)$, entonces hay $\boldsymbol{y} \in W$ no nulo 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 $f(t) = a_n(t-\alpha_1)(t-\alpha_2)\dots(t-\alpha_n)$, donde $n = \operatorname{gr} f(t)$ 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 Algebra.

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,

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

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

En la última Proposición, la hipótesis de que $p_T(t)$ escinde en $\mathbb{F}[t]$ no es indispensable (aunque fue usada en el último paso de la demostración). 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 $\mathbb{C}[t]$.

Lema 2.39. *Sea $V = W \oplus U$, con $W = \ker h(T)$ y $U = \ker k(T)$, la descomposición de V obtenida de una factorización $p_T(t) = h(t)\,k(t)$ —en factores relativamente primos— del polinomio característico de un operador lineal $T \in \mathrm{End}(V)$. Entonces hay una factorización correspondiente del polinomio mínimo $q_T(t) = r(t)\,s(t)$ en factores relativamente primos, donde $r(t) \setminus h(t)$, $s(t) \setminus k(t)$, el operador $r(T)$ se anula en W y $s(T)$ se anula en U.*

Demostración. Elíjase una base de $V = W \oplus U$ tal que T tenga una matriz en bloques de la forma (2.20). Entonces la relación $q_T(T) = 0$ conlleva la relación

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

así que $q_T(A) = O$ y $q_T(B) = O$. Sean $r(t) := \mathrm{mcd}(q_T(t), h(t))$ y $s(t) := \mathrm{mcd}(q_T(t), k(t))$; debe de ser claro que $r(t)$ y $s(t)$ son relativamente primos y que $r(t)s(t) = \mathrm{mcd}((q_T(t), h(t)k(t)) = q_T(t)$, porque $q_T(t)$ divide $p_T(t) = h(t)k(t)$.

También es $h(A) = p_A(A) = O$ y el Lema 2.23 dice que $r(t) = a(t)q_T(t) + b(t)h(t)$ para algunos polinomios $a(t), b(t)$. Luego $r(A) = O$. En otras palabras, $r(T') = 0$ en $\mathrm{End}(W)$, si T' es la restricción de T a W. Del mismo modo, se obtiene $s(B) = O$ y $s(T'') = 0$ en $\mathrm{End}(U)$, si T'' es la restricción de T a U. □

Corolario 2.40. *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, es $p_T(t) = (t-\lambda)^m\,k(t)$ para algún $m \in \{1,2,3,\dots\}$, con $k(\lambda) \neq 0$. El Lema anterior muestra que $q_T(t) = (t-\lambda)^l\,s(t)$, con $l \leq m$, donde $(T-\lambda I)^l$ anula $\ker((T-\lambda I)^m)$, que sería imposible si fuera $l = 0$. (Por el cálculo funcional de la Definición 2.25, es $T^0 := I$ 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 2.41. *Si el polinomio* mínimo *de un operador lineal $T \in \operatorname{End}(V)$ se factoriza en $q_T(t) = r(t)\,s(t)$ con $r(t)$, $s(t)$ mónicos y* $\operatorname{mcd}(r(t), s(t)) = 1$, *entonces $V = W' \oplus U'$, donde $W' = \ker r(T)$ y $U' = \ker s(T)$ son subespacios invariantes para T. Los polinomios mínimos de las restricciones de T a W' y U' son $r(t)$ y $s(t)$, respectivamente.*

Demostración. La demostración de la Proposición 2.37 se repite en forma casi idéntica, con $q_T(t)$, $r(t)$, $s(t)$, W', U' en los lugares respectivos de $p_T(t)$, $h(t)$, $k(t)$, W y U. En vez de usar $p_T(T) = O$ por el teorema de Cayley y Hamilton, se usa $q_T(T) = O$ por la definición del polinomio mínimo. Por tanto, es $V = W' \oplus U'$ con $W' = \ker r(T)$ y $U' = \ker s(T)$. Además, esta demostración conlleva las igualdades $W' = \operatorname{im} s(T)$ y $U' = \operatorname{im} r(T)$.

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

Si $f(t) \in \mathbb{F}[t]$ es un polinomio tal que $f(T') = 0$ en $\operatorname{End}(W')$, entonces $f(T)(\boldsymbol{y}) = \mathbf{0}$ para todo $\boldsymbol{y} = s(T)(\boldsymbol{x}) \in W'$. Por tanto, $f(T)\big(s(T)(\boldsymbol{x})\big) = \mathbf{0}$ para todo $\boldsymbol{x} \in V$; esto es, $f(T)s(T) = 0$ en $\operatorname{End}(V)$. Luego $r(t)s(t) = q_T(t) \setminus f(t)s(t)$; esto es, hay un polinomio $g(t)$ tal que $f(t)s(t) = g(t)r(t)s(t)$ y por ende[16] es $f(t) = g(t)r(t)$. En resumen: si $f(T') = 0$, entonces $r(t) \setminus f(t)$; esto dice que $r(t)$, el cual es mónico, es el polinomio mínimo del operador lineal T'.

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

Teorema 2.42 (Descomposición primaria)**.** *Si el polinomio mínimo de un operador lineal $T \in \operatorname{End}(V)$ se factoriza como $q_T(t) = h_1(t)\dots h_r(t)$ en factores mónicos relativamente primos $h_1(t), \dots, h_r(t)$, entonces $\underline{V = W_1 \oplus \cdots \oplus W_r}$, donde $\underline{W_i = \ker h_i(T)}$ para $i = 1, \dots, r$. Además, cada $h_i(t)$ es el polinomio mínimo de la restricción de T al subespacio invariante W_i.*

Demostración. Por inducción sobre r. Sea $k_1(t) := h_2(t)\dots h_r(t)$, así que $q_T(t) = h_1(t)k_1(t)$ con $\operatorname{mcd}(h_1(t), k_1(t)) = 1$. La Proposición 2.41 muestra que $V = W_1 \oplus U_1$, donde $W_1 := \ker h_1(T)$ y $U_1 := \ker k_1(T) = \operatorname{im} h_1(T)$.

Además, la Proposición 2.41 muestra que las restricciones de T a W_1 y U_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 U_1, con polinomio mínimo $k_1(t) = h_2(t)\dots h_r(t)$. Se obtiene $U_1 = W_2 \oplus \cdots \oplus W_r$, donde $W_i = \ker h_i(T)$, con polinomio mínimo $h_i(T)$ en cada subespacio W_i, para $i = 2, \dots, r$. El resultado ahora es evidente. □

Corolario 2.43. *Sea $T \in \operatorname{End}(V)$ un operador lineal cuyo polinomio característico $p_T(t)$ escinde en $\mathbb{F}[t]$. Sea*

$$p_T(t) = (t - \nu_1)^{k_1}(t - \nu_2)^{k_2} \dots (t - \nu_r)^{k_r} \tag{2.21}$$

su factorización completa, con $\nu_1, \dots, \nu_r \in \mathbb{F}$ distintos. Entonces su polinomio mínimo es de la forma

$$q_T(t) = (t - \nu_1)^{l_1}(t - \nu_2)^{l_2} \dots (t - \nu_r)^{l_r}, \tag{2.22}$$

con $l_1, \dots, l_r \in \mathbb{N}$ y $1 \leq l_i \leq k_i$ para $i = 1, \dots, r$. Sea $V = W_1 \oplus \cdots \oplus W_r$ la descomposición primaria correspondiente, donde $W_i := \ker((T - \nu_i I)^{l_r})$. Entonces $(t - \nu_i)^{k_i}$ es el polinomio característico de la restricción de T al subespacio invariante W_i.

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

Demostración. Por inducción sobre r; el resultado es obvio si $r = 1$. Sea $h(t) := (t-\nu_1)^{k_1}$, $k(t) := (t-\nu_2)^{k_2}\dots(t-\nu_r)^{k_r}$, de modo que $V = W \oplus U$ con $W = \ker h(T)$ y $U = \ker k(T)$, por la Proposición 2.37. El Lema 2.39 muestra que $q_T(t) = r(t)s(t)$, donde $r(t) = (t-\nu_1)^{l_1}$ y $s(t) = (t-\nu_2)^{l_2}\dots(t-\nu_r)^{l_r}$ con $l_i \leq k_i$ para cada i; además, $r(T)$ anula W y $s(T)$ anula U. El Corolario 2.40 muestra que $l_1 \geq 1$.

Como $r(t)$ divide $h(t)$, es inmediato que $W = \ker h(T) \subseteq \ker r(T) = W_1$. 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 $z \in W_1 \cap U$, entonces

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

Cualquier $\boldsymbol{x} \in V$ es de la forma $\boldsymbol{x} = \boldsymbol{y} + \boldsymbol{z}$ con $\boldsymbol{y} \in W$, $\boldsymbol{z} \in U$. Si $\boldsymbol{x} \in W_1$, entonces $\boldsymbol{y} \in W_1$ y $\boldsymbol{z} = \boldsymbol{x} - \boldsymbol{y}$ queda también en W_1 y por tanto $\boldsymbol{z} = \mathbf{0}$ y $\boldsymbol{x} = \boldsymbol{y}$. Se ha mostrado que $W_1 = W$.

Se concluye que $W_1 \oplus (W_2 \oplus \cdots \oplus W_r)$ es exactamente la descomposición de V que corresponde a la factorización $p_T(t) = h(t)\,k(t)$ por la Proposición 2.37. La Proposición 2.38 ahora muestra que el polinomio característico de T restringido a W_1 es $h(t) = (t-\nu_1)^{k_1}$. Además, el polinomio característico de T restringido a $W_2 \oplus \cdots \oplus W_r$ es $k(t) = (t-\nu_2)^{k_2}\dots(t-\nu_r)^{k_r}$, que es lo que se requiere para poder aplicar la hipótesis inductiva. □

► Ahora es posible ofrecer otro criterio de diagonalizabilidad.

Proposición 2.44. *Una matriz $A \in M_n(\mathbb{F})$ es diagonalizable si y sólo 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 (2.18), con factores distintos de primer grado.

Inversamente, si $q_A(t) = (t-\nu_1)(t-\nu_2)\dots(t-\nu_r)$ con $\nu_1, \dots, \nu_r$ distintos, el Teorema 2.42 muestra que $\mathbb{F}^n$ posee una descomposición primaria de la forma

$$\mathbb{F}^n = W_1 \oplus W_2 \oplus \cdots \oplus W_r, \quad \text{donde cada} \quad W_k := \ker(T_A - \nu_k I).$$

Sea $\mathcal{B}_k$ una base del subespacio $Z_A(\nu_k)$, 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 $A\boldsymbol{x} - \nu_k\boldsymbol{x} = (T_A - \nu_k I)(\boldsymbol{x}) = \mathbf{0}$, así que $A\boldsymbol{x} = \nu_k\boldsymbol{x}$: cada elemento de la base $\mathcal{B}$ es una *autovector* de A. La Proposición 2.34 muestra que A es diagonalizable. Concretamente, el cambio de la base estándar $\mathcal{E}$ a la base $\mathcal{B}$ diagonaliza la matriz A:

$$P^{-1}AP = [T_A]_{\mathcal{B}}^{\mathcal{B}} = \begin{bmatrix} \nu_1 & \dots & 0 & & & & & & & \\ \vdots & \ddots & \vdots & & & & & & & \\ 0 & \dots & \nu_1 & & & & & & & \\ & & & \nu_2 & \dots & 0 & & & & \\ & & & \vdots & \ddots & \vdots & & & & \\ & & & 0 & \dots & \nu_2 & & & & \\ & & & & & & \ddots & & & \\ & & & & & & & \nu_r & \dots & 0 \\ & & & & & & & \vdots & \ddots & \vdots \\ & & & & & & & 0 & \dots & \nu_r \end{bmatrix}$$ □

Ejemplo 2.45. 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 es $p_J(t) = t^2 + 1$, el cual es irreducible sobre $\mathbb{R}$. Como $q_J(t) \setminus p_J(t)$, también es $q_J(t) = t^2 + 1$, que no posee factores de primer grado sobre $\mathbb{R}$.

En cambio, si se toma $\mathbb{F} = \mathbb{C}$, la factorización $p_J(t) = (t - i)(t + i)$ muestra que $q_J(t) = (t - i)(t + i)$, ya que las posibilidades $q_J(t) = t \pm i$ quedan excluidas porque $J \pm iI \neq O$. Luego J es diagonalizable, con forma diagonal $\operatorname{diag}[i, -i]$, como ya se ha visto en el Ejemplo 2.13.

Ejemplo 2.46. Considérese la matriz triangular

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

con $\lambda \in \mathbb{F}$ cualquiera (y $\mathbb{F}$ un cuerpo cualquiera); esta matriz *no es diagonalizable*. En efecto, es $p_A(t) = (t - \lambda)^2$, 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 $A - \lambda I \neq O$; por otro lado, es $(A - \lambda I)^2 = O$ por el teorema de Cayley y Hamilton o bien por un cálculo directo.

Luego este polinomio $q_A(t) = (t - \lambda)^2$ no es un producto de factores *distintos* de primer grado.

2.5 La forma de Jordan de una matriz compleja

El Ejemplo 2.45 pone de manifiesto que la diagonalizabilidad de una matriz en $M_n(\mathbb{F})$ depende del cuerpo $\mathbb{F}$. El cuerpo de números complejos posee una propiedad fundamental, a veces llamado el Teorema Fundamental del Algebra:[17] *cualquier polinomio de grado n en $\mathbb{C}[t]$ posee n raíces complejas (no necesariamente distintas)*, o lo que es lo mismo, cualquier $f(t) \in \mathbb{C}[t]$ posee una factorización completa $f(t) = a_n(t - \alpha_1) \dots (t - \alpha_n)$, con $\alpha_1, \dots, \alpha_n \in \mathbb{C}$ no necesariamente distintos.

Para simplificar un poco la discusión, conviene suponer por ahora 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 2.42 de la descomposición primaria y su Corolario 2.43 muestran que cualquier operador lineal posee una matriz en $M_n(\mathbb{C})$ que es una suma directa de bloques diagonales:

$$[T]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} A_1 & O & \dots & O \\ O & A_2 & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & A_r \end{bmatrix}. \tag{2.23}$$

Es cuestión de elegir bases $\mathcal{B}_1, \dots, \mathcal{B}_r$ para los subespacios $W_1, \dots, W_r$ de la descomposición primaria y tomar la base $\mathcal{B}$ de V como su unión disjunta: $\mathcal{B} = \mathcal{B}_1 \uplus \dots \uplus \mathcal{B}_r$. Como cada

[17] Hay varias demostraciones de este teorema; parece que la primera demostración rigurosa fue dada por 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 $f(t) \in \mathbb{C}[t]$ no posee raíz alguna, entonces $z \mapsto f(z)$ es una función holomorfa acotada y por tanto constante, así que $\operatorname{gr} f(t) = 0$. Obsérvese que si $\operatorname{gr} f(t) \geq 1$, sólo hace falta que $f(t)$ tenga *una* raíz α, porque se puede considerar el cociente $f(t)/(t - \alpha)$ para obtener otra raíz, y así sucesivamente.

subespacio W_i reduce T, los bloques no diagonales son rectángulos de ceros. Cada A_i es una matriz con polinomio característico $(t-\nu_i)^{k_i}$.

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

$$N_i^{k_i} = (A_i - \nu_i I_{k_i})^{k_i} = O.$$

Definición 2.47. Un operador $T \in \operatorname{End}(V)$ es **nilpotente** si $T^k = 0$ para algún $k \in \{1,2,3,\dots\}$. Una matriz $A \in M_n(\mathbb{F})$ es una **matriz nilpotente** si $A^k = O$ para algún $k \in \{1,2,3,\dots\}$.

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

Proposición 2.48. *Cualquier matriz $A \in M_n(\mathbb{C})$ es de la forma $\underline{A = H + N}$, donde H es diagonalizable, N es nilpotente y $HN = NH$.*

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

$$p_A(t) = (t-\nu_1)^{k_1}(t-\nu_2)^{k_2}\dots(t-\nu_r)^{k_r},$$

donde $\nu_1,\dots,\nu_r \in \mathbb{C}$ 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 (2.23). Para simplificar, supóngase que la matriz A ya tiene esta forma. Sea N la matriz de bloques

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

Entonces para $k := \max\{k_1,\dots,k_r\}$ es $N^k = O$, o sea, N es nilpotente. Sea H la suma directa de bloques diagonales $\nu_i I_{k_i}$, la cual es una matriz diagonal con autovalores ν_i (repetidas con multiplicidades k_i). Es obvio que $A = H + N$. En cada bloque, el producto HN se reduce al producto de una matriz escalar $\nu_i I_{k_i}$ con una matriz $N_i \in M_{k_i}(\mathbb{C})$, de donde sigue $HN = NH$.

En el caso general, se reemplazan H por $P^{-1}HP$ y N por $P^{-1}NP$. Claramente, $P^{-1}HP$ es diagonalizable con forma diagonal H. La matriz $P^{-1}NP$ es nilpotente, porque

$$(P^{-1}NP)^k = P^{-1}NP\,P^{-1}NP\dots P^{-1}NP = P^{-1}N^kP = P^{-1}OP = O.$$

También es evidente que $(P^{-1}HP)(P^{-1}NP) = P^{-1}HNP = P^{-1}NHP = (P^{-1}NP)(P^{-1}HP)$. $\square$

Falta averiguar la estructura de una matriz nilpotente. Sea $N \in M_n(\mathbb{C})$ una matriz tal que $N^k = O$, $N^{k-1} \neq O$ para k un entero positivo (que depende de N). Entonces hay al menos un vector no nulo $\boldsymbol{x} \in \mathbb{C}^n$ tal que $N^{k-1}\boldsymbol{x} \neq \boldsymbol{0}$. Para cada $l = 0, 1, \dots, k$, considérese el subespacio

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

Entonces $V_0 = \{\boldsymbol{0}\}$, $V_k = \mathbb{C}^n$ y además $V_{l-1} \subseteq V_l$ para $l = 1, \dots, k$, porque $N^{l-1}\boldsymbol{x} = \boldsymbol{0}$ implica que $N^l\boldsymbol{x} = N(N^{l-1}\boldsymbol{x}) = N\boldsymbol{0} = \boldsymbol{0}$. De este modo, los V_l forman una *cadena de subespacios*:

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

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

$$0 = m_0 \leq m_1 \leq \ldots \leq m_{k-1} \leq m_k = n.$$

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

Lema 2.49. *Sea V un espacio vectorial de dimensión n sobre $\mathbb{F}$ y sea W un subespacio de V, con $\dim W = m$. Se puede elegir vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_{n-m} \in V$ que son **linealmente independientes sobre** W, es decir,*

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

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

Proposición 2.50. *Sea $N \in M_n(\mathbb{C})$ una matriz nilpotente. Si k es el menor entero positivo tal que $N^k = O$, 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 $N\boldsymbol{x} \in \mathcal{B}$ o bien $N\boldsymbol{x} = 0$.

Demostración. Denótese $r_1 := m_k - m_{k-1}$. Por el Lema 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 $N\boldsymbol{x}_1, \dots, N\boldsymbol{x}_{r_1}$ quedan en V_{k-1}, porque $N^{k-1}(N\boldsymbol{x}_j) = N^k\boldsymbol{x}_j = \boldsymbol{0}$ para cada j. Resulta que estos vectores son linealmente independientes sobre V_{k-2}. En efecto,

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

Por lo tanto, es $m_{k-1} - m_{k-2} \geq p_1$. Escríbase $r_2 := m_{k-1} - m_{k-2}$.

Si $r_2 > r_1$, en vista del Lema 2.49, se puede encontrar vectores $\boldsymbol{x}_{r_1+1}, \dots, \boldsymbol{x}_{r_2}$ tales que el conjunto $\{N\boldsymbol{x}_1, \dots, N\boldsymbol{x}_{r_1}, \boldsymbol{x}_{r_1+1}, \dots, \boldsymbol{x}_{r_2}\} \subset V_{k-1}$ sea 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-2}. De hecho,

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

Se concluye que $r_3 := m_{k-2} - m_{k-2}$ cumple $r_3 \geq r_2$. En el caso de que $r_3 > r_2$, hay vectores $\boldsymbol{x}_{r_2+1}, \dots, \boldsymbol{x}_{r_3}$ tales que el conjunto $\{N^2\boldsymbol{x}_1, \dots, N^2\boldsymbol{x}_{r_1}, N\boldsymbol{x}_{r_1+1}, \dots, N\boldsymbol{x}_{r_2}, \boldsymbol{x}_{r_2+1}, \dots, \boldsymbol{x}_{r_3}\} \subset V_{k-2}$ sea linealmente independiente sobre V_{k-3}.

Ad (a): Al repetir este proceso k veces, se obtiene la siguiente tabla de vectores en $\mathbb{C}^n$:

$$\begin{array}{ccccccccccccc} \boldsymbol{x}_1, & \dots, & \boldsymbol{x}_{r_1}, & & & & & & & & & & \\ N\boldsymbol{x}_1, & \dots, & N\boldsymbol{x}_{r_1}, & \boldsymbol{x}_{r_1+1}, & \dots, & \boldsymbol{x}_{r_2}, & & & & & & & \\ N^2\boldsymbol{x}_1, & \dots, & N^2\boldsymbol{x}_{r_1}, & N\boldsymbol{x}_{r_1+1}, & \dots, & N\boldsymbol{x}_{r_2}, & \boldsymbol{x}_{r_2+1}, & \dots, & \boldsymbol{x}_{r_3}, & & & & \\ \vdots & & \vdots & \vdots & & \vdots & \vdots & & \vdots & & & & \\ N^{k-1}\boldsymbol{x}_1, & \dots, & N^{k-1}\boldsymbol{x}_{r_1}, & N^{k-2}\boldsymbol{x}_{r_1+1}, & \dots, & N^{k-2}\boldsymbol{x}_{r_2}, & N^{k-3}\boldsymbol{x}_{r_2+1}, & \dots, & N^{k-3}\boldsymbol{x}_{r_3}, & \dots, & \boldsymbol{x}_{r_{k-1}+1}, & \dots, & \boldsymbol{x}_{r_k}. \end{array}$$

Aquí se ha escrito $r_j := m_{k-j+1} - m_{k-j}$ para $j = 1, \dots, k$. Los vectores en esta tabla son linealmente independientes, y las últimas l filas pertenecen a V_l para cada l. Ahora,

$$r_1 + r_2 + \cdots + r_k = (m_k - m_{k-1}) + (m_{k-1} - m_{k-2}) + \cdots + (m_2 - m_1) + m_1 = n,$$

así que todos estos vectores forman una **base** $\mathcal{B}$ de $\mathbb{C}^n$. Además, para cada l las sumas telescópicas $r_1 + \cdots + r_l = m_l = \dim V_l$ implican que las últimas l filas forman una base del subespacio V_l.

Ad (b): 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 tabla anterior de vectores, cada *columna* genera un subespacio invariante para T_N. De hecho, este subespacio *reduce* T_N porque las demás columnas generan un subespacio suplementario, también invariante. Se puede entonces reordenar la base $\mathcal{B}$, colocando las columnas de izquierda a derecha y dentro de cada columna leyendo las columnas de abajo hacia arriba:

$$\mathcal{B} = \{N^{k-1}\boldsymbol{x}_1, \dots, N\boldsymbol{x}_1, \boldsymbol{x}_1, N^{k-1}\boldsymbol{x}_2, \dots, N\boldsymbol{x}_2, \boldsymbol{x}_2, \dots, \boldsymbol{x}_{r_1}, N^{k-2}\boldsymbol{x}_{r_1+1}, \dots, \boldsymbol{x}_{r_2}, N^{k-3}\boldsymbol{x}_{r_2+1}, \dots, \boldsymbol{x}_{r_k}\}.$$

Las r_k columnas de la tabla determinan r_k subespacios que reducen T_N. Luego la matriz $[T_N]^{\mathcal{B}}_{\mathcal{B}}$ (que es semejante a N, desde luego) es una *suma directa de bloques*: hay r_1 bloques $k \times k$, seguido de $r_2 - r_1$ bloques de tamaño $(k-1) \times (k-1)$, etc., hasta $r_{k-1} - r_{k-2}$ bloques

2×2. Las últimas $r_k - r_{k-1}$ columnas de la tabla están en $V_1 = \ker T_N$ y se combinan para proporcionar un bloque de ceros en la esquina inferior derecha de esta matriz.

Cada uno de estos bloques tiene una estructura sencilla. Basta examinar el primer bloque, que queda 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 escribir $\boldsymbol{y}_j := N^{k-j}\boldsymbol{x}_1$, de modo que $\boldsymbol{y}_1,\dots,\boldsymbol{y}_k$ es la base ordenada del primer subespacio invariante de la lista, se obtiene

$$T_N(\boldsymbol{y}_1) = \boldsymbol{0}, \quad T_N(\boldsymbol{y}_2) = \boldsymbol{y}_1, \quad T_N(\boldsymbol{y}_3) = \boldsymbol{y}_2, \quad \dots \quad T_N(\boldsymbol{y}_k) = \boldsymbol{y}_{k-1},$$

y la matriz correspondiente es el bloque

$$J_k(0) := \begin{bmatrix} 0 & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & 0 & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 0 & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & 0 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix} \tag{2.24}$$

es decir, una matriz triangular con ceros en la diagonal y unos en la *subdiagonal superior*: $a_{12} = a_{23} = \cdots = a_{k-1,k} = 1$.

La matriz de T_N es la base elegida es entonces la suma directa de r_1 bloques $J_k(0)$, $(r_2 - r_1)$ bloques $J_{k-1}(0)$, etc., hasta $(r_{k-1} - r_{k-2})$ bloques $J_2(0)$, más un bloque cuadrado de ceros de lado $(r_k - r_{k-1})$.

Definición 2.51. Sea $k \in \{2,3,\dots\}$ y sea $\lambda \in \mathbb{C}$. El **bloque de Jordan** $J_k(\lambda)$ es la matriz triangular en $M_k(\mathbb{C})$ dado por

$$J_k(\lambda) := \lambda I_k + J_k(0) = \begin{bmatrix} \lambda & 1 & 0 & \dots & 0 & 0 & 0 \\ 0 & \lambda & 1 & \dots & 0 & 0 & 0 \\ 0 & 0 & \lambda & \dots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \lambda & 1 & 0 \\ 0 & 0 & 0 & \dots & 0 & \lambda & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & \lambda \end{bmatrix}. \tag{2.25}$$

Conviene denotar $J_1(\lambda) := [\lambda] \in M_1(\mathbb{C})$, como caso trivial.

Lema 2.52. *El polinomio mínimo de un bloque de Jordan es igual que 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 $(J_k(\lambda) - \lambda I_k)^l \neq O$ cuando $k \geq 2$ y $l < k$.

Ahora $J_k(\lambda) - \lambda I_k = J_k(0)$ es la matriz triangular nilpotente (2.24). 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

$$B = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad B^2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}, \qquad B^3 = \begin{bmatrix} 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}.$$

En particular, $B^l \neq O$ para $l = 1, 2, \dots, k-1$, pero $B^l = O$. Luego $q_B(t) = t^k$ y $q_A(t) = (t-\lambda)^k$. □

Teorema 2.53. *Cualquier matriz $A \in M_n(\mathbb{C})$ es semejante a una suma directa de bloques de Jordan de la forma $J_l(\nu_i)$, donde $\{\nu_1, \dots, \nu_r\}$ son los autovalores distintos de A; para cada ν_i, el mayor lado l de los bloques $J_l(\nu_i)$ en esta suma directa es el exponente l_i del factor $(t-\nu_i)^{l_i}$ 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 (2.21):

$$p_A(t) = (t-\nu_1)^{k_1}(t-\nu_2)^{k_2}\dots(t-\nu_r)^{k_r},$$

donde $\nu_1, \dots, \nu_r \in \mathbb{C}$ son distintos y $k_1 + \cdots + k_r = n$. Su polinomio mínimo tiene la forma (2.22):

$$q_A(t) = (t-\nu_1)^{l_1}(t-\nu_2)^{l_2}\dots(t-\nu_r)^{l_r},$$

donde $1 \leq l_i \leq k_i$ en cada caso, en vista del Corolario 2.43.

Sea $\mathbb{C}^n = W_1 \oplus \cdots \oplus W_r$ la descomposición primaria debida a esta factorización de $p_A(t)$. Elíjase una base $\mathcal{B}_i$ para cada W_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 obtiene que la matriz A es semejante a una suma directa $A = A_1 \oplus \cdots \oplus A_r$, como en (2.23). Además, cada bloque A_i tiene la forma $A_i = \nu_i I_{k_i} + N_i$ donde N_i es una matriz cuadrada nilpotente, con $N_i^{l_i} = O$ en $M_{k_i}(\mathbb{C})$.

Ahora, la matriz escalar $\nu_i I_{k_i}$ no sufre cambio alguno al reemplazar la base $\mathcal{B}_i$ por cualquier otra base de W_i. Se puede entonces suponer que $\mathcal{B}_i$ es aquélla que expresa T_{N_i} como suma directa de bloques de Jordan $J_l(0)$, como en (2.24). Por la construcción de esta base, se ve que $l \leq l_i$ en cada caso y que hay al menos un bloque de lado l_i. Al sumarles los bloques escalares $\nu_i I_l$, se obtiene que cada A_i es una suma directa de bloques de Jordan $J_l(\nu_i)$. □

Con el Teorema anterior, se dispone de una descripción completa de la estructura de una matriz cuadrada compleja, o bien la de un operador lineal sobre un espacio vectorial complejo finitodimensional. De hecho, la descripción es aplicable a matrices u operadores con otros cuerpos $\mathbb{F}$ de escalares, toda vez que sus polinomios característicos escinden en $\mathbb{F}[t]$. Hace falta, sin embargo, un proceso algorítmico para hallar los polinomios mínimos.[18]

[18]Tales procesos existen, pero quedan fuera del ámbito de este curso. Véase, por ejemplo, el libro de Anatoly I. Maltsev, *Fundamentos de Algebra Lineal*, Mir, Moscú, 1972.

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 λ_1, λ_2, λ_3 de la matriz

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

Resolver las ecuaciones $(\lambda_j I_3 - A)\boldsymbol{x}_j = \boldsymbol{0}$, $j = 1,2,3$, para obtener tres autovectores $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\}$ de A. Sea $P := [\boldsymbol{x}_1\ \boldsymbol{x}_2\ \boldsymbol{x}_3]$. Verificar que la matriz $P^{-1}AP$ es diagonal y que sus elementos diagonales son los autovalores de A.

Ejercicio 2.3. Calcular los autovalores de la matriz

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

Obtener una matriz inversible P cuyas columnas son autovectores de A y verificar que las transpuestas de las filas de P^{-1} son autovectores de A^t.

Ejercicio 2.4. 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 y de cada columna es la misma. Verificar que $\frac{1}{2}n(n^2+1)$ es un autovalor de esta matriz.

Ejercicio 2.5. 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 $-\pi < \theta \leq \pi$, $t \in \mathbb{R}$, y $\omega = e^{2\pi i/3} = \frac{1}{2}(-1 + i\sqrt{3})$.

Ejercicio 2.6. Calcular el polinomio característico de la matriz

$$A = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 & -a_0 \\ 1 & 0 & 0 & \dots & 0 & -a_1 \\ 0 & 1 & 0 & \dots & 0 & -a_2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1 & -a_{n-1} \end{bmatrix}$$

y concluir que todo polinomio $f(t)$ es el polinomio característico de alguna matriz.

Ejercicio 2.7. (a) Si $P^{-1}AP = D$ es una matriz diagonal, demostrar que $A^k = PD^kP^{-1}$ para todo $k \in \mathbb{N}$.

(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 2.8. Tómese $\mathbb{F} = \mathbb{R}$ ó $\mathbb{C}$. Las fórmulas $A^k = PD^kP^{-1}$ son casos particulares de la receta

$$f(A) = Pf(D)P^{-1} \quad \text{toda vez que} \quad D = P^{-1}AP,$$

la cual es válida para funciones f que pueden desarrollarse en series de potencias (con radio de convergencia infinita, digamos), 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: de hecho, es $f(D) = \operatorname{diag}[f(\lambda_1),\dots,f(\lambda_n)]$. Al tomar $f(t) := e^t = \sum_{k\geq 0}(1/k!)\,t^k$, se define la **exponencial** de una matriz $A \in M_n(\mathbb{F})$ como $\underline{\exp A} := \sum_{k\geq 0}(1/k!)\,A^k$. Comprobar el cálculo siguiente:

$$\exp\begin{bmatrix} -5 & 3 \\ -6 & 4 \end{bmatrix} = \begin{bmatrix} 2e^{-2} - e & e - e^{-2} \\ 2e^{-2} - 2e & 2e - e^{-2} \end{bmatrix}.$$

Ejercicio 2.9. Una **cadena de Markov** es un proceso probabilístico con un número finito n de *estados* caracterizado por números reales no negativos a_{ij} (que representa la probabilidad de cambiar del estado i al estado j en un 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 cambiar del estado i al estado j en k pasos es la entrada (i, j) de la matriz A^k. Comprobar que

$$A = \begin{bmatrix} \frac12 & \frac12 & 0 \\ \frac14 & \frac12 & \frac14 \\ 0 & \frac12 & \frac12 \end{bmatrix} \implies A^k = \begin{bmatrix} \frac14 + (\frac12)^{k+1} & \frac12 & \frac14 - (\frac12)^{k+1} \\ \frac14 & \frac12 & \frac14 \\ \frac14 - (\frac12)^{k+1} & \frac12 & \frac14 + (\frac12)^{k+1} \end{bmatrix}.$$

Ejercicio 2.10. Sea $A, B \in M_n(\mathbb{C})$ dos matrices cualesquiera. Demostrar que los polinomios característicos $p_{AB}(t)$ y $p_{BA}(t)$ coinciden.

⟦ Indicación: Si $\det A \neq 0$, es $BA = A^{-1}(AB)A$. Si $\det A = 0$, demostrar que $\det(A - \mu I_n) \neq 0$ para casi todo $\mu \in \mathbb{C}$ y concluir que para cada λ fijo, la expresión

$$t \mapsto \det(\lambda I_n - (A - tI_n)B) - \det(\lambda I - B(A - tI_n))$$

es un polinomio con más de n raíces. ⟧

Ejercicio 2.11. 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 2.12. Sea $S \in \operatorname{End}(M_2(\mathbb{F}))$ el *operador de transposición*, es decir $S(A) := A^t$ (véase el Ejercicio 1.14). Calcular los polinomios característico y mínimo de S. Exhibir una base de autovectores para el operador S.

Ejercicio 2.13. (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 $\operatorname{tr}(A^k) = \lambda_1^k + \cdots + \lambda_n^k$, para todo $k \in \mathbb{N}$.

Ejercicio 2.14. Sea $A \in M_n(\mathbb{F})$ una matriz inversible. Demostrar que el coeficiente de t en el polinomio característico $p_A(t)$ es $(-1)^{n-1} \det A \operatorname{tr}(A^{-1})$.

Ejercicio 2.15. Decimos que una matriz $B \in M_n(\mathbb{F})$ es *idempotente* si $B^2 = B$. Comprobar que la matriz $(I_n - B)$ es también idempotente. Demostrar que los autovalores distintos de B son $\{0, 1\}$, excepto si $B = O$ ó $B = I$.

¿Qué puede afirmarse acerca de la forma de Jordan de una matriz idempotente B?

Verificar que $r(B) = \operatorname{tr} B$ cuando B es idempotente.

Ejercicio 2.16. Una matriz $A \in M_n(\mathbb{F})$, sobre un cuerpo $\mathbb{F}$ cualquiera, se llama *semisimple* si su polinomio mínimo es un producto de *factores irreducibles distintos*. Comprobar que una matriz compleja (el caso $\mathbb{F} = \mathbb{C}$) es semisimple si y sólo si es diagonalizable.

Exhibir una matriz $B \in M_4(\mathbb{R})$ que es semisimple pero no diagonalizable.

Ejercicio 2.17. Si $A \in M_n(\mathbb{R})$ la matriz simétrica con entradas $a_{ij} = [\![i=j+1]\!] + [\![j=i+1]\!]$, es decir, tiene entradas 1 en las dos subdiagonales principales, y las demás entradas cero. Para $n = 5$, por ejemplo, es

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

Sea $B \in M_n(\mathbb{R})$ la matriz con entradas $b_{ij} = \operatorname{sen}\left(\dfrac{ij\pi}{n+1}\right)$. Verificar que las columnas de B son autovectores de A. ¿Cuáles son los autovalores correspondientes? es la matriz A diagonalizable?

Ejercicio 2.18. Calcular los polinomios característico y mínimo de la matriz

$$A = \begin{bmatrix} 5 & -1 & 0 & 0 & 0 \\ 6 & 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & -1 & 1 \end{bmatrix}.$$

Ejercicio 2.19. Calcular la forma 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 $(A - I_4)$, $(A - I_4)^2$, $(A - I_4)^3$, respectivamente.

(c) Hallar un vector $\boldsymbol{x} \in V_3 \setminus V_2$ tal que $\{(A - I_4)^2\boldsymbol{x}, (A - I_4)\boldsymbol{x}, \boldsymbol{x}, \boldsymbol{y}\}$ 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 2.20. 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

$$C(t) = Q(t)\,(tI_n - A) + C(A);$$

es decir, que es residuo de la "división a la derecha" de $C(t)$ por $(tI_n - A)$ es la matriz

$$C(A) := C_r A^r + C_{r-1}A^{r-1} + \cdots + C_1 A + C_0.$$

Ejercicio 2.21. 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 los menores $(n-1)\times(n-1)$ de A, esto es, el máximo común divisor de las entradas de $\operatorname{adj}(tI_n - A)$.

(a) Comprobar que $d_{n-1}(t)$ divide $p_A(t)$. Si

$$\tilde{q}(t) := \frac{p_A(t)}{d_{n-1}(t)},$$

verificar que $\tilde{q}(A) = O$ y como consecuencia, que $q_A(t)$ divide $\tilde{q}(t)$. ⟦ Indicación: Usar el ejercicio anterior. ⟧

(b) Si $\tilde{q}(t) = s(t)\,q_A(t)$ en $\mathbb{F}[t]$, demostrar que $s(t) \equiv 1$ y concluir que $q_A(t) = p_A(t)/d_{n-1}(t)$.[19]

[19]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*, Chelsea, New York, 1959.

3 Ortogonalidad y Teoría Espectral

Hasta ahora, el cuerpo subyacente a los espacios vectoriales y matrices ha sido arbitrario, y los conceptos principales han sido la independencia lineal de vectores y la semejanza de matrices cuadradas. (La única excepción a esta universalidad del cuerpo $\mathbb{F}$ de escalares ocurrió cuando fue necesario suponer que los polinomios característicos escinden en $\mathbb{F}[t]$ para obtener la forma de Jordan, en cuyo caso se tomó $\mathbb{F} = \mathbb{C}$ por comodidad.) De ahora en adelante, se adoptará un punto de vista más estrecho, porque los conceptos de *ortogonalidad* y *positividad* son más ligados al uso de *escalares reales o complejos*.

Así pues, en este capítulo el cuerpo de base será $\mathbb{R}$, el cuerpo de los números reales, o bien $\mathbb{C}$, el de los números complejos. Cuando una discusión se aplica en los dos casos, se usará la letra $\mathbb{F}$ para denotar $\mathbb{F} = \mathbb{R}$ ó $\mathbb{F} = \mathbb{C}$ indiferentemente.

Si $\alpha \in \mathbb{C}$, $\alpha = s + it$ con $s, t \in \mathbb{R}$, denótese por $\underline{\bar{\alpha} = s - it}$ su *conjugado complejo*; desde luego, es $\bar{\alpha} = \alpha$ si y sólo si $\alpha \in \mathbb{R}$.

3.1 Productos escalares reales y complejos

Definición 3.1. Sea V un espacio vectorial sobre $\mathbb{F} = \mathbb{R}$ ó $\mathbb{C}$. Un **producto escalar** en V es una operación que a cada par de vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ asocia un escalar $\langle \boldsymbol{x}, \boldsymbol{y} \rangle \in \mathbb{F}$, 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) $\langle \boldsymbol{x}, \boldsymbol{x} \rangle \geq 0$, con igualdad sólo 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

$$\langle \boldsymbol{x}, \boldsymbol{y} \rangle \equiv \boldsymbol{x} \cdot \boldsymbol{y} := x_1 y_1 + x_2 y_2 + \cdots + x_n y_n.$$

Esto es un producto escalar real.

Si $\mathbb{F} = \mathbb{C}$ y $V = \mathbb{C}^n$, se define análogamente

$$\langle \boldsymbol{z}, \boldsymbol{w} \rangle \equiv \bar{\boldsymbol{z}} \cdot \boldsymbol{w} := \bar{z}_1 w_1 + \bar{z}_2 w_2 + \cdots + \bar{z}_n w_n.$$

Esto es un producto escalar complejo.

[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*. De hecho, en la literatura matemática, abundan los productos internos y externos, como también los productos interiores y exteriores. Para no complicar las cosas antes de tiempo, es mejor evitar esta terminología.

Definición 3.3. Una aplicación $T\colon V \to W$ entre dos espacios vectoriales complejas (en el caso $\mathbb{F} = \mathbb{C}$) se llama **semilineal** (o bien *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

- $\boldsymbol{x} \mapsto T(\boldsymbol{x}, \boldsymbol{y})$ 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 $\boldsymbol{x} \in V$.

Es decir, T es bilineal si es *lineal en cada variable por separado*.

Si V, W, Z son tres espacios vectoriales complejos, 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* $\mathbb{F} = \mathbb{C}$. (Fíjese 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, hay que advertir que la mayoría de los textos de 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 de 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 esta es un producto escalar real.

En el espacio vectorial complejo $C_{\mathbb{C}}[a, b]$ de 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, $\langle f, f \rangle = \int_a^b |f(t)|^2\, dt \geq 0$, con igualdad si y sólo si la función continua f es idénticamente cero en el intervalo $[a, b]$.

[2] El prefijo *sesqui-* significa "$1\frac{1}{2}$ veces".

[3] *Paul Adrien Maurice Dirac* (1902–1984), físico inglés, obtuvo una ecuación que describe el comportamiento relativista del electrón. En 1930, publicó su *Principles of Quantum Mechanics*, que sentó el formalismo básico de la física cuántica (e incluye sus convenios notacionales).

Los espacios vectoriales del ejemplo anterior son infinitodimensionales. Para evitar las matices de análisis tales como la convergencia de integrales y series, en adelante se asumirá que todos los espacios vectoriales son de dimensión finita. Sin embargo, buena parte de la discusión 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

$$\langle A, B\rangle := \operatorname{tr}(A^t B).$$

Esta expresión es evidentemente lineal en A y en B. Para verificar su positividad, es cuestión de notar que

$$\operatorname{tr}(A^t A) = \sum_{j=1}^{n}\sum_{i=1}^{m} a_{ij}^2 \geq 0,$$

con igualdad si y sólo si todo $a_{ij} = 0$, es decir, $A = O$.

La positividad del producto escalar, en la propiedad (d) de la Definición 3.1, permite introducir el concepto de longitud de un vector. En adelante, en este capítulo, V será un espacio vectorial finitodimensional, real o complejo según el contexto, dotado de un producto escalar $\langle\cdot,\cdot\rangle$ fijo.

Definición 3.6. Se define la **norma** (o *longitud*) de un vector $\boldsymbol{x} \in V$ por

$$\|\boldsymbol{x}\| := \sqrt{\langle \boldsymbol{x}, \boldsymbol{x}\rangle}. \tag{3.1}$$

Proposición 3.7. *Se verifica la* ***desigualdad de Schwarz:***

$$\bigl|\langle \boldsymbol{x}, \boldsymbol{y}\rangle\bigr| \leq \|\boldsymbol{x}\|\,\|\boldsymbol{y}\| \quad \textit{para todo} \quad \boldsymbol{x}, \boldsymbol{y} \in V. \tag{3.2}$$

con igualdad si y sólo si $\boldsymbol{x}, \boldsymbol{y}$ son proporcionales.

Demostración. Es claro que (3.2) se cumple con igualdad si $\boldsymbol{x} = \boldsymbol{0}$ o bien si $\boldsymbol{y} = \boldsymbol{0}$ (el vector $\boldsymbol{0}$ es proporcional a cualquier vector $\boldsymbol{x}$ porque $\boldsymbol{0} = 0\boldsymbol{x}$). Supóngase entonces que $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{y} \neq \boldsymbol{0}$. Supóngase además, en el caso $\mathbb{F} = \mathbb{C}$, que $\langle \boldsymbol{x}, \boldsymbol{y}\rangle \in \mathbb{R}$.

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 $f(t) \geq 0$ 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 $b^2 - 4ac \leq 0$, es decir,

$$4\langle \boldsymbol{x}, \boldsymbol{y}\rangle^2 - 4\|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2 \leq 0,$$

o bien $\langle \boldsymbol{x},\boldsymbol{y}\rangle^2 \le \|\boldsymbol{x}\|^2\|\boldsymbol{y}\|^2$. Al tomar la raíz cuadrada positiva de ambos lados, se obtiene la desigualdad deseada:

$$\left|\langle \boldsymbol{x},\boldsymbol{y}\rangle\right| \le \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|,$$

con igualdad sólo si $b^2-4ac=0$, es decir, sólo 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 en consecuencia $\boldsymbol{x}+t_0\boldsymbol{y}=\boldsymbol{0}$ 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.

En el caso $\mathbb{F}=\mathbb{C}$, si $\langle \boldsymbol{x},\boldsymbol{y}\rangle\notin\mathbb{R}$, entonces $\langle \boldsymbol{x},\boldsymbol{y}\rangle=r\,e^{i\theta}$ con $r>0$, $\theta\in\mathbb{R}$. Colóquese $\boldsymbol{z}:=e^{-i\theta}\boldsymbol{y}$, así que $\langle \boldsymbol{x},\boldsymbol{z}\rangle=r\in\mathbb{R}$. De (3.2), ya verificado para ese caso, se obtiene $|\langle \boldsymbol{x},\boldsymbol{z}\rangle|\le\|\boldsymbol{x}\|\,\|\boldsymbol{z}\|$; pero $\|\boldsymbol{z}\|=|e^{-i\theta}|\,\|\boldsymbol{y}\|=\|\boldsymbol{y}\|$, así que

$$\left|\langle \boldsymbol{x},\boldsymbol{y}\rangle\right| = \left|e^{i\theta}\langle \boldsymbol{x},\boldsymbol{z}\rangle\right| = \left|\langle \boldsymbol{x},\boldsymbol{z}\rangle\right| \le \|\boldsymbol{x}\|\,\|\boldsymbol{z}\| = \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|. \qquad \square$$

Corolario 3.8 (Desigualdad de Cauchy)**.** *Si $x_1,\dots,x_n,y_1,\dots,y_n\in\mathbb{R}$, entonces*

$$(x_1y_1+\cdots+x_ny_n)^2 \le (x_1^2+\cdots+x_n^2)(y_1^2+\cdots+y_n^2). \qquad \boxminus$$

Proposición 3.9. *La norma de un vector tiene las siguientes propiedades; si $\boldsymbol{x},\boldsymbol{y}\in V$, $\alpha\in\mathbb{F}$, entonces:*

(a) $\|\alpha\boldsymbol{x}\|=|\alpha|\,\|\boldsymbol{x}\|$ *(homogeneidad positiva),*

(b) $\|\boldsymbol{x}+\boldsymbol{y}\|\le\|\boldsymbol{x}\|+\|\boldsymbol{y}\|$ *(desigualdad triangular),*

(c) $\|\boldsymbol{x}\|\ge 0$, *con igualdad sólo si $\boldsymbol{x}=\boldsymbol{0}$ en V.*

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 que $\sqrt{\bar\alpha\alpha}=|\alpha|$.

Ad (b): Al usar la desigualdad de Schwarz (3.2), se obtiene

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|^2 &= \langle \boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}+\boldsymbol{y}\rangle^2\\ &= \langle \boldsymbol{x},\boldsymbol{x}\rangle^2+\langle \boldsymbol{x},\boldsymbol{y}\rangle+\langle \boldsymbol{y},\boldsymbol{x}\rangle+\langle \boldsymbol{y},\boldsymbol{y}\rangle^2\\ &\le \|\boldsymbol{x}\|^2+2|\langle \boldsymbol{x},\boldsymbol{y}\rangle|+\|\boldsymbol{y}\|^2\\ &\le \|\boldsymbol{x}\|^2+2\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|+\|\boldsymbol{y}\|^2=(\|\boldsymbol{x}\|+\|\boldsymbol{y}\|)^2.\end{aligned}$$

Ad (c): Esto es inmediato de la propiedad (d) de la Definición 3.1. $\square$

Lema 3.10 (Ley del paralelogramo)**.** *Si V es un espacio vectorial con un producto escalar y si $\boldsymbol{x},\boldsymbol{y}y\in V$, entonces*

$$\|\boldsymbol{x}+\boldsymbol{y}\|^2+\|\boldsymbol{x}-\boldsymbol{y}\|^2=2\|\boldsymbol{x}\|^2+2\|\boldsymbol{y}\|^2.$$

Demostración. Este es un cálculo sencillo:

$$\begin{aligned}\|\boldsymbol{x}+\boldsymbol{y}\|^2+\|\boldsymbol{x}-\boldsymbol{y}\|^2 &= \langle \boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}+\boldsymbol{y}\rangle+\langle \boldsymbol{x}-\boldsymbol{y},\boldsymbol{x}-\boldsymbol{y}\rangle\\ &= 2\langle \boldsymbol{x},\boldsymbol{x}\rangle+2\langle \boldsymbol{y},\boldsymbol{y}\rangle=2\|\boldsymbol{x}\|^2+2\|\boldsymbol{y}\|^2. \qquad \square\end{aligned}$$

Hay otras normas sobre $\mathbb{R}^n$ ó $\mathbb{C}^n$, es decir, funciones $\boldsymbol{x} \mapsto \|\boldsymbol{x}\|$ que cumplen las propiedades (a), (b), (c) de la Proposición 3.9. Dos ejemplos son

$$\|\boldsymbol{x}\|_1 := |x_1| + |x_2| + \cdots + |x_n|,$$
$$\|\boldsymbol{x}\|_\infty := \max\{|x_1|, |x_2|, \dots, |x_n|\}. \tag{3.3}$$

No es difícil chequear que estas normas *no* cumplen la ley del paralelogramo. De hecho, se sabe que cualquier norma sobre $\mathbb{R}^n$ ó $\mathbb{C}^n$ que cumple esa ley determina un producto escalar tal que $\langle \boldsymbol{x}, \boldsymbol{x} \rangle = \|\boldsymbol{x}\|^2$.

Lema 3.11 (Polarización)**.** *Si V es un espacio vectorial con un producto escalar, la norma* (3.1) *determina el producto escalar por* polarización*:*

$$\textit{Caso } \mathbb{F} = \mathbb{R}: \quad \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \tfrac{1}{4}\|\boldsymbol{x} + \boldsymbol{y}\|^2 - \tfrac{1}{4}\|\boldsymbol{x} - \boldsymbol{y}\|^2.$$
$$\textit{Caso } \mathbb{F} = \mathbb{C}: \quad \langle \boldsymbol{x}, \boldsymbol{y} \rangle = \tfrac{1}{4}\|\boldsymbol{x} + \boldsymbol{y}\|^2 - \tfrac{1}{4}\|\boldsymbol{x} - \boldsymbol{y}\|^2 + \tfrac{i}{4}\|i\boldsymbol{x} + \boldsymbol{y}\|^2 - \tfrac{i}{4}\|i\boldsymbol{x} - \boldsymbol{y}\|^2.$$

Demostración. En ambos casos, se obtiene $\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ al hacer una expansión directa del lado derecho de la ecuación. □

Definición 3.12. La longitud $\|\boldsymbol{x} - \boldsymbol{y}\|$ se llama la **distancia** entre dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$.

En el caso $\mathbb{F} = \mathbb{R}$, si $\boldsymbol{x} \neq \boldsymbol{0}$ y $\boldsymbol{y} \neq \boldsymbol{0}$, la desigualdad de Schwarz implica que

$$-1 \leq \frac{\langle \boldsymbol{x}, \boldsymbol{y} \rangle}{\|\boldsymbol{x}\|\,\|\boldsymbol{y}\|} \leq 1,$$

así que hay un único ángulo θ con $0 \leq \theta \leq \pi$ tal que $\cos\theta := \langle \boldsymbol{x}, \boldsymbol{y} \rangle / \|\boldsymbol{x}\|\,\|\boldsymbol{y}\|$. Se dice que θ es el *ángulo entre los vectores* no nulos $\boldsymbol{x}, \boldsymbol{y}$. Se verifica la relación $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = \|\boldsymbol{x}\|\,\|\boldsymbol{y}\| \cos\theta$, pero es tautológica.

Definición 3.13. Dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ son **ortogonales** si $\langle \boldsymbol{x}, \boldsymbol{y} \rangle = 0$. Se escribe $\boldsymbol{x} \perp \boldsymbol{y}$ para significar que $\boldsymbol{x}, \boldsymbol{y}$ son ortogonales.

Se dice que los vectores no nulos $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in V$ forman un **conjunto ortogonal** si $\boldsymbol{x}_i \perp \boldsymbol{x}_j$ para $i \neq j$.

Lema 3.14. *Un conjunto ortogonal de vectores es linealmente independiente.*

Demostración. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ un conjunto ortogonal de vectores. Sean $c_1, \dots, c_m \in \mathbb{F}$ tales que $c_1\boldsymbol{x}_1 + \cdots + c_m\boldsymbol{x}_m = \boldsymbol{0}$. Para cada índice $j = 1, \dots, m$, vale

$$0 = \langle \boldsymbol{x}_j, \boldsymbol{0} \rangle = \langle \boldsymbol{x}_j, c_1\boldsymbol{x}_1 + \cdots + c_m\boldsymbol{x}_m \rangle = \langle \boldsymbol{x}_j, c_j\boldsymbol{x}_j \rangle = c_j\|\boldsymbol{x}_j\|^2,$$

y por tanto $c_j = 0$. Luego $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$ es linealmente independiente. □

Definición 3.15. Si M es un subespacio de V, su **complemento ortogonal** $M^\perp$ es el subespacio de V definido por

$$M^\perp := \{\boldsymbol{y} \in V : \langle \boldsymbol{y}, \boldsymbol{x} \rangle = 0 \text{ para todo } \boldsymbol{x} \in M\}.$$

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

$$\langle \boldsymbol{y} + \boldsymbol{z}, \boldsymbol{x} \rangle = \langle \boldsymbol{y}, \boldsymbol{x} \rangle + \langle \boldsymbol{z}, \boldsymbol{x} \rangle = 0 + 0 = 0,$$
$$\langle \alpha\boldsymbol{y}, \boldsymbol{x} \rangle = \bar{\alpha}\langle \boldsymbol{y}, \boldsymbol{x} \rangle = \bar{\alpha}0 = 0.$$

3.2 Bases ortonormales

Definición 3.16. Se dice que una base $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de V es una **base ortonormal** 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:

$$\langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = [\![j = k]\!] \quad \text{para todo} \quad j,k = 1, \dots, n. \tag{3.4}$$

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.

Si V posee una base ortonormal $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$, sean $\boldsymbol{x} = \sum_{j=1}^n x_j \boldsymbol{e}_j$, $\boldsymbol{y} = \sum_{k=1}^n y_k \boldsymbol{e}_k$ las expresiones de dos vectores $\boldsymbol{x}, \boldsymbol{y} \in V$ en términos de esta base. Entonces

$$\begin{aligned} \langle \boldsymbol{x}, \boldsymbol{y} \rangle &= \Big\langle \sum_{j=1}^n x_j \boldsymbol{e}_j, \sum_{k=1}^n y_k \boldsymbol{e}_k \Big\rangle = \sum_{j=1}^n \sum_{k=1}^n \langle x_j \boldsymbol{e}_j, y_k \boldsymbol{e}_k \rangle \\ &= \sum_{j=1}^n \sum_{k=1}^n \bar{x}_j y_k \langle \boldsymbol{e}_j, \boldsymbol{e}_k \rangle = \sum_{k=1}^n \bar{x}_k y_k = \bar{x}_1 y_1 + \bar{x}_2 y_2 + \dots + \bar{x}_n y_n, \end{aligned}$$

porque los términos de la doble suma con $j \neq k$ se anulan. De este modo, se recupera la forma explícita del Ejemplo 3.2 para el producto escalar al usar *coordenadas respecto de 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

$$f_k(\boldsymbol{x}) := \langle \boldsymbol{e}_k, \boldsymbol{x} \rangle \quad \text{para todo} \quad k = 1, \dots, n.$$

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 $J(\sum_{k=1}^n y_k \boldsymbol{e}_k) := \sum_{k=1}^n \bar{y}_k f_k$. Por un análogo de la Proposición 1.14, la aplicación semilineal J queda determinado 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

$$J(\boldsymbol{y}) = \sum_{k=1}^n \bar{y}_k f_k, \qquad J(\boldsymbol{y})(\boldsymbol{x}) = \sum_{k=1}^n \bar{y}_k f_k(\boldsymbol{x}) = \sum_{k=1}^n \bar{y}_k x_k = \langle \boldsymbol{y}, \boldsymbol{x} \rangle. \tag{3.5}$$

Resulta, entonces, que $J(\boldsymbol{y})$ es la forma *lineal* $\underline{\boldsymbol{x} \mapsto \langle \boldsymbol{y}, \boldsymbol{x} \rangle}$. El resultado 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* (3.5) *de V^* con V, el complemento ortogonal $M^\perp$ de un subespacio $M \subseteq V$ coincide con el* anulador $M^\perp \leq V^*$ *de la Definición 1.21.*

En consecuencia, $\dim(M^\perp) = \dim V - \dim M$.

Demostración. Es cuestión de notar que

$$\begin{aligned} J(\boldsymbol{y}) \in M^\perp \text{ (anulador)} &\iff J(\boldsymbol{y})(\boldsymbol{x}) = 0 \quad \text{para todo } \boldsymbol{x} \in M \\ &\iff \langle \boldsymbol{y}, \boldsymbol{x} \rangle = 0 \quad \text{para todo } \boldsymbol{x} \in M \\ &\iff \boldsymbol{y} \in M^\perp \text{ (complemento ortogonal).} \end{aligned}$$

□

► 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*.[4]

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. Sea $\boldsymbol{e}_1 := \boldsymbol{x}_1/\|\boldsymbol{x}_1\|$; enseguida defínase, para $k \leq n$,*

$$\begin{aligned} \boldsymbol{y}_2 &:= \boldsymbol{x}_2 - \langle \boldsymbol{e}_1, \boldsymbol{x}_2 \rangle \boldsymbol{e}_1, & \boldsymbol{e}_2 &:= \boldsymbol{y}_2/\|\boldsymbol{y}_2\|; \\ \boldsymbol{y}_3 &:= \boldsymbol{x}_3 - \langle \boldsymbol{e}_1, \boldsymbol{x}_3 \rangle \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_3 \rangle \boldsymbol{e}_2, & \boldsymbol{e}_3 &:= \boldsymbol{y}_3/\|\boldsymbol{y}_3\|; \\ &\vdots & &\vdots \\ \boldsymbol{y}_k &:= \boldsymbol{x}_k - \langle \boldsymbol{e}_1, \boldsymbol{x}_k \rangle \boldsymbol{e}_1 - \langle \boldsymbol{e}_2, \boldsymbol{x}_k \rangle \boldsymbol{e}_2 - \cdots - \langle \boldsymbol{e}_{k-1}, \boldsymbol{x}_k \rangle \boldsymbol{e}_{k-1}, & \boldsymbol{e}_k &:= \boldsymbol{y}_k/\|\boldsymbol{y}_k\|. \end{aligned} \tag{3.6}$$

Entonces $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_k\}$ es una base ortonormal del subespacio $\operatorname{lin}\langle \boldsymbol{x}_1, \dots, \boldsymbol{x}_k \rangle$*; en particular, $\mathcal{E} := \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ es una base ortonormal de V.*

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 $\boldsymbol{e}_1, \dots, \boldsymbol{e}_{k-1}$ han sido elegidos por el procedimiento indicado, y que 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

$$\begin{aligned} \langle \boldsymbol{e}_j, \boldsymbol{y}_k \rangle = \Big\langle \boldsymbol{e}_j, \boldsymbol{x}_k - \sum_{i=1}^{k-1} \langle \boldsymbol{e}_i, \boldsymbol{x}_k \rangle \boldsymbol{e}_i \Big\rangle &= \langle \boldsymbol{e}_j, \boldsymbol{x}_k \rangle - \sum_{i=1}^{k-1} \langle \boldsymbol{e}_i, \boldsymbol{x}_k \rangle \langle \boldsymbol{e}_j, \boldsymbol{e}_i \rangle \\ &= \langle \boldsymbol{e}_j, \boldsymbol{x}_k \rangle - \sum_{i=1}^{k-1} \langle \boldsymbol{e}_i, \boldsymbol{x}_k \rangle [\![j=i]\!] = \langle \boldsymbol{e}_j, \boldsymbol{x}_k \rangle - \langle \boldsymbol{e}_j, \boldsymbol{x}_k \rangle = 0. \end{aligned}$$

[4]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*, 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.

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 $\boldsymbol{y}_k$ fuera $\boldsymbol{0}$, por (3.6) $\boldsymbol{x}_k$ sería una combinación lineal de $\boldsymbol{e}_1,\dots,\boldsymbol{e}_{k-1}$, los cuales son a su vez combinaciones lineales de $\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k-1}$. Luego $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k$ serían linealmente dependientes, que 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, tiene norma 1 y 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 de V puede completarse para obtener 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 la Proposición 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 $\boldsymbol{e}_1,\dots,\boldsymbol{e}_m$ originales. □

3.3 Matrices ortogonales, unitarias y positivas

La *transpuesta* de una aplicación lineal $T \in \mathcal{L}(V,W)$ es, según la Definición 1.18, la aplicación lineal $T^t \in \mathcal{L}(W^*,V^*)$ dada por $T^t(g) := g \circ T$, para todo $g \in W^*$. Si V y W son espacios vectoriales *reales*, dotados con productos escalares, las identificaciones canónicas $J_V \colon V \to V^*$ y $J_W \colon W \to W^*$ son $\mathbb{R}$-lineales permiten reemplazar $T^t \colon W^* \to V^*$ por una aplicación lineal de W en V. En efecto, si $\boldsymbol{y} \in W$ y $\boldsymbol{x} \in V$, se ve que

$$\begin{aligned}\langle J_V^{-1} T^t J_W(\boldsymbol{y}), \boldsymbol{x}\rangle_V &= T^t J_W(\boldsymbol{y})(\boldsymbol{x}) \\ &= J_W(\boldsymbol{y})(T(\boldsymbol{x})) = \langle \boldsymbol{y}, T(\boldsymbol{x})\rangle_W.\end{aligned}$$

En este cálculo se ha etiquetado las productos escalares y las identificaciones J con subíndices V y W, para énfasis; pero es usual omitir estos subíndices por comodidad.

Definición 3.20. En la presencia de productos escalares para los espacios vectoriales V y W sobre $\mathbb{R}$, se identifica la aplicación $T^t \colon W^* \to V^*$ con la aplicación $J^{-1}T^tJ \colon W \to V$. Así, se puede *redefinir* la transpuesta de $T \colon V \to W$ como el operador $T^t \colon W \to V$ determinado por la fórmula

$$\langle T^t(\boldsymbol{y}), \boldsymbol{x}\rangle := \langle \boldsymbol{y}, T(\boldsymbol{x})\rangle.$$

Para el caso $V = \mathbb{R}^n$, $W = \mathbb{R}^m$, es más cómodo usar la notación del producto punto en vez de los corchetes angulares para denotar los productos escalares. Los vectores en el 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*,[5] es decir, los matrices $1 \times n$. De ahora en adelante, *se escribirá $\boldsymbol{x}^t$ para denotar el vector de fila* que es la transpuesta del

[5] Algunos textos franceses escriben $\mathbb{R}_n$, en lugar de $(\mathbb{R}^n)^*$, para denotar el espacio dual de $\mathbb{R}^n$.

vector de columna $\boldsymbol{x}$. El producto escalar se convierte en $\underline{\boldsymbol{x}^t\boldsymbol{y}} = \boldsymbol{x}\cdot\boldsymbol{y}$. La simetría del producto escalar dice que

$$\boldsymbol{x}^t\boldsymbol{y} = \boldsymbol{x}\cdot\boldsymbol{y} = \boldsymbol{y}\cdot\boldsymbol{x} = \boldsymbol{y}^t\boldsymbol{x}.$$

La base estándar de $\mathbb{R}^n$ es una base ortonormal, con respecto al producto punto de $\mathbb{R}^n$. Si $A \in \mathbb{R}^{m\times n}$ es la matriz de $T \in \mathcal{L}(\mathbb{R}^n, \mathbb{R}^m)$ con respecto a las bases estándares, la fórmula que define la aplicación transpuesta es entonces equivalente a la fórmula matricial:

$$\boldsymbol{x}\cdot A^t\boldsymbol{y} = \boldsymbol{x}^t A^t\boldsymbol{y} = (A^t\boldsymbol{y})^t\boldsymbol{x} = \boldsymbol{y}^t A\boldsymbol{x} = \boldsymbol{y}\cdot A\boldsymbol{x}. \tag{3.7}$$

► Para espacios vectoriales *complejos* (es decir, cuando $\mathbb{F} = \mathbb{C}$), hay que tomar en cuenta que las identificaciones canónicas $J_V : V \to V^*$ y $J_W : W \to W^*$ no son $\mathbb{C}$-lineales, sino *semilineales*. Esta circunstancia obliga a un cambio de notación.

Definición 3.21. Sean V y W espacios vectoriales sobre $\mathbb{C}$. Si $T \in \mathcal{L}(V,W)$, se define la **aplicación adjunta** $T^* \in \mathcal{L}(W,V)$ por

$$T^* := J_V^{-1} T^t J_W \in \mathcal{L}(W,V). \tag{3.8}$$

(Fíjese que la composición de dos aplicaciones semilineales es una aplicación lineal.) Si $\boldsymbol{y} \in W$ y $\boldsymbol{x} \in V$, se ve que

$$\langle J_V^{-1} T^t J_W(\boldsymbol{y}), \boldsymbol{x}\rangle_V = T^t J_W(\boldsymbol{y})(\boldsymbol{x}) = J_W(\boldsymbol{y})(T(\boldsymbol{x})) = \langle \boldsymbol{y}, T(\boldsymbol{x})\rangle_W,$$

al igual que en el caso real. La aplicación adjunta $T^* \in \mathcal{L}(W,V)$ queda determinada por la fórmula

$$\boxed{\langle T^*(\boldsymbol{y}), \boldsymbol{x}\rangle := \langle \boldsymbol{y}, T(\boldsymbol{x})\rangle,} \tag{3.9}$$

en donde se suprimen los índices de los productos escalares.

En cálculos prácticos, se puede mover una aplicación lineal de un lado a otro de un producto escalar, reemplazándola por su aplicación adjunta al otro lado.

Sean $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ y $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_m\}$ unas bases ortonormales para V y W, respectivamente. Sean $A = [T]^{\mathcal{U}}_{\mathcal{E}}$ y $B = [T^*]^{\mathcal{E}}_{\mathcal{U}}$ las matrices correspondientes. De acuerdo con la fórmula (1.6), se ve que

$$\begin{aligned} T(\boldsymbol{e}_j) &= \sum_{i=1}^m a_{ij}\boldsymbol{u}_i, \quad \text{así que} \quad a_{ij} = \langle \boldsymbol{u}_i, T(\boldsymbol{e}_j)\rangle, \\ T^*(\boldsymbol{u}_s) &= \sum_{r=1}^n b_{rs}\boldsymbol{e}_r, \quad \text{así que} \quad b_{rs} = \langle \boldsymbol{e}_r, T^*(\boldsymbol{u}_s)\rangle, \end{aligned}$$

para todo $j = 1, \dots, m$ y $s = 1, \dots, n$. Por lo tanto,

$$b_{rs} := \langle \boldsymbol{e}_r, T^*(\boldsymbol{u}_s)\rangle = \langle T(\boldsymbol{e}_r), \boldsymbol{u}_s\rangle = \overline{\langle \boldsymbol{u}_s, T(\boldsymbol{e}_r)\rangle} = \bar{a}_{sr}.$$

Es decir, además de transponer la matriz A hay que tomar el conjugado complejo de cada uno de sus elementos.

Definición 3.22. Sea $A = [a_{ij}] \in \mathbb{C}^{m\times n}$ una matriz compleja $m \times n$. Se denota por $\overline{A} := [\bar{a}_{ij}]$ su **matriz conjugada**, obtenida de A al tomar el conjugado complejo de cada uno de sus elementos. Obsérvese que $\overline{A} = A$ si y sólo si todas las entradas de A son reales.

Se denota por $A^* := [\bar{a}_{ji}] \in \mathbb{C}^{n\times m}$ la **matriz adjunta** de A, la cual es la matriz transpuesta de $\overline{A}$ o equivalentemente, la matriz conjugada de A^t:

$$A^* := (\overline{A})^t = \overline{A^t}.$$

Fíjese que las aplicaciones $A \mapsto \overline{A}$ 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.

Definición 3.23. Sea $A \in M_n(\mathbb{F})$ una matriz cuadrada sobre un cuerpo cualquiera. se dice que A es una **matriz simétrica** si $A^t = A$.

En el caso complejo, se dice que una matriz cuadrada $A \in M_n(\mathbb{C})$ es una **matriz hermítica** si $A^* = A$. Una matriz simétrica real es también hermítica.[6]

Ejemplo 3.24. Si $A \in \mathbb{R}^{m\times n}$, las matrices cuadradas $A^tA \in M_n(\mathbb{R})$ y $AA^t \in M_m(\mathbb{R})$ son simétricas. En efecto,

$$(A^tA)^t = A^tA^{tt} = A^tA, \qquad (AA^t)^t = A^{tt}A^t = AA^t.$$

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,

$$(B^*B)^* = B^*B^{**} = B^*B, \qquad (BB^*)^* = B^{**}B^* = BB^*.$$

Para las matrices hermíticas, en particular para las matrices reales simétricas, los *autovectores* linealmente independientes que corresponden a autovalores distintos son de hecho *ortogonales*.

Proposición 3.25. *Sea $A \in M_n(\mathbb{R})$ una matriz real simétrica, o bien sea $A \in M_n(\mathbb{C})$ una matriz hermítica. Entonces sus autovalores* son todos reales. *Además, si x_1, x_2 son autovectores de A que corresponden a autovalores distintos, entonces $\langle x_1, x_2\rangle = 0$.*

Si A posee n autovalores distintos $\{\lambda_1,\dots,\lambda_n\}$, sean $\{x_1,\dots,x_n\}$ unos autovectores de A tales que $Ax_j = \lambda_j x_j$ y además $\|x_j\| = 1$, para $j = 1,\dots,n$. Estos autovectores $\{x_1,\dots,x_n\}$ entonces forman una base ortonormal de $\mathbb{R}^n$ ó $\mathbb{C}^n$.

Demostración. Es suficiente considerar el caso $\mathbb{F} = \mathbb{C}$, en donde la matriz A cumple $A^* = A$. En términos del producto escalar estándar en $\mathbb{C}^n$, vale $\langle Ay, x\rangle = \langle y, Ax\rangle$ para todo $x, yy \in \mathbb{C}^n$. Si $x \in \mathbb{C}^n$ es un autovector para el autovalor λ de A, entonces

$$\begin{aligned} \lambda\|x\|^2 &= \lambda\langle x,x\rangle = \langle x,\lambda x\rangle \\ &= \langle x,Ax\rangle = \langle Ax,x\rangle \\ &= \langle \lambda x,x\rangle = \bar{\lambda}\langle x,x\rangle = \bar{\lambda}\|x\|^2, \end{aligned}$$

así que $\bar{\lambda} = \lambda$ porque $x \neq \mathbf{0}$. En otras palabras, vale $\lambda \in \mathbb{R}$.

[6]Esta es una de varios términos matemáticos nombrado por el francés *Charles Hermite* (1822–1901).

Ahora sean $A\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1$, $A\boldsymbol{x}_2 = \lambda_2\boldsymbol{x}_2$ con $\lambda_1 \neq \lambda_2$. Entonces

$$\begin{aligned}\lambda_2\langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle &= \langle\boldsymbol{x}_1,\lambda_2\boldsymbol{x}_2\rangle = \langle\boldsymbol{x}_1,A\boldsymbol{x}_2\rangle \\ &= \langle A\boldsymbol{x}_1,\boldsymbol{x}_2\rangle = \langle\lambda_1\boldsymbol{x}_1,\boldsymbol{x}_2\rangle = \lambda_1\langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle, \quad \text{ya que} \quad \lambda_1 \in \mathbb{R}.\end{aligned}$$

Como $\lambda_2 \neq \lambda_1$, se concluye que $\langle\boldsymbol{x}_1,\boldsymbol{x}_2\rangle = 0$.

Si A posee n autovalores distintos, sean $\boldsymbol{x}_1,\dots,\boldsymbol{x}_n$ unos autovectores tales que $A\boldsymbol{x}_j = \lambda_j\boldsymbol{x}_j$ para $j = 1,\dots,n$. Estos son vectores no nulos y además ortogonales entre sí, por el párrafo anterior. Al multiplicar cada $\boldsymbol{x}_j$ por un escalar positiva si fuera necesario, se puede también 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, es decir, 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. Para ver que A sea diagonalizable, basta comprobar que su polinomio mínimo no tiene factores repetidos. En vista de la descomposición primaria de $\mathbb{C}^n$ correspondiente al operador T_A, basta considerar un autovalor λ de A y un vector $\boldsymbol{x} \neq \boldsymbol{0}$ tal que $(A - \lambda I_n)^k\boldsymbol{x} = \boldsymbol{0}$ para algún $k \geq 2$ y mostrar que $\boldsymbol{x}$ es un autovector de A. (De lo contrario, podría haber un bloque de Jordan $J_k(\lambda)$ en la forma de Jordan de A.)

Si $2^{m-1} < k < 2^m$, entonces $(A - \lambda I_n)^k\boldsymbol{x} = \boldsymbol{0}$ implica que $(A - \lambda I_n)^{2^m}\boldsymbol{x} = \boldsymbol{0}$. Luego, se puede asumir que $k = 2^m$ para algún $m = 1,2,\dots$. Ahora, la matriz $(A - \lambda I_n)$ es hermítica porque $\lambda \in \mathbb{R}$, así que

$$0 = \langle\boldsymbol{x},\boldsymbol{0}\rangle = \langle\boldsymbol{x},(A-\lambda I_n)^{2^m}\boldsymbol{x}\rangle = \langle(A-\lambda I_n)^{2^{m-1}}\boldsymbol{x},(A-\lambda I_n)^{2^{m-1}}\boldsymbol{x}\rangle = \left\|(A-\lambda I_n)^{2^{m-1}}\boldsymbol{x}\right\|^2.$$

Se concluye que $(A - \lambda I_n)^{2^{m-1}}\boldsymbol{x} = \boldsymbol{0}$ en $\mathbb{C}^n$. Al repetir este argumento m veces, se obtiene $(A - \lambda I_n)\boldsymbol{x} = \boldsymbol{0}$, es decir, $\boldsymbol{x}$ es un autovector para el autovector λ.

Así las cosas, si $\nu_1,\dots,\nu_r$ son los autovalores distintos de A, entonces la descomposición primaria para T_A es $\mathbb{C}^n = W_1 \oplus W_2 \oplus \cdots \oplus W_r$, donde cada $W_i = \ker(T_A - \nu_i I)$ consta de todos los autovectores para el autovalor ν_i (más el vector nulo). Estos subespacios son *mutuamente ortogonales*: $\langle\boldsymbol{x}_i,\boldsymbol{x}_j\rangle = 0$ para $\boldsymbol{x}_i \in W_i$, $\boldsymbol{x}_j \in W_j$ con $i \neq j$. Ejíjase una base ortonormal en cada subespacio W_i (si es necesario, una base preexistente puede modificarse con el algoritmo de gram y Schmidt). Su unión es una base ortonormal de $\mathbb{C}^n$, formado por autovectores de A. □

► 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. Se generan gran cantidad de ejemplos por el algoritmo de Gram y Schmidt, como será evidente en los Ejercicios al final de este capítulo.

Definición 3.27. (a) Una matriz cuadrada $Q \in M_n(\mathbb{R})$ es una **matriz ortogonal** si sus columnas forman una base ortonormal de $\mathbb{R}^n$.

(b) Una matriz cuadrada $U \in M_n(\mathbb{C})$ es una **matriz unitaria** si sus columnas forman una base ortonormal de $\mathbb{C}^n$.

Proposición 3.28. *(a) Una matriz cuadrada $Q \in M_n(\mathbb{R})$ es ortogonal si y sólo si $\underline{Q^tQ = I_n}$.*

(b) Una matriz cuadrada $U \in M_n(\mathbb{C})$ es unitaria si y sólo si $\underline{U^*U = I_n}$.

Demostración. Es suficiente demostrar la parte (b), porque una matriz *real* es ortogonal si y sólo si es unitaria.

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 sólo si $\langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle = [\![i=j]\!]$ para $i, j = 1,\dots,n$.

La entrada (i, j) de la matriz $M = U^*U$ cumple

$$m_{ij} = \sum_{k=1}^{n} \bar{u}_{ki} u_{kj} = \langle \boldsymbol{u}_i, \boldsymbol{u}_j\rangle,$$

así que U es unitaria si y sólo si $m_{ij} = [\![i=j]\!]$ para todo i, j, si y sólo si $M = I_n$. □

Corolario 3.29. *(a) Una matriz cuadrada $Q \in M_n(\mathbb{R})$ es ortogonal si y sólo si es inversible, con $Q^{-1} = Q^t$.*

(b) Una matriz cuadrada $U \in M_n(\mathbb{C})$ es unitaria si y sólo si es inversible, con $U^{-1} = U^*$.

Demostración. Si $Q \in M_n(\mathbb{R})$ es ortogonal, entonces $r(Q) = n$ porque sus n columnas son linealmente independientes. Por tanto, Q es inversible. La fórmula $Q^tQ = I_n$ dice que Q^{-1} es necesariamente igual a Q^t; en otras palabras, vale $\underline{QQ^t = I_n}$ también.

Este argumento se aplica, *mutatis mutandis*, al caso unitario: vale $\underline{UU^* = I_n}$ también. □

► Las matrices de la forma A^tA (en el caso $\mathbb{F} = \mathbb{R}$) o bien A^*A (en el caso $\mathbb{F} = \mathbb{C}$) son de gran importancia en el álgebra lineal. En primer lugar, los rangos de A, A^*, A^*A y AA^* coinciden, como se demuestra a continuación.

Es parte del teorema de rango y nulidad (Proposición 1.23) que $r(T) = r(T^t)$ para una aplicación lineal T cualquiera. Bajo la identificación (3.8) de la transpuesta abstracta $T^t \in \mathcal{L}(W^*, V^*)$ y la aplicación adjunta $T^* \in \mathcal{L}(W, V)$, se obtiene $r(T) = r(T^*)$. La correspondencia $A \leftrightarrow T_A$ entre matrices y operadores conlleva las igualdades

$$r(A) = r(A^*) \text{ para } A \in \mathbb{C}^{m\times n}, \qquad r(B) = r(B^t) \text{ para } B \in \mathbb{R}^{m\times n}.$$

Lema 3.30. *Si $A \in \mathbb{C}^{m\times n}$, entonces $r(A) = r(\overline{A}) = r(A^*) = r(A^t)$.*

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:

$$\bar{c}_1\,\boldsymbol{a}_1 + \bar{c}_2\,\boldsymbol{a}_2 + \cdots + \bar{c}_n\,\boldsymbol{a}_n = \boldsymbol{0}.$$

De ahí se ve que $r(\overline{A}) = r(A)$.

La igualdad $r(A) = r(A^*)$ viene del teorema de rango y nulidad. Además, como $A^* = \overline{A^t}$ en $\mathbb{C}^{n\times m}$, el argumento del párrafo anterior muestra que $r(A^*) = r(A^t)$. □

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 $A\boldsymbol{x} = \boldsymbol{0} \implies A^*A\boldsymbol{x} = \boldsymbol{0}$; además,

$$A^*A\boldsymbol{x} = \boldsymbol{0} \implies \langle \boldsymbol{x}, A^*A\boldsymbol{x}\rangle = 0 \implies \langle A\boldsymbol{x}, A\boldsymbol{x}\rangle = 0 \implies A\boldsymbol{x} = \boldsymbol{0}.$$

De ahí se concluye que $A\boldsymbol{x} = \boldsymbol{0}$ en $\mathbb{C}^m$ si y sólo si $A^*A\boldsymbol{x} = \boldsymbol{0}$ en $\mathbb{C}^n$. Por lo tanto, los subespacios $\ker T_A$ y $\ker T_{A^*A}$ de $\mathbb{C}^n$ coinciden. Luego $n(A) = n(A^*A)$.

Del teorema de rango y nulidad, se obtiene $r(A^*A) = n - n(A^*A) = n - n(A) = r(A)$. □

Corolario 3.32. *Si $A \in \mathbb{R}^{m\times n}$, entonces $r(A^tA) = r(A)$.* ⊟

Corolario 3.33. *Si $A \in \mathbb{C}^{m\times n}$, entonces $r(AA^*) = r(A)$.*

Demostración. Al aplicar la Proposición 3.31 a la matriz compleja A^* en vez de A, se obtiene $r(AA^*) = r(A^*)$. Por el Lema 3.30, se concluye que $r(AA^*) = r(A^*) = r(A)$. □

Lema 3.34. *Una matriz hermítica $A \in M_n(\mathbb{C})$ es nula, $A = O$, si y sólo 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 sólo si $\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = 0$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$. En efecto, si se cumple esta condición, se puede tomar $\boldsymbol{y} = A\boldsymbol{x}$ para que $A\boldsymbol{x} = \boldsymbol{0}$ en $\mathbb{C}^n$ para todo $\boldsymbol{x} \in \mathbb{C}^n$, esto es, $T_A = 0$ en $\mathrm{End}(\mathbb{C}^n)$, lo cual implica que $A = O$.

Ahora sea A una matriz hermítica y supóngase que $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = 0$ para todo $\boldsymbol{x} \in \mathbb{C}^n$. Entonces, para todo $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$, vale

$$\langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle \boldsymbol{x}, A\boldsymbol{y}\rangle = \langle (\boldsymbol{x}+\boldsymbol{y}), A(\boldsymbol{x}-\boldsymbol{y})\rangle - \langle \boldsymbol{x}, A\boldsymbol{x}\rangle - \langle \boldsymbol{y}, A\boldsymbol{y}\rangle = 0$$

y por tanto $2\Re\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = \langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle A\boldsymbol{x}, \boldsymbol{y}\rangle = \langle \boldsymbol{y}, A\boldsymbol{x}\rangle + \langle \boldsymbol{x}, A\boldsymbol{y}\rangle = 0$ porque $A = A^*$.

Si $\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = re^{i\theta} \in \mathbb{C}$, entonces $|\langle \boldsymbol{y}, A\boldsymbol{x}\rangle| = r = e^{-i\theta}\langle \boldsymbol{y}, A\boldsymbol{x}\rangle = \langle e^{i\theta}\boldsymbol{y}, A\boldsymbol{x}\rangle$. Del párrafo anterior, con $e^{i\theta}\boldsymbol{y}$ en lugar de $\boldsymbol{y}$, se concluye 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 $A = O$. □

Lema 3.35. *Una matriz cuadrada $A \in M_n(\mathbb{C})$ es hermítica si y sólo si $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle \in \mathbb{R}$ para todo $\boldsymbol{x} \in \mathbb{C}^n$.*

Demostración. Si $A = A^*$, entonces para cada $\boldsymbol{x} \in \mathbb{C}^n$ vale

$$\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \langle A^*\boldsymbol{x}, \boldsymbol{x}\rangle = \langle A\boldsymbol{x}, \boldsymbol{x}\rangle = \overline{\langle \boldsymbol{x}, A\boldsymbol{x}\rangle},$$

así que $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle$ es real.

Por otro lado, si $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle \in \mathbb{R}$ para todo $\boldsymbol{x} \in \mathbb{C}^n$, entonces

$$\langle \boldsymbol{x}, (A - A^*)\boldsymbol{x}\rangle = \langle \boldsymbol{x}, A\boldsymbol{x}\rangle - \langle \boldsymbol{x}, A^*\boldsymbol{x}\rangle = \langle \boldsymbol{x}, A\boldsymbol{x}\rangle - \langle A\boldsymbol{x}, \boldsymbol{x}\rangle = 0 \quad \text{para todo} \quad \boldsymbol{x} \in \mathbb{C}^n.$$

Por el Lema anterior, se concluye que $A - A^* = O$ en $M_n(\mathbb{C})$. □

Definición 3.36. Una matriz cuadrada $A \in M_n(\mathbb{C})$ es una **matriz positiva** si es hermítica y si además

$$\langle \boldsymbol{x}, A\boldsymbol{x}\rangle \geq 0 \quad \text{para todo} \quad \boldsymbol{x} \in \mathbb{C}^n. \tag{3.10}$$

En particular, una matriz cuadrática *real* $A \in M_n(\mathbb{R})$ es positiva si y sólo si es *simétrica* y cumple (3.10), es decir, si y sólo si $\boldsymbol{x}^t A\boldsymbol{x} \geq 0$ para todo $\boldsymbol{x} \in \mathbb{R}^n$.

Se dice que una matriz positiva $A \in M_n(\mathbb{C})$ es **definida positiva**[7] si la desigualdad en (3.10) es estricta para vectores no nulos: $\underline{\langle \boldsymbol{x}, A\boldsymbol{x}\rangle > 0}$ para todo $\boldsymbol{x} \neq \boldsymbol{0}$ 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,

$$\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = \langle \boldsymbol{x}, B^*B\boldsymbol{x}\rangle = \langle B\boldsymbol{x}, B\boldsymbol{x}\rangle = \|B\boldsymbol{x}\|^2 \geq 0$$

para $\boldsymbol{x} \in \mathbb{C}^n$. (Resulta que este ejemplo es universal: para cada matriz positiva A, se puede mostrar la existencia de una matriz B tal que $A = B^*B$. Esto se verificará más adelante.)

Del mismo modo, cada matriz de la forma $\underline{A = CC^*}$ es positiva: tómese $B := C^*$.

Proposición 3.38. *Una matriz cuadrada $A \in M_n(\mathbb{C})$ es definida positiva si y sólo si es positiva e inversible.*

Demostración. Como A es hermítica, $\mathbb{C}^n$ posee una base ortonormal $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ formado por autovectores de A. Sean $\lambda_1, \dots, \lambda_n$ los autovalores correspondientes. La matriz A es inversible si y sólo si su forma diagonal $\operatorname{diag}[\lambda_1, \dots, \lambda_n]$ es inversible, si y sólo si $\lambda_i \neq 0$ para $i = 1, \dots, n$.

Entonces, si A no es inversible, hay al menos un autovalor nulo $\lambda_i = 0$. En este caso, el autovector $\boldsymbol{u}_i$ cumple $\langle \boldsymbol{u}_i, A\boldsymbol{u}_i\rangle = \langle \boldsymbol{u}_i, \boldsymbol{0}\rangle = 0$, así que A no es definida positiva.

Supóngase que A es positiva e inversible, 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

$$A\boldsymbol{x} = c_1\lambda_1\boldsymbol{u}_1 + \cdots + c_n\lambda_n\boldsymbol{u}_n; \qquad \langle \boldsymbol{x}, A\boldsymbol{x}\rangle = |c_1|^2\lambda_1 + \cdots + |c_n|^2\lambda_n,$$

de modo que $\langle \boldsymbol{x}, A\boldsymbol{x}\rangle = 0$ si y sólo si cada $c_i = \boldsymbol{0}$, si y sólo si $\boldsymbol{x} = \boldsymbol{0}$. Por tanto, A es definida positiva. □

► Cada matriz real simétrica es una matriz hermítica, de oficio. Por tanto, la Proposición anterior proporciona un criterio para detectar si una matriz (real, simétrica) dada, A, sea definida positiva o no. Para que ese criterio sea eficaz, hay que determinar si A es inversible (por ejemplo, al evaluar $\det A$); también hay que comprobar las relaciones de positividad (3.10). La eliminación gaussiana simple (sin intercambios de filas) permite resolver todas estas inquietudes simultáneamente.

Proposición 3.39. *Una matriz $A \in M_n(\mathbb{R})$ es definida positiva si y sólo si es simétrica y los pivotes sucesivos $a_{kk}^{(k)}$ en la eliminación gaussiana simple son todos positivos.*

[7] Una consecuencia desafortunada de esta terminología es que la matriz nula O es positiva, aunque no definida positiva. Por razones históricas, el término *matriz no negativa* ha sido reservado para otra noción: una matriz real (no necesariamente simétrica) es "no negativa" si todas sus entradas son números reales no negativas. Hay matrices reales simétricas que son "no negativas" pero no son positivas en el sentido de la Definición 3.36; hay otras que son positivas pero no son "no negativos".

Demostración. La eliminación gaussiana simple, cuando es aplicable con éxito a una matriz cuadrada inversible, produce una factorización $A = LDU$, en donde D es una matriz diagonal, L es triangular inferior unipotente y U es triangular superior unipotente; ésta descomposición es única. Obsérvese que $A^t = U^tDL^t$ es la descomposición correspondiente de la matriz transpuesta A^t. Por tanto, es $U = L^t$ cuando A es simétrica.

Dicha eliminación gaussiana *simple* funciona cuando se puede garantizar que ningún pivote $a_{kk}^{(k)}$ se anulará durante el proceso. Si A es definida positiva, entonces

$$a_{11}^{(1)} = a_{11} = \langle \boldsymbol{e}_1, A\boldsymbol{e}_1 \rangle > 0.$$

Para $k > 1$, $a_{kk}^{(k)}$ es el elemento (k,k) de una matriz M_kA, donde la premultiplicación $A \mapsto M_kA$ ejecuta las operaciones de fila, del tipo (b), que colocan ceros debajo de la diagonal en las primeras $(k-1)$ columnas de A. La *postmultiplicación* $M_kA \mapsto M_kAM_k^t$ ejecutaría las operaciones *de columna* correspondientes; y estas operaciones de columna no cambian el elemento (k,k) de las matrices intermedios, porque en cada paso se efectúa $a_{kk}^{(k)} \mapsto a_{kk}^{(k)} + 0$. Entonces $a_{kk}^{(k)}$ es el elemento (k,k) de la matriz simétrica $M_kAM_k^t$, así que

$$a_{kk}^{(k)} = \langle \boldsymbol{e}_k, M_kAM_k^t\boldsymbol{e}_k \rangle = \langle M_k^t\boldsymbol{e}_k, AM_k^t\boldsymbol{e}_k \rangle > 0$$

por ser A definida positiva. Fíjese que $M_k^t\boldsymbol{e}_k \neq \boldsymbol{0}$ porque $\langle M_k^t\boldsymbol{e}_k, \boldsymbol{e}_k \rangle = \langle \boldsymbol{e}_k, M_k\boldsymbol{e}_k \rangle = 1$: al aplicar las mismas operaciones de fila a la matriz identidad, no cambian la entrada 1 en la posición (k,k) de la matriz I_n. Por tanto, una matriz definida positiva tiene la factorización $A = LDL^t$ sin necesidad de hacer intercambios de filas. El factor diagonal es

$$D := \begin{bmatrix} a_{11} & 0 & 0 & \dots & 0 \\ 0 & a_{22}^{(2)} & 0 & \dots & 0 \\ 0 & 0 & a_{33}^{(3)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{nn}^{(n)} \end{bmatrix}.$$

Por otro lado, sea A una matriz simétrica e inversible que admite factorización $A = LDL^t$ por eliminación gaussiana simple. Sea $D =: [d_{ij}]$. Si $d_{kk} > 0$ para $k = 1,\dots,n$, entonces $D = D^{1/2}D^{1/2}$, al definir $D^{1/2} := \operatorname{diag}[\sqrt{d_{11}},\dots \sqrt{d_{nn}}]$. Entonces

$$A = LDL^t = LD^{1/2}D^{1/2}L^t = B^tB, \quad \text{donde} \quad B := D^{1/2}L^t.$$

Luego A es inversible y positiva; por la Proposición 3.38, A es definida positiva. □

Definición 3.40. Sea $A \in M_n(\mathbb{R})$ una matriz definida positiva y sea $A = LDL^t$ su factorización por eliminación gaussiana; sea $C := LD^{1/2}$, la cual es una matriz triangular inferior; entonces $A = CC^t$ se llama la **factorización de Cholesky** de A.

La idea de esta factorización es la siguiente: si $A = BB^t$ para *alguna* matriz B, hay diversas posibilidades para B; por ejemplo, se puede reemplazar B por $(-B)$. Entre estas posibilidades, hay una, $B = C$, que es triangular inferior, con entradas positivas en la diagonal. Se obtiene el factor C por una variante de la eliminación gaussiana.

3.4 Operadores sobre espacios hilbertianos

En esta sección, V denotará un espacio vectorial sobre $\mathbb{C}$, de dimensión finita n, dotado de un producto escalar $\langle\cdot,\cdot\rangle$. definido en V. La incómoda frase "espacio vectorial complejo con producto escalar" se puede abreviar por el término **espacio hilbertiano**, o bien "espacio de Hilbert".[8] Un espacio vectorial sobre $\mathbb{R}$ con un producto escalar recibe el nombre de **espacio euclidiano**.[9]

En vista de la fórmula (3.9), el **adjunto** de un operador lineal $T \in \operatorname{End}(V)$ es el operador lineal $T^* \in \operatorname{End}(V)$ determinado por

$$\langle T^*(\boldsymbol{y}), \boldsymbol{x}\rangle = \langle \boldsymbol{y}, T(\boldsymbol{x})\rangle, \quad \text{para todo} \quad \boldsymbol{x}, \boldsymbol{y} \in V. \tag{3.11}$$

Más generalmente, si $T \in \mathcal{L}(V, W)$, esta fórmula define $T^* \in \mathcal{L}(W, V)$.

Proposición 3.41. *Sean V, W son dos espacios hilbertianos y sea $T \in \mathcal{L}(V, W)$. Entonces* $\ker T^*$ *es el complemento ortogonal $T(V)^\perp$ de la imagen de T, y* $\ker T$ *es el complemento ortogonal $T^*(W)^\perp$ de la imagen de T^*.*

En consecuencia, T^* es uno-a-uno si y sólo si T es sobreyectivo y viceversa.

Demostración. Obsérvese que $T(V)^\perp \leq W$ y que $T^*(W)^\perp \leq V$.

Para un vector $\boldsymbol{y} \in W$, vale

$$\begin{aligned} \boldsymbol{y} \in \ker T^* &\iff T^*(\boldsymbol{y}) = \boldsymbol{0} \\ &\iff \langle T^*(\boldsymbol{y}), \boldsymbol{x}\rangle = 0 \text{ para todo } \boldsymbol{x} \in V \\ &\iff \langle \boldsymbol{y}, T(\boldsymbol{x})\rangle = 0 \text{ para todo } T(\boldsymbol{x}) \in T(V) \\ &\iff \boldsymbol{y} \in T(V)^\perp. \end{aligned}$$

De igual manera, para un vector $\boldsymbol{x} \in V$, vale

$$\begin{aligned} \boldsymbol{x} \in \ker T &\iff T(\boldsymbol{x}) = \boldsymbol{0} \\ &\iff \langle \boldsymbol{y}, T(\boldsymbol{x})\rangle = 0 \text{ para todo } \boldsymbol{y} \in W \\ &\iff \langle T^*(\boldsymbol{y}), \boldsymbol{x}\rangle = 0 \text{ para todo } T^*(\boldsymbol{y}) \in T^*(W) \\ &\iff \boldsymbol{x} \in T^*(W)^\perp. \end{aligned}$$

□

Definición 3.42. Un operador lineal $T \in \operatorname{End}(V)$ es **autoadjunto** si $\underline{T^* = T}$.

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 $M_n(\mathbb{C})$.

[8] Los trabajos de *David Hilbert* sobre ecuaciones integrales, que condujeron al estudio de espacios de funciones de cuadrado integrable, cobraron nueva relevancia con la formulación de la mecánica cuántica en los años 1925–30. El término "espacio de Hilbert" fue introducido por su estudiante *John von Neumann* (un húngaro, cuyo nombre original fue Neyman János) en 1929, para describir espacios vectoriales infinitodimensionales (completos) con producto escalar. Hoy en día, se usa ese término para al caso finitodimensional también.

[9] Algunos autores emplean el término *espacio unitario* para denotar un espacio finitodimensional complejo con producto escalar, en vez de "espacio hilbertiano". Es preferible limitar el adjetivo "unitario" a los operadores unitarios.

Lema 3.43. *Todo $T \in \mathrm{End}(V)$ puede escribirse, de forma única, como $T = T_1 + iT_2$ donde T_1 y T_2 son operadores autoadjuntos.*

Demostración. Si $T = T_1 + iT_2$ con T_1, T_2 autoadjuntos, entonces $T^* = T_1^* - iT_2^* = T_1 - iT_2$ por la semilinealidad de la correspondencia $T \mapsto T^*$. Por lo tanto, vale

$$T_1 = \tfrac{1}{2}(T^* + T), \qquad T_2 = \tfrac{i}{2}(T^* - T). \tag{3.12}$$

Estas fórmulas establecen la unicidad de la "parte real" T_1 y de la "parte imaginaria" T_2 del operador T.

Por otro lado, (3.12) permite *definir* T_1 y T_2 por $T_1 := \frac{1}{2}(T^* + T)$, $T_2 := \frac{i}{2}(T^* - T)$. Queda claro que $T_1^* = \frac{1}{2}(T + T^*) = T_1$ y $T_2^* = -\frac{i}{2}(T - T^*) = T_2$. Esto establece la existencia de la descomposición deseada. □

Lema 3.44. *Si $T \in \mathrm{End}(V)$ es autoadjunto, sus autovalores son reales.*

Demostración. Es cuestión de adaptar la demostración de la Proposición 3.25. Si $T^* = T$, si $\lambda \in \mathbb{C}$ es un autovalor de T y si $\boldsymbol{x} \neq \boldsymbol{0}$ es un autovector correspondiente, 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, $\lambda \in \mathbb{R}$. □

Obsérvese también que si $T^* = T$, entonces $\langle \boldsymbol{x}, T(\boldsymbol{x})\rangle \in \mathbb{R}$ 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 $T \in \mathrm{End}(V)$ en un espacio vectorial cualquiera se llama **idempotente** si $\underline{T^2 = T}$. Fíjese que $T^k = T$ también, para $k = 3, 4, \dots$.

Un operador lineal $P \in \mathrm{End}(V)$ en un espacio hilbertiano es un **proyector ortogonal**, o simplemente un **proyector**,[10] si P es *idempotente y autoadjunto*, es decir, si $\underline{P^2 = P = P^*}$.

Sea $P \in \mathrm{End}(V)$ un proyector ortogonal. La restricción de P a su imagen $P(V)$ es el operador identidad sobre $P(V)$, porque $P(P(\boldsymbol{x})) = P(\boldsymbol{x})$ para todo $\boldsymbol{x} \in V$. Por otro lado, es $P(V)^\perp = \ker P^* = \ker P$ por la Proposición 3.41, así que la restricción de P al complemento ortogonal de su imagen es el operador cero.

Los proyectores abundan en $\mathrm{End}(V)$, 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, hay un único proyector ortogonal P_M en $\mathrm{End}(V)$ tal que $P_M(V) = M$.*

---

[10]Cualquier operador idempotente $T \in \mathrm{End}(V)$ 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 la $*$-álgebra $\mathrm{End}(V)$ que cumple $P^2 = P = P^*$), aunque muchos autores las confunden. *Caveat lector*.

Demostración. Es evidente que $M \cap M^\perp = \{\mathbf{0}\}$, porque $\boldsymbol{x} \in M \cap M^\perp \implies \langle \boldsymbol{x}, \boldsymbol{x} \rangle = 0 \implies \boldsymbol{x} = \mathbf{0}$.

Si $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_m\}$ es una base ortonormal del subespacio M, hay vectores $\{\boldsymbol{e}_{m+1}, \dots, \boldsymbol{e}_n\}$ tales que la unión $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ sea una base ortonormal de V, por el Corolario 3.19. Es fácil comprobar que $\{\boldsymbol{e}_{m+1}, \dots, \boldsymbol{e}_n\}$ es una base ortonormal de $M^\perp$. Luego $\dim M + \dim(M^\perp) = \dim V$. Se concluye que $\underline{V = M \oplus M^\perp}$ como suma directa de espacios vectoriales complejos.

Defínase $P_M \in \operatorname{End}(V)$ por

$$P_M(\boldsymbol{x}) := \langle \boldsymbol{e}_1, \boldsymbol{x} \rangle \boldsymbol{e}_1 + \cdots + \langle \boldsymbol{e}_m, \boldsymbol{x} \rangle \boldsymbol{e}_m.$$

Queda claro que $P_M(V) = M$ y que $P_M^2 = P_M$.

Si $\boldsymbol{x}, \boldsymbol{y} \in V$, entonces

$$\begin{aligned}
\langle \boldsymbol{y}, P_M(\boldsymbol{x}) \rangle &= \Big\langle \sum_{j=1}^n \langle \boldsymbol{e}_j, \boldsymbol{y} \rangle \boldsymbol{e}_j, \sum_{i=1}^m \langle \boldsymbol{e}_i, \boldsymbol{x} \rangle \boldsymbol{e}_i \Big\rangle = \sum_{j=1}^n \sum_{i=1}^m \overline{\langle \boldsymbol{e}_j, \boldsymbol{y} \rangle} \langle \boldsymbol{e}_i, \boldsymbol{x} \rangle \langle \boldsymbol{e}_j, \boldsymbol{e}_i \rangle \\
&= \sum_{i=1}^m \overline{\langle \boldsymbol{e}_i, \boldsymbol{y} \rangle} \langle \boldsymbol{e}_i, \boldsymbol{x} \rangle = \Big\langle \sum_{j=1}^m \langle \boldsymbol{e}_j, \boldsymbol{y} \rangle \boldsymbol{e}_j, \sum_{i=1}^n \langle \boldsymbol{e}_i, \boldsymbol{x} \rangle \boldsymbol{e}_i \Big\rangle = \langle P_M(\boldsymbol{y}), \boldsymbol{x} \rangle,
\end{aligned}$$

así que $P_M^* = P_M$. Ahora es evidente que la matriz de P_M, respecto de la base ortonormal $\mathcal{E}$, es la matriz de bloques

$$[P_M]^{\mathcal{E}}_{\mathcal{E}} = \begin{bmatrix} I_m & O \\ O & O \end{bmatrix} = \operatorname{diag}[\underbrace{1, \dots, 1}_{m}, \underbrace{0, \dots, 0}_{n-m}].$$

Para la unicidad, obsérvese que cualquier proyector ortogonal P con $P(V) = M$ cumple $P(\boldsymbol{z}) = \boldsymbol{z}$ para $\boldsymbol{z} \in M$ y $P(\boldsymbol{y}) = \mathbf{0}$ 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 $P_M(\boldsymbol{x}) = P_M(\boldsymbol{z}) + P_M(\boldsymbol{y}) = \boldsymbol{z} + \mathbf{0} = \boldsymbol{z}$. Se ha mostrado que $P(\boldsymbol{x}) = P_M(\boldsymbol{x})$ para todo $\boldsymbol{x} \in V$, así que $P = P_M$. □

Lema 3.47. *Dos subespacios M y N de V son ortogonales, es decir, $\langle \boldsymbol{z}, \boldsymbol{x} \rangle = 0$ para $\boldsymbol{x} \in M$ y $\boldsymbol{z} \in N$, si y sólo si $P_M P_N = 0$ en* $\operatorname{End}(V)$.

Demostración. Si $\boldsymbol{x} \in M$ y $\boldsymbol{z} \in N$, entonces

$$\langle \boldsymbol{z}, \boldsymbol{x} \rangle = \langle P_N \boldsymbol{z}, P_M \boldsymbol{x} \rangle = \langle P_M^* P_N \boldsymbol{z}, \boldsymbol{x} \rangle = \langle P_M P_N \boldsymbol{z}, \boldsymbol{x} \rangle.$$

Luego $P_M P_N = 0$ implica que $\langle \boldsymbol{z}, \boldsymbol{x} \rangle = 0$ para $\boldsymbol{x} \in M$, $\boldsymbol{z} \in N$, es decir, que $M \perp N$.

Por otro lado, si $M \perp N$, entonces $N \subseteq M^\perp$. Si $\boldsymbol{y} \in V$, entonces $P_N(\boldsymbol{y}) \in N$ y por ende $P_N(\boldsymbol{y}) \in M^\perp$, por lo tanto $P_M(P_N(\boldsymbol{y})) = \mathbf{0}$. Como $\boldsymbol{y}$ es arbitrario, esto dice que $P_M P_N = 0$. □

► En un espacio hilbertiano, la *semejanza* de matrices $A \sim P^{-1}AP$ no es la clasificación más natural, pues P podría ser una matriz inversible cualquiera sin tomar en cuenta el producto escalar. Las matrices cuadradas en $M_n(\mathbb{C})$ se clasifican mejor por *semejanza unitaria*, es decir, por la relación $\underline{A \sim U^{-1}AU}$ 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* $T \in \mathrm{End}(V)$*, entonces hay una cadena de subespacios*

$$\{\mathbf{0}\} \leq R_1 \leq \cdots \leq R_{n-1} \leq R_n = V$$

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. Por inducción sobre $n = \dim V$; el resultado es evidente si $\dim V = 1$. Supóngase que la Proposición sea válida para espacios de dimensión $(n-1)$.

Sea $\mu \in \mathbb{C}$ un autovalor[11] de T^* y $\boldsymbol{y} \in V$ un autovector con $T^*(\boldsymbol{y}) = \mu\boldsymbol{y}$. El subespacio unidimensional $\mathrm{lin}\langle\boldsymbol{y}\rangle = \{\alpha\boldsymbol{y} : \alpha \in \mathbb{C}\}$ es invariante bajo T^*. Defínase $R_{n-1} := \mathrm{lin}\langle\boldsymbol{y}\rangle^\perp$. 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 $T(\boldsymbol{x}) \in R_{n-1}$. Además, es $\dim R_{n-1} = n - \dim(\mathrm{lin}\langle\boldsymbol{y}\rangle) = n-1$.

Sea $S \in \mathrm{End}(R_{n-1})$ la restricción de T a este subespacio: es $S(\boldsymbol{x}) := T(\boldsymbol{x}) \in R_{n-1}$ para $\boldsymbol{x} \in R_{n-1}$. Al reemplazar V por R_{n-1} y T por S, la hipótesis inductiva muestra que hay una cadena de subespacios $\{\mathbf{0}\} \leq R_1 \leq \cdots \leq R_{n-1}$ con $\dim R_k = k$ y $S(R_k) \subset R_k$ para $k = 1, \dots, n-1$. Como $S(\boldsymbol{x}) = T(\boldsymbol{x})$ para cada $\boldsymbol{x} \in R_k$, el resultado queda demostrado para $\dim V = n$.

Elíjase una base ortonormal $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ de V como sigue. 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 $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\} \subset R_k$, tómese $\boldsymbol{u}_{k+1} \in R_{k+1} \cap R_k^\perp$ tal que $\|\boldsymbol{u}_{k+1}\| = 1$. La existencia del 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, es claro que $\{\boldsymbol{u}_1, \dots, \boldsymbol{u}_k\}$ es una base ortonormal para R_k, con $k = 1, \dots, n$. La invariancia $T(R_k) \subseteq R_k$ significa que $T(\boldsymbol{u}_k) = \sum_{i=1}^{k} a_{ik}\boldsymbol{u}_i$ para $k = 1, \dots, n$. En otras palabras, la matriz de T respecto de esta base es triangular:

$$[T]^{\mathcal{U}}_{\mathcal{U}} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & \dots & a_{1n} \\ 0 & a_{22} & a_{23} & \dots & a_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & a_{n-1,n} \\ 0 & 0 & 0 & \dots & a_{nn} \end{bmatrix}. \qquad \square$$

Corolario 3.49. *Sea* $A \in M_n(\mathbb{C})$. *Entonces hay una matriz unitaria* U *tal que la matriz* $\underline{U^*AU} = U^{-1}AU$ *sea triangular.*

[11]Esta Proposición no tiene un análogo en el caso real $\mathbb{F} = \mathbb{R}$, porque no siempre puede garantizarse la existencia de autovalores para operadores en un espacio euclidiano. Podría suceder que el polinomio característico de T^t no tenga factores irreducibles de primer grado.

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 anterior para el operador $T_A \in \operatorname{End}(\mathbb{C}^n)$. La matriz de cambio de base dada por (1.9) es

$$U := [I]_{\mathcal{U}}^{\mathcal{E}} = [\boldsymbol{u}_1\ \boldsymbol{u}_2\ \dots\ \boldsymbol{u}_n], \tag{3.13}$$

la cual es unitaria porque sus columnas forman una base ortonormal de $\mathbb{C}^n$. Ahora

$$U^* A U = U^{-1} A U = [I]_{\mathcal{E}}^{\mathcal{U}}\, [T_A]_{\mathcal{E}}^{\mathcal{E}}\, [I]_{\mathcal{U}}^{\mathcal{E}} = [T_A]_{\mathcal{U}}^{\mathcal{U}}$$

es una matriz triangular. □

3.5 El teorema espectral y sus consecuencias

Cualquier operador lineal sobre un espacio hilbertiano 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. Es importante obtener un criterio para que una matriz cuadrada sea más bien *diagonalizable*. Se busca una condición sobre una matriz A que garantice que $U^* A U = D$ sea una matriz diagonal para alguna matriz unitaria U.

El caso más importante es el de las matrices hermíticas, que corresponden a operadores autoadjuntos. Se sabe, por la Proposición 3.26, que una matriz hermítica tiene una base ortonormal de autovectores $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$. La matriz U de (3.13) cuyas columnas son estos autovectores cumple $AU = UD$, donde $D = \operatorname{diag}[\lambda_1, \dots, \lambda_n]$ es la matriz diagonal que recoge los autovalores de A. Por lo tanto, en este caso la diagonalización $U^* A U = D$ es factible.

Resulta muy útil reformular esta diagonalización en términos de una colección de proyectores ortogonales. De hecho, es oportuno expresar el resultado en términos de la estructura abstracta de operadores autoadjuntos, mediante el **teorema espectral**, que se demuestra a continuación.

Teorema 3.50 (Teorema Espectral)**.** *Sea V un espacio hilbertiano, y sea $T = T^* \in \operatorname{End}(V)$ un operador autoadjunto. Entonces se puede escribir*

$$T = \mu_1 P_1 + \cdots + \mu_r P_r, \tag{3.14}$$

donde los $\mu_j \in \mathbb{R}$ son los autovalores distintos de T, los P_j son proyectores ortogonales no nulos tales que

$$P_i P_j = 0 \quad si \quad i \neq j; \qquad P_1 + \cdots + P_r = I.$$

Demostración. Sea $\mathcal{U} = \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ una base ortonormal de V respecto de la cual T tenga una matriz triangular $A = [T]_{\mathcal{U}}^{\mathcal{U}}$, en vista de la Proposición 3.48: es $a_{ij} = 0$ 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 a_{kk} cumplen $\bar{a}_{kk} = a_{kk}$, es decir, son reales.

Estos elementos diagonales son autovalores de la matriz A y también del operador T. Sean $\mu_1, \dots, \mu_r$ los elementos *distintos* de la lista $(a_{11}, \dots, a_{nn})$. Denótese por $M_j := \ker(T - \mu_j I)$ el subespacio generado por los $\boldsymbol{u}_k$ tales que $a_{kk} = \mu_j$. Sea P_j el proyector ortogonal 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 dice que $P_iP_j = 0$.

En consecuencia, la suma $P_1 + \cdots + P_r$ es también un proyector ortogonal: de hecho, esta suma es autoadjunto, y vale

$$(P_1 + \cdots + P_r)^2 = (P_1^2 + \cdots + P_r^2) + \sum_{i<j}(P_iP_j + P_jP_i) = P_1 + \cdots + P_r.$$

Cada $\boldsymbol{u}_k$ pertenece a un sólo subespacio M_j, así que $(P_1 + \cdots + P_r)(\boldsymbol{u}_k) = P_j(\boldsymbol{u}_k) = \boldsymbol{u}_k$. Por lo tanto, vale $P_1 + \cdots + P_r = I$ en $\operatorname{End}(V)$.

Para cada $k = 1,\dots,n$, vale

$$\begin{aligned}(\mu_1P_1 + \cdots + \mu_rP_r)(\boldsymbol{u}_k) &= \sum_{j=1}^{r}\mu_jP_j(\boldsymbol{u}_k) = \sum_{j=1}^{r}\mu_j\,[\![\boldsymbol{u}_k \in M_j]\!]\,\boldsymbol{u}_k \\ &= \sum_{j=1}^{r}\mu_j\,[\![\mu_j = a_{kk}]\!]\,\boldsymbol{u}_k = a_{kk}\boldsymbol{u}_k = T(\boldsymbol{u}_k),\end{aligned}$$

porque $\mu_j = a_{kk}$ para un solo índice j. Luego $T = \mu_1P_1 + \cdots + \mu_rP_r$. □

Obsérvese que en la demostración anterior, no se aprovechó la Proposición 3.26, que construyó una base ortonormal de autovectores para una matriz hermítica a partir de una forma normal de Jordan. El procedimiento actual es más directo y sencillo: dada una matriz A con $A = A^*$, se trigonaliza A por cambio de base ortonormal (Corolario 3.49) y se observe que la matriz resultante es de hecho diagonal, por ser simultáneamente triangular y hermítica. Conviene declarar esa consecuencia como corolario del teorema espectral.

Corolario 3.51. *Una matriz hermítica $A = A^* \in M_n(\mathbb{C})$ es diagonalizable, mediante conjugación $A \mapsto U^*AU$ por una matriz unitaria U.* ⊟

Definición 3.52. El conjunto de autovalores distintos $\{\mu_1,\dots,\mu_r\}$ de un operador lineal $T \in \operatorname{End}(V)$ [respectivamente, de una matriz cuadrada $A \in M_n(\mathbb{C})$] se llama el **espectro** de T [o bien de A].[12]

La descomposición $T = \mu_1P_1 + \cdots + \mu_rP_r$ de T [respectivamente, la suma directa de matrices $U^*AU = \mu_1I_{m_1} \oplus \cdots \oplus \mu_rI_{m_r}$] dados por el Teorema 3.50 y el Corolario 3.51 se llaman **descomposiciones espectrales**.

[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 la 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 no es continuo, sino que 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 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. De ahí viene la costumbre de llamar “espectro” al conjunto de autovalores (o de autovalores generalizados, en el caso infinitodimensional) de un operador lineal cualquiera.

► 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 matrices hermíticas distintas, la base ortonormal que diagonaliza la primera no diagonaliza la segunda. Sería bueno tener un criterio para poder elegir una sola base ortonormal respecto del cual las dos matrices tengan forma diagonal. Resulta que esto es posible si y sólo 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 sólo si $AB = BA$.*

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 $A = B^*$, $B = B^*$ y además $AB = BA$, sea $T_A = \mu_1 P_1 + \cdots + \mu_r P_r$ la descomposición espectral de T_A. Sea $M_j = P_j(\mathbb{C}^n) = \ker(T_A - \mu_j I)$, de manera que $\mathbb{C}^n = M_1 \oplus \cdots \oplus M_r$ es la descomposición primaria de $\mathbb{C}^n$ para el operador T_A. Si $\boldsymbol{x} \in M_j$, entonces

$$AB\boldsymbol{x} = BA\boldsymbol{x} = B(\mu_j \boldsymbol{x}) = \mu_j B\boldsymbol{x},$$

lo cual dice que el vector $B\boldsymbol{x}$ pertenece al subespacio M_j; por tanto, es $T_B(M_j) \subseteq M_j$. 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 $\{\tilde{\boldsymbol{u}}_1, \dots, \tilde{\boldsymbol{u}}_n\}$ es una base ortonormal de autovectores de A formado por concatenar 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 además $\widetilde{U}^*B\widetilde{U} = B_1 \oplus \cdots \oplus B_r$ es una suma directa de bloques. Cada bloque B_j es una matriz hermítica en $M_{m_j}(\mathbb{C})$, donde $m_j = \dim M_j$.

Por un cambio de base ortonormal en cada subespacio M_j por separado, se puede diagonalizar cada B_j. Luego, hay una matriz unitaria de la forma $V = V_1 \oplus \cdots \oplus V_r$ tal que $V^*(\widetilde{U}^*B\widetilde{U})V$ sea diagonal. Sea $U := \widetilde{U}V$ y nótese que U es unitaria por ser el producto de dos matrices unitarias. Cada V_j conmuta con el bloque escalar $\mu_j I_{m_j}$ de la matriz diagonal $\widetilde{U}^*A\widetilde{U}$, así que $U^*AU = V(\widetilde{U}^*A\widetilde{U})V = \widetilde{U}^*A\widetilde{U}$. Luego U^*BU y U^*AU ambas son matrices diagonales. □

Corolario 3.54. *Dos operadores autoadjuntos sobre un espacio hilbertiano $S, T \in \operatorname{End}(V)$ poseen matrices diagonales respecto de una misma base ortonormal si y sólo si $ST = TS$.* □

► El teorema espectral sigue válido para una clase de operadores lineales más amplia que los operadores autoadjuntos. Considérese un operador $T \in \operatorname{End}(V)$ que posee una base ortonormal de autovectores con autovalores distintos $\mu_1, \dots, \mu_r \in \mathbb{C}$ que no son necesariamente reales. Entonces se puede escribir:

$$T = \mu_1 P_1 + \cdots + \mu_r P_r, \qquad T^* = \bar{\mu}_1 P_1 + \cdots + \bar{\mu}_r P_r.$$

Si algún $\mu_j \notin \mathbb{R}$, entonces T no es autoadjunto, pero se cumple la relación

$$T^*T = |\mu_1|^2 P_1 + \cdots + |\mu_r|^2 P_r = TT^*.$$

Definición 3.55. Un operador lineal sobre un espacio hilbertiano $T \in \operatorname{End}(V)$ es un **operador normal** si y sólo si $\underline{T^*T = TT^*}$.

Una matriz cuadrada compleja $A \in M_n(\mathbb{C})$ es una **matriz normal** si y solo si $\underline{A^*A = AA^*}$.

En particular, los operadores unitarios o autoadjuntos [respectivamente, las matrices unitarias o hermíticas] son normales.

Proposición 3.56 (Teorema Espectral II)**.** *Sea V un espacio hilbertiano, y sea $T \in \operatorname{End}(V)$ un operador* normal. *Entonces se puede escribir $T = \mu_1 P_1 + \cdots + \mu_r P_r$ como en* (3.14)*, donde los $\mu_j \in \mathbb{C}$ son los autovalores distintos de T, los P_j son proyectores ortogonales no nulos tales que $P_i P_j = 0$ para $i \neq j$, y además $P_1 + \cdots + P_r = I$.*

Hay dos maneras alternativas de comprobar este resultado.

Demostración 1. Por el Lema 3.43, vale $T = T_1 + iT_2$ donde T_1 y T_2 son operadores autoadjuntos. Queda claro que $T^* = T_1 - iT_2$. Luego T es normal si y sólo si sus partes real e imaginaria conmutan, es decir, $T_1T_2 = T_2T_1$.

Por el Corolario 3.54, T es normal si y sólo si hay una base ortonormal $\mathcal{U}$ para V, para la cual las matrices $[T_1]^{\mathcal{U}}_{\mathcal{U}}$ y $[T_2]^{\mathcal{U}}_{\mathcal{U}}$ tienen 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 $\mu_1, \dots, \mu_r$ los autovalores distintos de T y sea $P_j \in \operatorname{End}(V)$ el proyector ortogonal cuyo imagen es $M_j := \ker(T - \mu_j I)$. Las propiedades enunciadas de los proyectores P_i se verifican al igual que en la demostración del Teorema 3.50. $\square$

Demostración 2. Obsérvese que la hipótesis $T = T^*$ fue utilizado en la demostración del Teorema 3.50 únicamente para mostrar que la matriz triangular de T es necesariamente diagonal.

La Proposición 3.48 construye, para cualquier $T \in \operatorname{End}(V)$, una base ortonormal $\mathcal{U}$ de V tal que $A := [T]^{\mathcal{U}}_{\mathcal{U}}$ sea una matriz triangular superior: es $a_{ij} = 0$ 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

$$(A^*A)_{kk} = \sum_{i \leq k} \bar{a}_{ik} a_{ik} = \sum_{i \leq k} |a_{ik}|^2, \qquad (AA^*)_{kk} = \sum_{j \geq k} a_{kj} \bar{a}_{kj} = \sum_{j \geq k} |a_{kj}|^2. \tag{3.15}$$

Si T es normal, entonces $A^*A = AA^*$. En particular, el caso $k = 1$ de (3.15) da

$$|a_{11}|^2 = |a_{11}|^2 + |a_{12}|^2 + |a_{13}|^2 + \ldots + |a_{1n}|^2,$$

por ende $a_{12} = a_{13} = \cdots = a_{1n} = 0$. El caso $k = 2$ es entonces

$$0 + |a_{22}|^2 = |a_{22}|^2 + |a_{23}|^2 + \ldots + |a_{2n}|^2,$$

luego $a_{23} = a_{24} = \cdots = a_{2n} = 0$. Al repetir este argumento para $k = 3, \dots, n$, se comprueba que $a_{kj} = 0$ toda vez que $j > k$; se concluye que la matriz A es diagonal. El resto de la demostración del Teorema 3.50 es aplicable sin otro cambio. $\square$

► Para cualquier matriz normal $A \in M_n(\mathbb{C})$, sea $T_A = \mu_1 P_1 + \cdots + \mu_r P_r$ la descomposición espectral del operador $T_A \colon \boldsymbol{x} \mapsto A\boldsymbol{x}$. Si $\mathcal{E}$ es la base estándar de $\mathbb{C}^n$, sea $E_j := [P_j]^{\mathcal{E}}_{\mathcal{E}}$ la matriz del proyector ortogonal P_j, para $j = 1, \dots, r$. Estas matrices cumplen $E_j^2 = E_j = E_j^*$ en $M_n(\mathbb{C})$. La fórmula

$$A = \mu_1 E_1 + \cdots + \mu_r E_r, \quad \text{con} \quad \begin{cases} E_j^2 = E_j = E_j^*, & \text{para } j = 1, \dots, r, \\ E_i E_j = O & \text{si } i \neq j, \\ E_1 + \cdots + E_r = I_n \end{cases} \tag{3.16}$$

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, por restricción, en $M_n(\mathbb{R})$.

Proposición 3.57. *Una matriz $A \in M_n(\mathbb{C})$ es positiva si y sólo si A es hermítica y todas sus autovalores son números no negativos. La matriz A es* definida positiva *si y sólo si A es hermítica y todas sus autovalores son números* positivos.

Demostración. Si A es una matriz positiva, entonces A es hermítica por definición. Sea $A = \mu_1 E_1 + \cdots + \mu_r E_r$ su descomposición espectral, con $\mu_1, \dots, \mu_r \in \mathbb{R}$.

Si $\boldsymbol{x} \in \mathbb{C}^n$ y si $j \in \{1, \dots, r\}$, 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

$$\langle \boldsymbol{x}, A\boldsymbol{x} \rangle = \sum_{j=1}^{r} \mu_j \langle \boldsymbol{x}, E_j \boldsymbol{x} \rangle = \sum_{j=1}^{r} \mu_j \|E_j \boldsymbol{x}\|^2. \tag{3.17}$$

Ahora, si $\mu_j \geq 0$ para cada j, esta relación (3.17) muestra que A es una matriz positiva. Por otro lado, 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 $\langle \boldsymbol{x}_j, A\boldsymbol{x}_j \rangle = \mu_j \|\boldsymbol{x}_j\|^2$. Luego, la positividad de A implica que $\mu_j \geq 0$ para cada j.

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 (3.17) se ve que $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle > 0$ para todo $\boldsymbol{x} \neq \boldsymbol{0}$ si y sólo si $\mu_j > 0$ para cada $j = 1, \dots, r$. □

Proposición 3.58. *Para una matriz $A \in M_n(\mathbb{C})$, son equivalentes las siguientes condiciones:*

(a) A es una matriz positiva: $A = A^*$ y $\langle \boldsymbol{x}, A\boldsymbol{x} \rangle \geq 0$ 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 anterior. Las implicaciones (d) $\implies$ (c) $\implies$ (a) son evidentes.

Ad(b) $\implies$ (d): Sea $A = \mu_1 E_1 + \cdots + \mu_r E_r$ la descomposición espectral de A con cada $\mu_j \geq 0$, y defínase

$$R := \sqrt{\mu_1}\, E_1 + \sqrt{\mu_2}\, E_2 + \cdots + \sqrt{\mu_r}\, E_r.$$

La matriz R cumple la condición (b) y por tanto es una matriz positiva (en particular, R es hermítica). Además,

$$R^2 = \sum_{j=1}^{r} \mu_j E_j^2 + \sum_{i \neq j} \sqrt{\mu_i \mu_j}\,(E_i E_j + E_j E_i) = \sum_{j=1}^{r} \mu_j E_j = A.$$

Para ver la unicidad de R, sea S otra matriz positiva con $S^2 = A$. Si $S = \nu_1 F_1 + \cdots + \nu_s F_s$ es la descomposición espectral de S, entonces

$$\nu_1^2 F_1 + \cdots + \nu_s^2 F_s = S^2 = A = \mu_1 E_1 + \cdots + \mu_r E_r,$$

así que $s = r$ es el número de autovalores distintos de A; los conjuntos $\{\nu_1^2, \dots, \nu_r^2\}$ y $\{\mu_1, \dots, \mu_r\}$ coinciden porque ellos son los autovalores de A. Después de una permutación de este conjunto, se puede suponer que $\nu_j^2 = \mu_j$ para $j = 1, \dots, r$, así que $\nu_j = \sqrt{\mu_j}$ para cada j. Finalmente, los autovectores de S y de A corresponden:

$$\{\, \boldsymbol{x} \in \mathbb{C}^n : F_j \boldsymbol{x} = \boldsymbol{x} \,\} = \ker(T_S - \nu_j I) = \ker(T_A - \mu_j I) = \{\, \boldsymbol{x} \in \mathbb{C}^n : E_j \boldsymbol{x} = \boldsymbol{x} \,\},$$

así que $F_j = E_j$ para cada j. Por lo tanto, es $S = R$. □

Definición 3.59. Un operador lineal $T \in \operatorname{End}(V)$ sobre un espacio hilbertiano es un **operador positivo** si T es autoadjunto y si $\langle \boldsymbol{x}, T(\boldsymbol{x}) \rangle \geq 0$ 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.

Por la Proposición anterior, aplicada a la matriz de T en cualquier base ortonormal de V, un operador T es [definido] positivos si y sólo si $T = T^*$ y todos sus autovalores son no negativos [resp., positivos]; si y sólo si $T = S^* S$ para algún operador $S \in \operatorname{End}(V)$; si y sólo si $T = Q^2$ para algún operador autojunto $Q \in \operatorname{End}(V)$; si y sólo si $T = R^2$ para algún operador positivo $R \in \operatorname{End}(V)$.

Esta raíz cuadrada positiva $R =: T^{1/2}$ es única, y vale $R = \sqrt{\mu_1}\, P_1 + \cdots + \sqrt{\mu_r}\, P_r$ si la descomposición espectral de T es $T = \mu_1 P_1 + \cdots + \mu_r P_r$.

Proposición 3.60. *Una matriz $U \in M_n(\mathbb{C})$ es* unitaria *si y sólo si U es normal y cada autovalor λ de U cumple $|\lambda| = 1$.*

Demostración. Si U es unitaria, entonces $U^* U = I_n = U U^*$, 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

$$|\alpha_1|^2 E_1 + \cdots + |\alpha_r|^2 E_r = U^* U = I_n = E_1 + \cdots + E_r.$$

Puesto que $E_iE_j = O$ 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 $U = \alpha_1E_1 + \cdots + \alpha_rE_r$ con cada $|\alpha_j| = 1$. Por tanto, vale

$$U^*U = UU^* = |\alpha_1|^2E_1 + \cdots + |\alpha_r|^2E_r = E_1 + \cdots + E_r = I_n,$$

lo cual muestra que U es unitaria. □

Definición 3.61. Un operador lineal $U \in \operatorname{End}(V)$ es una **isometría parcial** si $\underline{UU^*U = U}$.

Cada operador unitario es una isometría parcial, pero una isometría parcial con $\ker U \neq \{\boldsymbol{0}\}$ no es unitaria.[13] Como $U = UU^*U$ implica $U^* = (UU^*U)^* = U^*UU^*$, se ve que el operador adjunto U^* es también unitario.

Obsérvese que los operadores UU^* y U^*U son proyectores ortogonales. Es fácil comprobar que UU^* es el proyector cuya imagen es $U(V)$, mientras que U^*U es el proyector cuya imagen es $U^*(V) = (\ker U)^\perp$.

Notación. Si $T \in \operatorname{End}(V)$ es un operador lineal cualquiera, el operador T^*T es positivo. Escríbase

$$|T| := (T^*T)^{1/2}$$

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 sólo si T es un operador positivo.[14]

Resulta que cualquier operador lineal es el producto de una isometría parcial y un operador positivo (su módulo).

Teorema 3.62 (Descomposición Polar)**.** *Sea V un espacio hilbertiano y sea $T \in \operatorname{End}(V)$. Entonces hay una única isometría parcial $U \in \operatorname{End}(V)$ tal que $\underline{\ker U = \ker T}$ y $\underline{T = U|T|}$.*

El operador T es inversible si y sólo si U es unitaria y $|T|$ es definido positivo.

Demostración. Para todo $\boldsymbol{x} \in V$, vale

$$\langle |T|(\boldsymbol{x}), |T|(\boldsymbol{x})\rangle = \langle \boldsymbol{x}, |T|^2(\boldsymbol{x})\rangle = \langle \boldsymbol{x}, T^*T(\boldsymbol{x})\rangle = \langle T(\boldsymbol{x}), T(\boldsymbol{x})\rangle. \tag{3.18}$$

Por lo tanto, la correspondencia $|T|(\boldsymbol{x}) \mapsto T(\boldsymbol{x})$ define una aplicación lineal biyectiva de $|T|(V)$ en $T(V)$.

La ecuación (3.18) también muestra que $\ker|T| = \ker T$. De la Proposición 3.41, se obtiene $|T|(V)^\perp = \ker(|T|^*) = \ker|T|$, así que $|T|(V)^\perp = \ker T$.

[13] Sobre un espacio de Hilbert infinitodimensional, un operador U se llama una **isometría** si $U^*U = I$. Esto es equivalente a la condición de que $\|U(\boldsymbol{x})\| = \|\boldsymbol{x}\|$ para cada vector $\boldsymbol{x}$, porque $\|U(\boldsymbol{x})\|^2 = \langle U\boldsymbol{x}, U(\boldsymbol{x})\rangle = \langle \boldsymbol{x}, U^*U(\boldsymbol{x})\rangle$. En particular, una isometría es un operador inyectivo. En el contexto infinitodimensional, hay isometrías que no son sobreyectivos, luego no son inversibles; pero en espacios hilbertianos finitodimensionales, cualquier isometría es unitaria.

[14] En el caso unidimensional $V = \mathbb{C}$, cualquier elemento de $\operatorname{End}(\mathbb{C})$ es de la forma $w \mapsto zw$ para algún $z \in \mathbb{C}$. El módulo se obtiene al reemplazar z por su valor absoluto $|z| = \sqrt{\bar{z}z}$.

Defínase $U \in \mathrm{End}(V)$ por

$$U(\boldsymbol{y}) := \begin{cases} T(\boldsymbol{x}), & \text{si } \boldsymbol{y} = |T|(\boldsymbol{x}), \\ \boldsymbol{0}, & \text{si } \boldsymbol{y} \in \ker T. \end{cases} \tag{3.19}$$

Como $V = |T|(V) \oplus |T|(V)^{\perp} = |T|(V) \oplus \ker T$, el operador U queda bien definido por esta fórmula. Es inmediato de esta definición que $U|T| = T$.

Como $T(\boldsymbol{x}) = \boldsymbol{0}$ si y sólo si $|T|(\boldsymbol{x}) = \boldsymbol{0}$, se ve que $U(\boldsymbol{y}) = \boldsymbol{0}$ si y sólo si $\boldsymbol{y} \in \ker T$. Por lo tanto, es $\ker U = \ker |T| = \ker T$.

Si $\boldsymbol{x}, \boldsymbol{y} \in V$, hay $z \in V$, $\boldsymbol{w} \in \ker T$ tales que $\boldsymbol{x} = |T|(z) + \boldsymbol{w}$. Entonces

$$\begin{aligned} \langle \boldsymbol{x}, U^* U(|T|(\boldsymbol{y}))\rangle &= \langle U|T|(z) + U(\boldsymbol{w}), U|T|(\boldsymbol{y})\rangle = \langle U|T|(z), U|T|(\boldsymbol{y})\rangle \\ &= \langle T(z), T(\boldsymbol{y})\rangle = \langle z, T^*T(\boldsymbol{y})\rangle = \langle |T|(z), |T|(\boldsymbol{y})\rangle \\ &= \langle |T|(z) + \boldsymbol{w}, |T|(\boldsymbol{y})\rangle = \langle \boldsymbol{x}, |T|(\boldsymbol{y})\rangle. \end{aligned}$$

Luego $U^*U(|T|(\boldsymbol{y})) = |T|(\boldsymbol{y})$ para todo $\boldsymbol{y} \in V$. Por tanto, $|T|(V)$ es un subespacio invariante para el operador U^*U y que la restricción de U^*U a este subespacio es la identidad. Además,

$$UU^*U(\boldsymbol{x}) = UU^*U(|T|(z) + \boldsymbol{w}) = U(|T|(z) + U^*U(\boldsymbol{w})) = U(|T|(z)) = U(|T|(z) + \boldsymbol{w}) = U(\boldsymbol{x}).$$

Se concluye que $UU^*U = U$, es decir, U es una isometría parcial.

Si $T = W|T|$ para cualquier isometría parcial con $\ker W = \ker T$, entonces $W(|T|(\boldsymbol{x})) = T(\boldsymbol{x})$ para $\boldsymbol{x} \in V$ y $W(\boldsymbol{y}) = \boldsymbol{0}$ para $\boldsymbol{y} \in \ker T$, así que $W = U$.

Finalmente, T es inversible si y sólo si T^*T es inversible y $\ker T = \{\boldsymbol{0}\}$, si y sólo si T^*T es definido positivo y $\ker U = \{\boldsymbol{0}\}$, si y sólo si $|T|$ es definido positivo y la isometría parcial U es unitaria. □

Corolario 3.63. *Si $T \in \mathrm{End}(V)$, hay una única isometría parcial $W \in \mathrm{End}(V)$ que cumple $\underline{W(V) = T(V)}$ y $\underline{T = |T^*|W}$.*

Demostración. Si $T^* = U_1|T^*|$ es la descomposición polar de T^*, tómese $W := U_1^*$. Entonces $T = (T^*)^* = |T^*|^* U_1^* = |T^*|W$ porque el operador positivo $|T^*|$ es autoadjunto. Además, vale $W(V) = (\ker U_1)^{\perp} = (\ker T^*)^{\perp} = T(V)$.

La unicidad de la factorización $T = |T^*|W$ es consecuencia de la unicidad de la descomposición polar $T^* = W^*|T^*|$. □

Es inmediato de la fórmula (3.19) que $U(V) = T(V)$ cuando $T = U|T|$ es una descomposición polar.

Proposición 3.64. *Un operador lineal $T \in \mathrm{End}(V)$ es normal si y sólo si $|T^*| = |T|$ si y sólo si los factores de su descomposición polar conmutan: $T = U|T| = |T|U$.*

Se deja la demostración como un ejercicio. ⊟

► 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.

Proposición 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*

$$P^{-1}QP = P^tQP = R_{\theta_1} \oplus \cdots \oplus R_{\theta_t} \oplus (-I_r) \oplus I_s \tag{3.20}$$

*donde cada R_{θ_j} es una **rotación** de la forma*

$$R_\theta := \begin{bmatrix} \cos\theta & \operatorname{sen}\theta \\ -\operatorname{sen}\theta & \cos\theta \end{bmatrix}. \tag{3.21}$$

Demostración. La matriz ortogonal Q cumple $Q^tQ = QQ^t = I_n$ en $M_n(\mathbb{R})$. Al consider Q como elemento de $M_n(\mathbb{C})$ con entradas reales, Q es también unitaria en $M_n(\mathbb{C})$. Por la Proposicion 3.56 y 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 $T_Q = \mu_1 P_1 + \ldots + \mu_l P_l$, entonces cada λ_i pertenece a $\{\mu_1, \dots, \mu_l\}$.

Los polinomios característicos de Q y de Q^t coinciden, así que los autovalores de $Q^* = Q^t$ son también $\mu_1, \dots, \mu_l$. Un autovalor real 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 $z \in \mathbb{C}^n$ es un autovector de Q para un autovalor ± 1, sea $z =: \boldsymbol{x} + i\boldsymbol{y}$ con $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$. Si $\boldsymbol{y} = \boldsymbol{0}$, es $z = \boldsymbol{x} \in \mathbb{R}^n$; si $\boldsymbol{x} = \boldsymbol{0}$, entonces $(-i)z = \boldsymbol{y}$ es un autovector real para Q; y si $\boldsymbol{x} \neq \boldsymbol{0}$, $\boldsymbol{y} \neq \boldsymbol{0}$, entonces $\operatorname{lin}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ es un subespacio de autovectores de Q de dimensión 2 (sobre $\mathbb{R}$). En todo caso, los subespacios $\ker(T_Q + I) \subseteq \mathbb{R}^n$ y $\ker(T_Q - I) \subseteq \mathbb{R}^n$ poseen bases ortonormales de autovectores reales para los autovalores respectivos ∓ 1. Sean $r := n(T_Q + I)$, $s := n(T_Q - I)$ sus respectivas dimensiones.

Ahora sea $z \in \mathbb{C}^n$ un autovector para el autovalor complejo $e^{i\theta}$ con $0 < \theta < \pi$, con longitud $\|z\| = \sqrt{2}$. Escríbase $z =: \boldsymbol{x} + i\boldsymbol{y}$ con $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ y nótese que $\bar z = \boldsymbol{x} - i\boldsymbol{y}$ es un autovector para el autovalor $e^{-i\theta}$ porque $Qz = e^{i\theta} z$ conlleva $Q\bar z = e^{-i\theta}\bar z$. Ahora

$$e^{i\theta}\langle \bar z, z \rangle = \langle \bar z, Qz \rangle = \langle Q^* \bar z, z \rangle = \langle Q^{-1}\bar z, z \rangle = \langle e^{i\theta}\bar z, z \rangle = e^{-i\theta}\langle \bar z, z \rangle,$$

así que $\langle \bar z, z \rangle = 0$ porque $e^{i\theta} \neq e^{-i\theta}$. Por tanto, vale

$$0 = \langle \bar z, z \rangle = \langle \boldsymbol{x} - i\boldsymbol{y}, \boldsymbol{x} + i\boldsymbol{y} \rangle = \langle \boldsymbol{x}, \boldsymbol{x} \rangle - \langle \boldsymbol{y}, \boldsymbol{y} \rangle + 2i\langle \boldsymbol{x}, \boldsymbol{y} \rangle$$

y además $\langle \boldsymbol{x}, \boldsymbol{x} \rangle + \langle \boldsymbol{y}, \boldsymbol{y} \rangle = \langle z, z \rangle = 2$. Se concluye 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 *real* generado por z y $\bar z$. Al tomar partes reales e imaginarios de ecuación $Qz = e^{i\theta} z$, se obtiene

$$\begin{aligned} Q\boldsymbol{x} &= (\cos\theta)\,\boldsymbol{x} - (\operatorname{sen}\theta)\,\boldsymbol{y}, \\ Q\boldsymbol{y} &= (\operatorname{sen}\theta)\,\boldsymbol{x} + (\cos\theta)\,\boldsymbol{y}. \end{aligned}$$

Luego $\operatorname{lin}\langle \boldsymbol{x}, \boldsymbol{y} \rangle$ es un subespacio invariante para Q, en donde la restricción de T_Q posee la matriz R_θ de (3.21).

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 pareja $(e^{i\theta_k}, e^{-i\theta_k})$, 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 (3.20). La matriz de cambio de bases $P := [I]^{\mathcal{E}}_{\mathcal{V}}$ es ortogonal. $\square$

3.6 Ejercicios sobre ortogonalidad y teoría espectral

Ejercicio 3.1. Encontrar el complemento ortogonal $M^\perp \subset \mathbb{R}^3$ del subespacio $M = \text{lin}\langle \boldsymbol{x}_1, \boldsymbol{x}_2\rangle$ de $\mathbb{R}^3$ generado por

$$\boldsymbol{x}_1 := \begin{bmatrix} 1 \\ 1 \\ 2 \end{bmatrix}, \qquad \boldsymbol{x}_2 := \begin{bmatrix} 1 \\ 2 \\ 3 \end{bmatrix}.$$

[[Indicación: Sea $A \in \mathbb{R}^{2\times 3}$ la matriz cuyas filas son $\boldsymbol{x}_1^t$ y $\boldsymbol{x}_2^t$; resolver la ecuación $A\boldsymbol{x} = \boldsymbol{0}$.]]

Ejercicio 3.2. Encontrar una base ortonormal para el subespacio $W = \text{lin}\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\rangle$ de $\mathbb{R}^5$, 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 3.3. Sea $W := \text{lin}\langle \boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3\rangle$ 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 $w \in W$ tal que $\langle \boldsymbol{w}, \boldsymbol{x}\rangle = \langle \boldsymbol{w}, \boldsymbol{y}\rangle = 0$.

Ejercicio 3.4. En el espacio vectorial $\mathbb{R}[t]$ de polinomios reales, considérese el producto escalar

$$\langle f(t), g(t)\rangle := \int_{-1}^{1} f(x)g(x)\,dx.$$

El algoritmo de Gram y Schmidt, aplicado a la base $\{1, t, t^2, t^3, \dots\}$ de $\mathbb{R}[t]$, produce una familia ortonormal de polinomios $\{p_0(t), p_1(t), p_2(t), \dots\}$, donde cada $p_k(t)$ es un polinomio de grado k. Verificar que

$$p_0(t) = \frac{1}{\sqrt{2}}, \qquad p_1(t) = \frac{\sqrt{3}}{\sqrt{2}}t, \qquad p_2(t) = \frac{\sqrt{5}}{2\sqrt{2}}(3t^2 - 1), \qquad p_3(t) = \frac{\sqrt{7}}{2\sqrt{2}}(5t^3 - 3t).$$

Calcular el polinomio $p_4(t)$. Explicar por qué estos “polinomios de Legendre” son alternadamente funciones pares e impares de t.

Ejercicio 3.5. Usualmente, se define el **polinomio de Legendre** de grado n por la *fórmula de Rodrigues*:

$$P_n(t) := \frac{1}{2^n\, n!}\,\frac{d^n}{dt^n}((t^2 - 1)^n).$$

Con el mismo producto escalar del Ejercicio anterior, verificar que

$$\langle P_m(t), P_n(t)\rangle = 0 \quad \text{si} \quad m \neq n; \qquad \langle P_n(t), P_n(t)\rangle = \frac{2}{2n+1}.$$

[[Indicación: Integración por partes.]]

Comprobar que $\operatorname{lin}\langle P_0(t), P_1(t), \dots, P_n(t)\rangle = \operatorname{lin}\langle 1, t, \dots, t^n\rangle$ para $n \in \mathbb{N}$, por inducción sobre n. Concluir que los $p_n(t)$ del Ejercicio anterior cumplen $p_n(t) = \sqrt{(2n+1)/2}\, P_n(t)$.

Ejercicio 3.6. En el espacio vectorial $\mathbb{R}[t]$, considérese otro producto escalar

$$\langle\!\langle f(t), g(t)\rangle\!\rangle := \int_{-1}^{1} \frac{f(x)g(x)}{\sqrt{1-x^2}}\,dx.$$

El **polinomio de Chebyshev** $T_n(t)$ es aquél polinomio de grado n que verifica la identidad

$$T_n(\cos\theta) \equiv \cos(n\theta).$$

Demostrar que estos polinomios son ortogonales con respecto al producto escalar $\langle\!\langle \cdot, \cdot\rangle\!\rangle$.

Ejercicio 3.7. 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 3.8. Decidir (con razonamiento) si la siguiente matriz es ortogonal o no:

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

Ejercicio 3.9. Demostrar que la siguiente matriz es una matriz ortogonal:

$$A = \begin{bmatrix} \operatorname{sen}\theta\operatorname{sen}\phi & \operatorname{sen}\theta\cos\phi & \cos\theta\cos\phi & -\cos\theta\operatorname{sen}\phi \\ \operatorname{sen}\theta\cos\phi & -\operatorname{sen}\theta\operatorname{sen}\phi & \cos\theta\operatorname{sen}\phi & \cos\theta\cos\phi \\ \cos\theta\operatorname{sen}\phi & \cos\theta\cos\phi & -\operatorname{sen}\theta\cos\phi & \operatorname{sen}\theta\operatorname{sen}\phi \\ \cos\theta\cos\phi & -\cos\theta\operatorname{sen}\phi & -\operatorname{sen}\theta\operatorname{sen}\phi & -\operatorname{sen}\theta\cos\phi \end{bmatrix}.$$

Ejercicio 3.10. (a) Si $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^n$ son vectores de columna no nulos, demostrar que $\boldsymbol{x}\boldsymbol{y}^t$ 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** determinado por $\boldsymbol{x}$ es

$$H_{\boldsymbol{x}} := I_n - 2\boldsymbol{x}\boldsymbol{x}^t.$$

Demostrar que $H_{\boldsymbol{x}}$ es simétrica y ortogonal, y que $H_{\boldsymbol{x}}^2 = I_n$. [[Indicación: Es $\boldsymbol{x}^t\boldsymbol{x} = \|\boldsymbol{x}\|^2$.]]

Ejercicio 3.11. (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 una matriz unitaria, demostrar que U es inversible y que U^{-1} es también unitaria.[15]

(c) Demostrar que $A \in M_n(\mathbb{C})$ es unitaria si y sólo si $\|A\boldsymbol{x}\| = \|\boldsymbol{x}\|$ para todo $\boldsymbol{x} \in \mathbb{C}^n$.

Ejercicio 3.12. (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 sean autovectores de A, de modo que $P^tAP = P^{-1}AP$ sea diagonal.

(b) Calcular A^5, usando esta forma diagonal $D = P^tAP$.

Ejercicio 3.13. (a) Sea $A := \begin{bmatrix} a & b \\ b & c \end{bmatrix}$ una matriz simétrica en $M_2(\mathbb{R})$. Demostrar que es posible factorizar $A = LDL^t$, con L triangular inferior unipotente y D diagonal inversible, si y sólo si $a \neq 0$ y $c \neq b^2/a$.

(b) Concluir que A es definida positiva si y sólo si $a > 0$, $c > 0$ y $ac - b^2 > 0$.

(c) Obtener la factorización $A = LDL^t$ para $A := \begin{bmatrix} 4 & 12 \\ 12 & 45 \end{bmatrix}$. Usar esta factorización para comprobar que $\boldsymbol{x}^tA\boldsymbol{x} = (2x_1 + 6x_2)^2 + (3x_2)^2$ si $\boldsymbol{x} = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \in \mathbb{R}^2$.

Ejercicio 3.14. 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 $A \in M_n(\mathbb{R})$ es definida positiva, demostrar que todos sus menores principales delanteras son números positivos.

(b) Inversamente, si $A \in M_n(\mathbb{R})$ es tal que $D_k > 0$ para $k = 1, 2, \dots, n$, demostrar que A es definida positiva. ⟦ Indicación: Eliminación gaussiana simple. ⟧

Ejercicio 3.15. Determinar si cada una de las matrices

$$A := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \qquad B := \begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 2 \\ 1 & 2 & 3 \end{bmatrix}, \qquad C := \begin{bmatrix} 9 & -6 & 2 \\ -6 & 8 & -4 \\ 2 & -4 & 4 \end{bmatrix}$$

es (a) positiva; (b) definida positiva.

[15] Las partes (a) y (b) dicen que la totalidad de matrices unitarias $n \times n$ es un *grupo*, llamado $U(n)$.

Ejercicio 3.16. Considérese la siguiente matriz $A \in M_8(\mathbb{R})$:

$$A := \begin{bmatrix} b & 0 & -1 & 0 & 0 & 0 & 0 & 1 \\ 0 & b & 0 & -1 & 0 & 0 & -1 & 0 \\ -1 & 0 & c & 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & 0 & c & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & b & 0 & 1 & 0 \\ 0 & 0 & -1 & 0 & 0 & b & 0 & 1 \\ 0 & -1 & 0 & 0 & 1 & 0 & c & 0 \\ 1 & 0 & 0 & 0 & 0 & 1 & 0 & c \end{bmatrix}. \tag{3.22}$$

Demostrar que esta matriz es definida positiva si y sólo si $b > 0$, $c > 0$ y $bc > 4$. ⟦ Esta matriz apareció en un problema de la mecánica cuántica.[16] ⟧

Ejercicio 3.17. Dadas m vectores $\boldsymbol{x}_1, \dots, \boldsymbol{x}_m \in \mathbb{R}^n$, su **determinante de Gram** es

$$\det A = \begin{vmatrix} \boldsymbol{x}_1 \cdot \boldsymbol{x}_1 & \boldsymbol{x}_1 \cdot \boldsymbol{x}_2 & \dots & \boldsymbol{x}_1 \cdot \boldsymbol{x}_m \\ \boldsymbol{x}_2 \cdot \boldsymbol{x}_1 & \boldsymbol{x}_2 \cdot \boldsymbol{x}_2 & \dots & \boldsymbol{x}_2 \cdot \boldsymbol{x}_m \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{x}_m \cdot \boldsymbol{x}_1 & \boldsymbol{x}_m \cdot \boldsymbol{x}_2 & \dots & \boldsymbol{x}_m \cdot \boldsymbol{x}_m \end{vmatrix}$$

el cual 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 sólo 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 $A = B^t B$. ⟧

Ejercicio 3.18. Sea $\zeta := e^{2\pi i/n}$ una *raíz n-ésima de* 1 y sea $F \in M_n(\mathbb{C})$ la matriz con entrada $f_{ij} := \zeta^{(i-1)(j-1)}$:

$$F := \frac{1}{\sqrt{n}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \zeta & \zeta^2 & \dots & \zeta^{n-1} \\ 1 & \zeta^2 & \zeta^4 & \dots & \zeta^{2(n-1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \zeta^{n-1} & \zeta^{2(n-1)} & \dots & \zeta^{(n-1)^2} \end{bmatrix}.$$

La matriz F se llama la **transformación de Fourier finita** (TFF) de orden n.

(a) Escribir F explícitamente en los casos $n = 2, 3, 4$.

(b) Demostrar que $F^* F = I_n$ y concluir que la matriz F es unitaria.

(c) Calcula la matriz F^2 y mostrar que $F^4 = I_n$. Concluir que los únicos autovalores posibles[17] para F son $\lambda = 1, -1, i, -i$.

[16] Referencia: Ileana Castillo Arias, *Productos cuánticos en espacios de funciones analíticas*, tesis de licenciatura, UCR, 1988.

[17] La multiplicidad de cada uno de estos autovalores, para n cualquiera, es "aproximadamente $\frac{1}{4}n$". 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.

Ejercicio 3.19. (a) Sea V un espacio hilbertiano. Si $S \in \mathrm{End}(V)$ es un operador autoadjunto tal que $\langle \boldsymbol{x}, S(\boldsymbol{x})\rangle = 0$ para todo $\boldsymbol{x} \in V$, demostrar que $S = 0$.

(b) Si además $T \in \mathrm{End}(V)$ es un operador lineal cualquiera, demostrar que $T = T^*$ si y sólo si $\langle \boldsymbol{x}, T(\boldsymbol{x})\rangle \in \mathbb{R}$ para todo $\boldsymbol{x} \in V$. [[Indicación: Considerar $S := i(T^* - T)$.]]

Ejercicio 3.20. (a) Sea $S \in \mathrm{End}(V)$ un operador autoadjunto. Demostrar que hay dos operadores positivos $S_+, S_- \in \mathrm{End}(V)$ tales que $S_+S_- = S_-S_+ = 0$ y $S = S_+ - S_-$.

(b) Concluir que cualquier operador lineal $T \in \mathrm{End}(V)$ es una combinación lineal de la forma $T = T_1 - T_2 + iT_3 - iT_4$ donde T_1, T_2, T_3, T_4 son operadores positivos.

Ejercicio 3.21. (a) Si $T \in \mathrm{End}(V)$ es un operador normal, demostrar que hay un polinomio $f(t) \in \mathbb{C}[t]$ tal que $f(T) = T^*$.

[[Indicación: Buscar un polinomio que cumple $f(\mu_j) = \bar{\mu}_j$ para cada autovalor μ_j de T.]]

(b) Si $S \in \mathrm{End}(V)$ es otro operador lineal (no necesariamente normal) tal que $ST = TS$, demostrar que $S^*T = TS^*$.

(c) Si $S, T \in \mathrm{End}(V)$ son operadores normales, concluir que el producto ST es también un operador normal.

Ejercicio 3.22. (a) Si $T \in \mathrm{End}(V)$ es un operador positivo, demostrar que hay un polinomio $g(t) \in \mathbb{R}[t]$ tal que $g(T) = T^{1/2}$.

[[Indicación: Buscar un polinomio que cumple $g(\mu_j) = \sqrt{\mu_j}$ para cada autovalor μ_j de T.]]

(b) Si $S \in \mathrm{End}(V)$ es otro operador positivo tal que $ST = TS$, demostrar que los operadores $S + T$ y ST son también positivos.

(c) Si $P, Q \in \mathrm{End}(V)$ son operadores positivos que no conmutan, demostrar que $P + Q$ es positivo pero que PQ no es necesariamente positivo.

[[Indicación: Dar un contraejemplo de dos matrices positivas cuyo producto no es una matriz positiva.]]

Ejercicio 3.23. 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:

$$\mathrm{tr}(A^*A) \geq |\lambda_1|^2 + \cdots + |\lambda_n|^2,$$

con igualdad si y sólo si A es una matriz normal.

Ejercicio 3.24. (a) Demostrar que un operador lineal $U \in \mathrm{End}(V)$ es unitario si y sólo si $\|U(\boldsymbol{x})\| = \|\boldsymbol{x}\|$ para todo $\boldsymbol{x} \in V$.

(b) Demostrar que un operador lineal $T \in \mathrm{End}(V)$ es normal si y sólo si $\|T(\boldsymbol{x})\| = \|T^*(\boldsymbol{x})\|$ para todo $\boldsymbol{x} \in V$.

Ejercicio 3.25. Obtener la descomposición polar $U|A|$ de la matriz $A = \begin{bmatrix} -1 & -2 \\ 2 & 1 \end{bmatrix} \in M_2(\mathbb{C})$.

4 Formas Bilineales

Las aplicaciones lineales entre dos espacios vectoriales no son los únicos objetos que pueden representarse por matrices. En este capítulo se examinará otra clase de funciones, las *formas bilineales* (o *sesquilineales*) sobre un espacio vectorial, que son una generalización del concepto de producto escalar. A cada forma bilineal se le asocia una matriz que la representa con respecto a una base particular.

Se distinguen tres clases importantes de estas formas: (a) las formas bilineales simétricas, (b) las formas bilineales alternadas o antisimétricas, y (c) las formas sesquilineales hermíticas (en el caso complejo). A cada forma se le asocia una matriz cuadrada, respectivamente simétrica, antisimétrica, o hermítica. Se busca una clasificación de estas formas hasta isomorfismo, o lo que es equivalente, la identificación de ciertos tipos de matrices que permiten distinguir entre formas inequivalentes.

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 $\boldsymbol{x} \mapsto f(\boldsymbol{x},\boldsymbol{y})$ y $\boldsymbol{y} \mapsto f(\boldsymbol{x},\boldsymbol{y})$ son formas lineales sobre $\mathbb{F}$; en otras palabras,

$$\begin{aligned} f(\boldsymbol{x}+\boldsymbol{z},\boldsymbol{y}) &= f(\boldsymbol{x},\boldsymbol{y}) + f(\boldsymbol{z},\boldsymbol{y}), \\ f(\boldsymbol{x},\boldsymbol{y}+\boldsymbol{w}) &= f(\boldsymbol{x},\boldsymbol{y}) + f(\boldsymbol{x},\boldsymbol{w}), \\ f(c\boldsymbol{x},\boldsymbol{y}) &= f(\boldsymbol{x},c\boldsymbol{y}) = c\,f(\boldsymbol{x},\boldsymbol{y}), \end{aligned}$$

para todo $\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{w} \in V$, $c \in \mathbb{F}$. En consecuencia,

$$f\Big(\sum_{i=1}^{n} c_i\boldsymbol{x}_i, \sum_{j=1}^{n} d_j\boldsymbol{x}_j\Big) = \sum_{i=1}^{n}\sum_{j=1}^{n} c_i d_j\, f(\boldsymbol{x}_i,\boldsymbol{x}_j). \tag{4.1}$$

Ejemplo 4.2. El producto punto de vectores $f(\boldsymbol{x},\boldsymbol{y}) := \underline{\boldsymbol{x}^t\boldsymbol{y}} = x_1y_1 + \cdots + x_ny_n$ es una forma bilineal sobre $\mathbb{F}^n$.

Más generalmente, si $A \in M_n(\mathbb{F})$ es una matriz cuadrada cualquiera, la receta

$$f(\boldsymbol{x},\boldsymbol{y}) := \boldsymbol{x}^t A\boldsymbol{y}$$

define una forma bilineal sobre $\mathbb{F}^n$.

Definición 4.3. Es evidente de (4.1) que f queda determinada por los valores $f(\boldsymbol{x}_i,\boldsymbol{x}_j)$ en una *base* $\mathcal{B} = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ del espacio vectorial V. Denótese estos **coeficientes** de f por

$$a_{ij} := f(\boldsymbol{x}_i,\boldsymbol{x}_j), \quad \text{para todo} \quad i,j = 1,\dots,n. \tag{4.2}$$

Entonces $f(\sum_i c_i\boldsymbol{x}_i, \sum_j d_j\boldsymbol{x}_j) = \sum_{i,j} c_i a_{ij} d_j$. 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 $\mathcal{B}$.

Para cada vector fijo $\boldsymbol{y} \in V$, la aplicación $\boldsymbol{x} \mapsto f(\boldsymbol{x}, \boldsymbol{y})$ es una forma lineal sobre V, es decir, un elemento $f_{\boldsymbol{y}}$ del espacio dual V^*. Por (4.1), si $\boldsymbol{y} = \sum_{j=1}^n d_j \boldsymbol{x}_j$ entonces $f_{\boldsymbol{y}}(\boldsymbol{x}) = \sum_{j=1}^n d_j f_{\boldsymbol{x}_j}(\boldsymbol{x})$ para todo $\boldsymbol{x} \in V$, y por ende $F \colon \boldsymbol{y} \mapsto f_{\boldsymbol{y}}$ es una aplicación lineal en $\mathcal{L}(V, V^*)$. Obsérvese 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^* que es dual a la base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ de V, entonces la matriz (1.6) de F, con respecto a estas bases de V y V^*, se obtiene de

$$F(\boldsymbol{x}_j) = f_{\boldsymbol{x}_j} = \sum_{i=1}^n f_{\boldsymbol{x}_j}(\boldsymbol{x}_i)\, f_i = \sum_{i=1}^n a_{ij} f_i.$$

Luego $[f]_{\mathcal{B}}^{\mathcal{B}} = A = [F]_{\mathcal{B}}^{\mathcal{B}^*}$. 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 $\mathcal{B}$ y su base dual.

De esta manera, se ve que las formas bilineales sobre V conforman un espacio vectorial, isomorfo al espacio vectorial $\mathcal{L}(V, V^*)$. Su dimensión es $(\dim V)(\dim V^*) = (\dim V)^2 = n^2 = \dim(M_n(\mathbb{F}))$.

Proposición 4.4. *Sea $f \colon V \times V \to \mathbb{F}$ es una forma bilineal y sean $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ y $\mathcal{B}' = \{\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_n\}$ dos bases de V. Sea P la matriz de cambio de base* (1.9)*, dado por $\boldsymbol{x}'_s =: \sum_{j=1}^n p_{js} \boldsymbol{x}_j$. Entonces las matrices respectivas $A = [f]_{\mathcal{B}}^{\mathcal{B}}$ y $B = [f]_{\mathcal{B}'}^{\mathcal{B}'}$ de la forma bilineal f cumplen*

$$B = P^t A P.$$

Demostración. La matriz A viene de (4.2) y las entradas de B obedecen $b_{rs} := f(\boldsymbol{x}'_r, \boldsymbol{x}'_s)$. En vista de (4.1), vale

$$b_{rs} = f\Big(\sum_{i=1}^n p_{ir} \boldsymbol{x}_i, \sum_{j=1}^n p_{js} \boldsymbol{x}_j\Big) = \sum_{i=1}^n \sum_{j=1}^n p_{ir}\, f(\boldsymbol{x}_i, \boldsymbol{x}_j)\, p_{js} = \sum_{i=1}^n \sum_{j=1}^n p_{ir} a_{ij} p_{js},$$

y se reconoce el lado derecho como la entrada (r, s) de la matriz $P^t A P$. □

Definición 4.5. Dos matrices cuadradas $A, B \in M_n(\mathbb{F})$ son **matrices congruentes**, escrito $A \doteqdot B$, si hay una matriz inversible $P \in M_n(\mathbb{F})$ tal que $\underline{B = P^t A P}$.

La igualdad $(P^{-1})^t = (P^t)^{-1}$ implica que la congruencia de matrices es una relación de equivalencia.[1] En efecto, esta relación es reflexiva, porque $A = I_n^t A I_n$; es antisimétrica, porque $B = P^t A P$ implica $A = (P^{-1})^t B P^{-1}$; y es transitiva, porque $B = P^t A P$ y $C = Q^t B Q$ implican $C = (PQ)^t A (PQ)$. La Proposición 4.4 dice que dos matrices que representan la misma forma bilineal respecto de dos bases distintas son congruentes, mediante la matriz P de cambio de base.

Obsérvese que dos matrices congruentes tienen el mismo rango: se sabe que el rango es invariante bajo cambios $A \mapsto QAP$ con Q, P inversibles; al tomar $Q = P^t$, se ve que $r(A) = r(B)$ cuando $A \doteqdot B$. En consecuencia, el rango de la matriz de una forma bilineal no depende de la base $\mathcal{B}$ de V: se puede hablar del **rango de la forma bilineal** f.

[1] Una minoría de autores escriben A^{-t} para denotar la matriz $(A^{-1})^t = (A^t)^{-1}$, por un abuso de las leyes de exponentes. Fíjese que $(AB)^{-t} = A^{-t} B^{-t}$ si A, B son inversibles. A^{-t} se llama la **matriz contragrediente** de A.

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;

(d) la aplicación lineal asociada $F\in\mathcal{L}(V,V^*)$ es sobreyectiva;

(e) si $\mathcal{B}$ es una base de V, la matriz $A=[f]^{\mathcal{B}}_{\mathcal{B}}$ es inversible.

Demostración. Las equivalencias (a) $\iff$ (b) $\iff$ (c) son consecuencias directas de las definiciones de $f_{\boldsymbol{y}}$ y F. Para (c) $\iff$ (d), fíjese que F es inyectiva si y sólo si es sobreyectiva, porque $\dim V=\dim V^*$. La equivalencia (d) $\iff$ (e) se debe a que $[f]^{\mathcal{B}}_{\mathcal{B}}=[F]^{\mathcal{B}^*}_{\mathcal{B}}$. □

La sobreyectividad $F(V)=V^*$, es decir, $\{f_{\boldsymbol{y}}:\boldsymbol{y}\in V\}=V^*$, dice que una forma bilineal f es no degenerada si y sólo si cualquier elemento de V^* es dada 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 $d\colon V\times V\to\mathbb{F}$ 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 $s\colon V\times V\to\mathbb{F}$ es **alternada** o bien **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 sólo si su matriz $A=[d]^{\mathcal{B}}_{\mathcal{B}}$ cumple $\underline{A^t=A}$; y que una forma bilineal s es alternada si y sólo si su matriz $B=[s]^{\mathcal{B}}_{\mathcal{B}}$ cumple $\underline{B^t=-B}$.

Lema 4.9. *Cada forma bilineal $f\colon V\times V\to\mathbb{F}$ es la suma, de manera única, de una forma bilineal simétrica y una forma bilineal alternada.*

Demostración. Escríbase

$$d(\boldsymbol{x},\boldsymbol{y}):=\tfrac12\big(f(\boldsymbol{x},\boldsymbol{y})+f(\boldsymbol{y},\boldsymbol{x})\big)\quad\text{y}\quad s(\boldsymbol{x},\boldsymbol{y}):=\tfrac12\big(f(\boldsymbol{x},\boldsymbol{y})-f(\boldsymbol{y},\boldsymbol{x})\big).$$

Es evidente que d es simétrica y s es alternada —aunque cualquiera de las dos podría ser idénticamente nula— y que $f(\boldsymbol{x},\boldsymbol{y})=d(\boldsymbol{x},\boldsymbol{y})+s(\boldsymbol{x},\boldsymbol{y})$.

Para la unicidad de d y s, es suficiente observar que si $f=d'+s'$ es otra suma similar, entonces $d-d'=s'-s$; esta es una igualdad entre una forma simétrica y otra alternada, lo cual solo es posible[2] si ambas formas son nulas: $d-d'=s'-s\equiv 0$; luego, $d'=d$ y $s'=s$. □

[2]Hay una excepción a esta afirmación, si en el cuerpo de escalares $\mathbb{F}$ vale $-1=+1$, o equivalentemente, si $1+1=0$. El ejemplo más conocido es el cuerpo de dos elementos $\mathbb{F}_2=\{\bar0,\bar1\}$ que es la base de la aritmética binaria. En tales casos, la distinción entre simétrica y antisimétrica carece de sentido. Para evitar esta excepción, en este capítulo se asume implícitamente que $1+1\neq 0$ en $\mathbb{F}$.

Definición 4.10. Sea $d\colon V \times V \to \mathbb{F}$ una forma bilineal simétrica. A cada subespacio $M \leq V$ le corresponde un **subespacio ortogonal** con respecto a d:

$$\begin{aligned} M^\perp &= \{\, \boldsymbol{x} \in V : d(\boldsymbol{x},\boldsymbol{y}) = 0 \text{ para todo } \boldsymbol{y} \in M \,\} \\ &= \{\, \boldsymbol{y} \in V : d(\boldsymbol{x},\boldsymbol{y}) = 0 \text{ para todo } \boldsymbol{x} \in M \,\}. \end{aligned} \tag{4.3}$$

Es claro que $N \leq M \leq V$ implica $M^\perp \leq N^\perp$ y que $\{\boldsymbol{0}\}^\perp = V$.

El subespacio $V^\perp = \{\, \boldsymbol{y} \in V : d_{\boldsymbol{y}} = 0 \,\} = \ker D$ es el **núcleo** de la forma bilineal d. (Aquí D es la aplicación en $\mathcal{L}(V,V^*)$ asociada con la forma d.) Fíjese que d es no degenerada si y sólo si $V^\perp = \{\boldsymbol{0}\}$.

Proposición 4.11. *Si d es una forma bilineal simétrica no degenerada sobre V, y si M es un subespacio de V tal que $M \cap M^\perp = \{\boldsymbol{0}\}$, 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 (4.3) se ve que $M^\perp = \{\, \boldsymbol{y} \in V : d'_{\boldsymbol{y}} = 0 \,\} = \ker D_M$.

Por otro lado, si $\{g_1,\dots,g_m\}$ es una base de M^* y $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_m\}$ es la base dual de M, se puede completar ésta a una base $\mathcal{B} = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ de V y obtener la base dual $\mathcal{B}^* = \{f_1,\dots,f_n\}$ de V^*. Para $j = 1,\dots,m$ la regla $f_j(\boldsymbol{x}_k) = [\![j{=}k]\!]$ 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 concluye 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 $\boldsymbol{x} \in M \cap M^\perp = \{\boldsymbol{0}\}$ ahora implica $\dim V = \dim(M \oplus M^\perp)$, así que $M \oplus M^\perp = V$. □

Definición 4.12. Sea $d\colon V \times V \to \mathbb{F}$ una forma bilineal simétrica. La **forma cuadrática** asociada con 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 dado por $q(\boldsymbol{x}) = \sum_{i,j=1}^n a_{ij} x_i x_j$, donde $A \in M_n(\mathbb{F})$ es una matriz de coeficientes. Al reemplazar a_{ij} por $\frac{1}{2}(a_{ij} + a_{ji})$ si fuera necesario, se puede suponer que $A^t = A$. Entonces $q(\boldsymbol{x}) = d(\boldsymbol{x},\boldsymbol{x})$, 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 $q(x,y,z) = 1$, donde

$$q(x,y,z) := ax^2 + 2hxy + by^2 + 2gxz + 2fyz + cz^2 = \begin{bmatrix} x & y & z \end{bmatrix} \begin{bmatrix} a & h & g \\ h & b & f \\ g & f & c \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix}.$$

Proposición 4.15. *Una forma bilineal simétrica d es determinada por su forma cuadrática asociada, por* polarización*:*

$$d(\boldsymbol{x},\boldsymbol{y}) = \tfrac{1}{2}\bigl(q(\boldsymbol{x}+\boldsymbol{y}) - q(\boldsymbol{x}) - q(\boldsymbol{y})\bigr). \tag{4.4}$$

Demostración. Por cálculo directo, vale

$$\begin{aligned} q(\boldsymbol{x}+\boldsymbol{y}) - q(\boldsymbol{x}) - q(\boldsymbol{y}) &= d(\boldsymbol{x}+\boldsymbol{y},\boldsymbol{x}+\boldsymbol{y}) - d(\boldsymbol{x},\boldsymbol{x}) - d(\boldsymbol{y},\boldsymbol{y}) \\ &= d(\boldsymbol{x},\boldsymbol{y}) + d(\boldsymbol{y},\boldsymbol{x}) = 2\,d(\boldsymbol{x},\boldsymbol{y}). \end{aligned}$$ □

► La clasificación de las formas cuadráticas, o lo que es o mismo, la clasificación de las formas bilineales simétricas, procede por convertir una matriz A en cierta matriz diagonal congruente con A. En contraste con la diagonalización por semejanza de una matriz real simétrica (mediante el teorema espectral), que requiere averiguar los autovalores de la matriz, la diagonalización por congruencia es más sencilla.

Proposición 4.16. *Sea $d\colon V\times V\to\mathbb{F}$ una forma bilineal simétrica, de rango $\underline{k}$. Entonces hay una base $\mathcal{B}=\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k,\boldsymbol{z}_1,\dots,\boldsymbol{z}_{n-k}\}$ 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 $A=O$ y $k=0$ por ser $d\equiv 0$, tómese una base $\{\boldsymbol{z}_1,\dots,\boldsymbol{z}_n\}$ cualquiera. Si $d\not\equiv 0$, entonces por (4.4) hay un vector $\boldsymbol{x}_1\in V$ con $q(\boldsymbol{x}_1)\neq 0$; colóquese $b_1:=q(\boldsymbol{x}_1)$.

Supóngase, para argumentar por inducción, que ya se haya elegido r vectores linealmente independientes $\boldsymbol{x}_1,\dots,\boldsymbol{x}_r\in V$ tales que $d(\boldsymbol{x}_i,\boldsymbol{x}_j)=b_i\,[\![i=j]\!]$ con $b_i\neq 0$, para $i,j=1,\dots,r$. Entonces la restricción de d al subespacio $M_r:=\operatorname{lin}\langle\boldsymbol{x}_1,\dots,\boldsymbol{x}_r\rangle$ posee una matriz inversible y por ende esta restricción es no degenerada. En consecuencia, vale $M_r\cap M_r^{\perp}=\{\boldsymbol{0}\}$.

Si d fuera no degenerada (es decir, si $k=n$), la Proposición 4.11 permitiría concluir que $V=M_r\oplus M_r^{\perp}$. Resulta que esta relación es válida aun para $k<n$. Para $\boldsymbol{x}\in V$, colóquese

$$\boldsymbol{y}:=\boldsymbol{x}-\sum_{i=1}^{r}\frac{d(\boldsymbol{x},\boldsymbol{x}_i)}{b_i}\boldsymbol{x}_i.$$

Entonces

$$d(\boldsymbol{y},\boldsymbol{x}_j)=d(\boldsymbol{x},\boldsymbol{x}_j)-\sum_{i=1}^{r}\frac{d(\boldsymbol{x},\boldsymbol{x}_i)}{b_i}b_i\,[\![i=j]\!]=d(\boldsymbol{x},\boldsymbol{x}_j)-d(\boldsymbol{x},\boldsymbol{x}_j)=0\quad\text{para}\quad j=1,\dots,r,$$

así que $\boldsymbol{y}\in M_r^{\perp}$. Por tanto,

$$\boldsymbol{x}=\sum_{i=1}^{r}b_i^{-1}d(\boldsymbol{x},\boldsymbol{x}_i)\boldsymbol{x}_i+\boldsymbol{y}\in M_r\oplus M_r^{\perp}.$$

Como $\boldsymbol{x}$ es arbitrario, se concluye que $V=M_r\oplus M_r^{\perp}$.

La matriz de d con respecto a esta descomposición de V es evidentemente de la forma $\begin{bmatrix} B_r & O \\ O & C_{n-r}\end{bmatrix}$, donde $B_r=\operatorname{diag}[b_1,\dots,b_r]$ y $C_{n-r}\in M_{n-r}(\mathbb{F})$.

Si $C_{n-r} \neq 0$ y por ende $r < k$, la restricción de d al subespacio $M_r^{\perp}$ no es nula y existe $\boldsymbol{x}_{r+1} \in M_r^{\perp}$ tal que $b_{r+1} := q(\boldsymbol{x}_{r+1}) \neq 0$. Ahora $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_{r+1}\}$ es la base de un subespacio M_{r+1} de V 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 al B_k, en cuyo caso $C_{n-k} = O$ 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 $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k, z_1, \dots, z_{n-k}\}$ de V. □

Corolario 4.17. *Una forma cuadrática sobre V de rango k puede escribirse como*

$$q(\boldsymbol{x}) = b_1 f_1(\boldsymbol{x})^2 + b_2 f_2(\boldsymbol{x})^2 + \cdots + b_k f_k(\boldsymbol{x})^2, \tag{4.5}$$

con formas lineales $f_1, \dots, f_k \in V^*$ *linealmente independientes y coeficientes* $b_1, \dots, b_k \in \mathbb{F}$.

Demostración. Por la demostración de la Proposición 4.16, es $q(\boldsymbol{x}) = b_1 \xi_1^2 + \ldots + b_k \xi_k^2$, con respecto a una base apropiada $\mathcal{B}$ de V. Sea $\mathcal{B}^* = \{f_1, \dots, f_n\}$ la base dual de V, así que $\xi_j = f_j(\boldsymbol{x})$ para todo $\boldsymbol{x} \in V$. De ahí resulta que $q(\boldsymbol{x}) = \sum_{j=1}^{k} b_j f_j(\boldsymbol{x})^2$. □

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 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 $t_j \mapsto x_j$ logran 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 (4.5) expresa q en términos de un juego de formas *lineales* $f_1, \dots, f_k \in (\mathbb{F}^n)^*$. Cada f_i es explícitamente $f_i(\boldsymbol{x}) =: \sum_{j=1}^{n} c_{ij} x_j$.

Defínanse unos polinomios de primer grado en n variables, sin términos constantes, también denotados f_i, por $f_i(t_1, \dots, t_n) := \sum_{j=1}^{n} c_{ij} t_j$. Entonces

$$q(t_1, \dots, t_n) = b_1 f_1(t_1, \dots, t_n)^2 + \cdots + b_k f_k(t_1, \dots, t_n)^2.$$ □

Fíjese que estos polinomios de primer grado pueden elegirse de modo que el polinomio f_i depende solamente de las variables $t_i, \dots, t_n$. Esto es consecuencia 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 $x_1, \dots, x_r$ 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 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 $\boldsymbol{e}_1 = (1,0,0,0)$, es $b_1 = q(\boldsymbol{e}_1) = 1$. Defínase $f_1(\boldsymbol{x}) := \underline{b_1^{-1}d(\boldsymbol{x},\boldsymbol{e}_1)} = x_1 - 2x_2 + x_3 - x_4$. Se ve que

$$q(\boldsymbol{x}) - b_1 f_1(\boldsymbol{x})^2 = -x_2^2 - x_3^2 + x_4^2 - 2x_2x_3 + 4x_2x_4 + 4x_3x_4 =: q'(\boldsymbol{x}),$$

Esta es una nueva forma cuadrática q' que no incluye la coordenada x_1. Hasta ahora se ha identificado $M_1 = \operatorname{lin}\langle \boldsymbol{e}_1\rangle$, $M_1^\perp = \operatorname{lin}\langle \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4\rangle$.

Con $\boldsymbol{e}_2 = (0,1,0,0)$, es $b_2 = q'(\boldsymbol{e}_2) = -1$. Defínase $f_2(\boldsymbol{x}) := \underline{b_2^{-1}d'(\boldsymbol{x},\boldsymbol{e}_2)} = x_2 + x_3 - 2x_4$. Ahora

$$q(\boldsymbol{x}) - b_1 f_1(\boldsymbol{x})^2 - b_2 f_2(\boldsymbol{x})^2 = 5x_4^2 =: q''(\boldsymbol{x}),$$

con $M_1 = \operatorname{lin}\langle \boldsymbol{e}_1, \boldsymbol{e}_2\rangle$, $M_1^\perp = \operatorname{lin}\langle \boldsymbol{e}_3, \boldsymbol{e}_4\rangle$. Ahora se toma $\boldsymbol{x}_3 := \boldsymbol{e}_4 = (0,0,0,1)$, $b_3 = q''(\boldsymbol{e}_4) = 5$, para obtener, finalmente:

$$q(\boldsymbol{x}) = (x_1 - 2x_2 + x_3 - x_4)^2 - (x_2 + x_3 - 2x_4)^2 + 5x_4^2.$$

La identificación de unas formas lineales $f_1, \dots, f_k$ que satisfacen la fórmula (4.5) puede hacerse en forma algorítmica, mediante un proceso conocido como la **reducción de Lagrange**, detallado a continuación.

Proposición 4.20. *La forma cuadrática sobre* $\mathbb{F}^n$ *dada por* $q(\boldsymbol{x}) = \boldsymbol{x}^t A\boldsymbol{x} = \sum_{i,j=1}^n a_{ij}x_ix_j$ *con una matriz simétrica* $A = A^t \in M_n(\mathbb{F})$ *puede expresarse en el formato de* (4.5) *mediante el siguiente algoritmo.*

(a) Si $a_{kk} \neq 0$ *para algún k, reordenar las variables para que* $a_{11} \neq 0$. *Entonces*

$$q(\boldsymbol{x}) = \frac{1}{a_{11}}\Bigl(\sum_{j=1}^n a_{1j}x_j\Bigr)^2 + q_1(\boldsymbol{x}),$$

donde $q_1(\boldsymbol{x})$ *depende solamente de* $(x_2, \dots, x_n)$.

(b) Si todo $a_{kk} = 0$, *reordenar las variables para que* $a_{12} \neq 0$. *Entonces*

$$q(\boldsymbol{x}) = \frac{1}{2a_{12}}\Bigl(\sum_{j=1}^n (a_{1j} + a_{2j})x_j\Bigr)^2 - \frac{1}{2a_{12}}\Bigl(\sum_{j=1}^n (a_{1j} - a_{2j})x_j\Bigr)^2 + q_2(\boldsymbol{x}),$$

donde $q_2(\boldsymbol{x})$ *depende solamente de* $(x_3, \dots, x_n)$.

(c) Repítase los pasos (a) y (b) con la nueva forma cuadrática q_1 *ó* q_2, *hasta llegar a una forma cuadrática residual nula.*

Demostración. Un cálculo directo verifica que $q_1(\boldsymbol{x})$ no depende de x_1 y que $q_2(\boldsymbol{x})$ no depende de x_1 ni de x_2. La fórmula (4.5) es evidente, ya que el paso (a) construye b_1 y $f_1(\boldsymbol{x})$, o bien en su defecto el paso (b) construye b_1, b_2 y $f_1(\boldsymbol{x})$, $f_2(\boldsymbol{x})$; los demás términos se obtienen al repetir el proceso hasta agotar las variables x_i.

La independencia lineal de las formas lineales $f_1, \dots, f_k$ se deja como ejercicio. □

La forma diagonal de la Proposición 4.16 no es única, porque hay cierta flexibilidad en la elección de los coeficientes b_j. Si, por ejemplo, $b_j = c_j a_j^2$ con $a_j \neq 0$, se puede sustituir $b_j \mapsto c_j$ y $\boldsymbol{x}_j \mapsto a_j^{-1} \boldsymbol{x}_j$, ya que

$$d(a_i^{-1}\boldsymbol{x}_i, a_j^{-1}\boldsymbol{x}_j) = a_i^{-1} a_j^{-1} b_j \, [\![i=j]\!] = c_j \, [\![i=j]\!].$$

Por tanto, la matriz de d en la nueva base es $\operatorname{diag}[c_1, \dots, c_k, 0, \dots, 0]$. Es posible (y deseable), entonces, “normalizar” los coeficientes diagonales no ceros al dividirlos por cuadrados convenientes.

Si $\mathbb{F} = \mathbb{C}$, cualquier número complejo no cero posee una raíz cuadrada compleja.[3] Sin embargo, cuando $\mathbb{F} = \mathbb{R}$, solamente los números positivos (el cero se excluye) poseen una raíz cuadrada real. Estas consideraciones bastan para demostrar el siguiente teorema de Sylvester.[4]

Teorema 4.21 (Ley de Inercia de Sylvester)**.** *Sea V un espacio vectorial 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 $\mathcal{B}$ de V para la cual la matriz de d es

$$[d]_{\mathcal{B}}^{\mathcal{B}} = \operatorname{diag}[1, \dots, 1, 0, \dots, 0] = I_k \oplus O_{n-k},$$

con k entradas diagonales iguales a 1.

(b) Si $\mathbb{F} = \mathbb{R}$, entonces hay una base $\mathcal{B}$ de V para la cual la matriz de d es

$$[d]_{\mathcal{B}}^{\mathcal{B}} = \operatorname{diag}[1, \dots, 1, -1, \dots, -1, 0, \dots, 0] = I_p \oplus -I_q \oplus O_{n-p-q},$$

con p entradas diagonales iguales a 1 *y q entradas diagonales iguales a* (-1)*, donde $p + q = k$.*

Demostración. Sea $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_k, z_1, \dots, z_{n-k}\}$ una base de V para la cual la matriz de d es $\operatorname{diag}[b_1, \dots, b_k, 0, \dots, 0]$, donde cada $b_j \neq 0$.

Ad (a): Si $\mathbb{F} = \mathbb{C}$, sea a_j una de las dos raíces cuadradas 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 $\mathcal{B} := \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_k, z_1, \dots, z_{n-k}\}$ es $I_k \oplus O_{n-k}$.

[3] De hecho, tiene *dos* raíces cuadradas, pues $(-\alpha)^2 = \alpha^2$.

[4] *James Joseph Sylvester* se considera, junto con su compatriota Arthur Cayley, como los padres fundadores del álgebra abstracta, hasta el punto de inventar buena parte de su terminología: el “discriminante” de un polinomio, la “función tociente” de Euler, y la “ley de inercia” para formas cuadráticas, fueron vocablos introducidos por Sylvester.

Ad (b): Si $\mathbb{F} = \mathbb{R}$, sea p el número de los b_j que son positivos y sea q el número de los b_j que son negativos. Es $p + q = k$. Por una permutación de los vectores $\boldsymbol{x}_j$, si fuera necesario, puede suponerse 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 $j = 1, \dots, k$ y defínase $\boldsymbol{y}_j := a_j^{-1} \boldsymbol{x}_j$; entonces la matriz de d respecto de la base $\mathcal{B} := \{\boldsymbol{y}_1, \dots, \boldsymbol{y}_k, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-k}\}$ es $I_p \oplus -I_q \oplus O_{n-p-q}$. □

Definición 4.22. Sea $A = A^t \in M_n(\mathbb{R})$ una matriz simétrica real. Si p es el número de autovalores positivos de A (repetidas según su multiplicidad) y si $q = r(A) - p$ el número de autovalores negativos de A, la diferencia[5] $s(A) := p - q$ es la **signatura** de la matriz A.

Proposición 4.23. *Dos matrices simétricas reales $A, B \in M_n(\mathbb{R})$ son congruentes si y sólo si tienen el mismo rango y la misma signatura.*

Demostración. Si $A \simeq B$, ya se sabe que A y B tienen el mismo rango. Denótese $k := r(A) = r(B)$ en ese caso.

Por el Teorema 4.21, hay enteros $p, p', q, q' \in \mathbb{N}$ con $p + q = p' + q' = k$ tales que $A \simeq I_p \oplus -I_q \oplus O_{n-k}$ mientras $B \simeq I_{p'} \oplus -I_{q'} \oplus O_{n-k}$. Basta verificar, entonces, en el caso $k = n$, que las matrices diagonales $I_p \oplus -I_q$ e $I_{p'} \oplus -I_{q'}$ son congruentes si y sólo si $p = p'$.

Supongamos que $k = n$ y considérese la forma bilineal simétrica sobre $\mathbb{R}^n$ dada por

$$d(\boldsymbol{x}, \boldsymbol{y}) := x_1 y_1 + \cdots + x_p y_p - x_{p+1} y_{p+1} - \cdots - x_{p+q} y_{p+q}. \tag{4.6}$$

Entonces la matriz de d en la base estándar $\mathcal{E} = \{\boldsymbol{e}_1, \dots, \boldsymbol{e}_n\}$ de $\mathbb{R}^n$ es $[d]_{\mathcal{E}}^{\mathcal{E}} = I_p \oplus -I_q$. Sea $\mathcal{U} := \{\boldsymbol{u}_1, \dots, \boldsymbol{u}_n\}$ otra base de $\mathbb{R}^n$ tal que $[d]_{\mathcal{U}}^{\mathcal{U}} = I_{p'} \oplus -I_{q'}$. Explícitamente, con $\boldsymbol{x} = \sum_{j=1}^n x'_j \boldsymbol{u}_j$, $\boldsymbol{y} = \sum_{j=1}^n y'_j \boldsymbol{u}_j$, supóngase que vale

$$d(\boldsymbol{x}, \boldsymbol{y}) = x'_1 y'_1 + \cdots + x'_{p'} y'_{p'} - x'_{p'+1} y'_{p'+1} - \cdots - x'_{p'+q'} y'_{p'+q'}.$$

Considérese los subespacios $M := \mathrm{lin}\langle \boldsymbol{e}_1, \dots, \boldsymbol{e}_p \rangle$ y $N' := \mathrm{lin}\langle \boldsymbol{u}_{p'+1}, \dots, \boldsymbol{u}_{p'+q'} \rangle$. Es claro que

$$d(\boldsymbol{y}, \boldsymbol{y}) > 0 \text{ si } \boldsymbol{y} \in M,\ \boldsymbol{y} \neq \boldsymbol{0}, \quad \text{mientras} \quad d(\boldsymbol{z}, \boldsymbol{z}) < 0 \text{ si } \boldsymbol{z} \in N',\ \boldsymbol{z} \neq \boldsymbol{0}.$$

Por lo tanto, es $M \cap N' = \{0\}$ y luego $M + N' = M \oplus N'$. Al contar dimensiones, se ve que $p + q' = \dim(M \oplus N') \leq n = p + q$, así que $q' \leq q$.

Del mismo modo, los subespacios $M' := \mathrm{lin}\langle \boldsymbol{u}_1, \dots, \boldsymbol{u}_{p'} \rangle$ y $N := \mathrm{lin}\langle \boldsymbol{e}_{p+1}, \dots, \boldsymbol{e}_{p+q} \rangle$ tienen intersección nula, lo cual implica que $p' + q = \dim(M' \oplus N) \leq n = p + q$ y por ende $p' \leq p$. Pero $p' + q' = n = p + q$, de donde $p' = p$ y $q' = q$ necesariamente. □

Corolario 4.24. *La signatura de una matriz de una forma bilineal simétrica real es independiente de la base elegida.* ⊟

Definición 4.25. Si $d \colon V \times V \to \mathbb{R}$ es una forma bilineal simétrica sobre un espacio vectorial, 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 asociada con d.

[5] Algunos autores llaman *signatura* al par ordenado (p, q); a la diferencia $p - q$ la llaman el **índice** de la forma bilineal simétrica $d(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^t A \boldsymbol{y}$.

Definición 4.26. Una forma bilineal simétrica real d es **positiva** si $d(\boldsymbol{x},\boldsymbol{x}) \geq 0$ para todo $\boldsymbol{x} \in V$. Una forma cuadrática real q es positiva si $q(\boldsymbol{x}) \geq 0]$ 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**.

Es evidente que d es positiva si y sólo si su rango y signatura coinciden (porque $q = 0$ si y sólo si $s = k$). Además, d es definida positiva si es positiva y no degenerada (el caso $s = k = n$).

Fíjese que *un producto escalar real* no es más que una forma bilineal simétrica real que es definida positiva. La teoría de espacios vectoriales euclidianos admite una generalización que consiste en reemplazar el producto escalar por usar formas bilineal simétrica *indefinida*. Por ejemplo, se podría reemplazar el producto punto en $\mathbb{R}^n$ por la forma (4.6).

Con respecto de estas formas indefinidas (es decir, cuando $p > 0$ y $q > 0$, o bien cuando $-n < s < n$), es posible definir bases ortonormales, matrices “pseudo-ortogonales”, etcétera, en analogía con el caso euclidiano. Un caso de particular interés es $p = 3$, $q = 1$, la forma bilineal “lorentziana”:

$$d(\boldsymbol{x},\boldsymbol{y}) = -x_0y_0 + x_1y_1 + x_2y_2 + x_3y_3, \quad \text{para} \quad \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^4$$

subyace la teoría “especial” de la relatividad einsteiniana.[6]

4.3 Formas hermíticas

Las formas bilineales simétricas complejas tienen menos estructura que las formas reales análogas, ya que su clasificación depende solamente de su rango y no de su signatura. Esto se debe a que la simetría de una forma bilineal compleja no es muy apropiada, porque ignora la conjugación compleja de las escalares en $\mathbb{C}$. Para incorporar la conjugación compleja, es oportuno sustituir la noción de forma bilineal por la de forma sesquilineal; dichas formas se clasificarán, como luego se verá, por su rango y signatura.

Definición 4.27. Sea V un espacio vectorial finitodimensional sobre el cuerpo $\mathbb{C}$. 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,

$$\begin{aligned} h(\boldsymbol{z}+\boldsymbol{z}',\boldsymbol{w}) &= h(\boldsymbol{z},\boldsymbol{w}) + h(\boldsymbol{z}',\boldsymbol{w}), & h(\boldsymbol{z},\alpha\boldsymbol{w}) &= \alpha\, h(\boldsymbol{z},\boldsymbol{w}), \\ h(\boldsymbol{z},\boldsymbol{w}+\boldsymbol{w}') &= h(\boldsymbol{z},\boldsymbol{w}) + h(\boldsymbol{z},\boldsymbol{w}'), & h(\alpha\boldsymbol{z},\boldsymbol{w}) &= \bar{\alpha}\, h(\boldsymbol{z},\boldsymbol{w}), \end{aligned}$$

para todo $\boldsymbol{z},\boldsymbol{z}',\boldsymbol{w},\boldsymbol{w}' \in V$, $\alpha \in \mathbb{C}$.

Una forma sesquilineal h es **hermítica** si $h(\boldsymbol{z},\boldsymbol{w}) = \overline{h(\boldsymbol{w},\boldsymbol{z})}$ para $\boldsymbol{z},\boldsymbol{w} \in V$.

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 sólo si su matriz cumple $A^* = A$.

[6] El principio de la relatividad fue formulado por Galileo Galilei, en: *Dialogo sopra i due massimi sistemi del mondo*, Firenze, 1632. Las ecuaciones de Maxwell para movimiento en campos electromagnéticos incumplen este principio, pero Einstein logró recuperar la relatividad al precio de postular que la velocidad de la luz es constante.

Si $\mathcal{B}' = \{z'_1, \dots, z'_n\}$ es otra base de V, y si P es la matriz de cambio de base (1.9), entonces las matrices respectivas $A = [h]^{\mathcal{B}}_{\mathcal{B}}$, $B = [h]^{\mathcal{B}'}_{\mathcal{B}'}$ de la forma sesquilineal h para estas bases cumplen

$$B = P^* AP.$$

Para demostrar esta fórmula, sólo hay que modificar la demostración de la Proposición 4.4, lo cual se deja como ejercicio.

La relación de equivalencia $A \rightleftarrows P^*AP$, con P inversible, a veces se llama "congruencia hermítica". Dos matrices hermíticamente congruentes tienen el mismo rango: este es el caso de $A \mapsto QAP$ con $Q = P^*$. 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*.

Si $h\colon V \times V \to \mathbb{C}$ es una forma hermítica y si M es un subespacio de V, su *subespacio ortogonal* con respecto a h es

$$\begin{aligned} M^\perp &= \{\, z \in V : h(z, w) = 0 \text{ para todo } w \in M \,\} \\ &= \{\, w \in V : h(z, w) = 0 \text{ para todo } z \in M \,\}. \end{aligned}$$

Ejemplo 4.28. Si V es un espacio hilbertiano (complejo), su producto escalar $h(z, w) := \langle z, w \rangle$ es una forma sesquilineal y hermítica, que además es definida positiva, es decir, es $h(z, z) > 0$ para $z \neq \mathbf{0}$ en V. Por tanto, una forma hermítica es una generalización de la noción de producto escalar, donde se omite el requisito de positividad.

Ejemplo 4.29. Si $p, q \in \mathbb{N}$ con $p + q = k \leq n$, la forma sesquilineal siguiente sobre $\mathbb{C}^n$ es hermítica:

$$h(z, w) := \bar{z}_1 w_1 + \cdots + \bar{z}_p w_p - \bar{z}_{p+1} w_{p+1} - \cdots - \bar{z}_{p+q} w_{p+q}.$$

Su matriz (respecto de la base estándar $\mathcal{E}$) es

$$[h]^{\mathcal{E}}_{\mathcal{E}} = \operatorname{diag}[1, \dots, 1, -1, \dots, -1, 0, \dots, 0] = I_p \oplus -I_q \oplus O_{n-p-q}.$$

El ejemplo anterior es típico: el siguiente teorema es una versión de la ley de inercia de Sylvester para formas hermíticas. Dichas formas se clasifican por su rango $p + q$ y su signatura $s := p - q$.

Proposición 4.30. *Sea V un espacio vectorial sobre $\mathbb{C}$ con $\dim V = n$. Sea $h\colon V \times V \to \mathbb{C}$ una forma hermítica, de rango k. Entonces hay una base $\mathcal{B}$ de V tal que $[h]^{\mathcal{B}}_{\mathcal{B}} = I_p \oplus -I_q \oplus O_{n-k}$, donde $p + q = k$. Además, si hay otra base $\mathcal{B}'$ de V tal que $[h]^{\mathcal{B}'}_{\mathcal{B}'} = I_{p'} \oplus -I_{q'} \oplus O_{n-k}$, entonces $p' = p$ y $q' = q$.*

Demostración. Las demostraciones de las Proposiciones 4.16 y 4.23 y del Teorema 4.21(b) se adaptan directamente al caso hermítico: se deja los detalles como un ejercicio.

(Fíjese que los b_j obtenidos de la Proposición 4.16 son reales, porque $h(z, z) \in \mathbb{R}$ para todo $z \in V$; y que además $h(\alpha^{-1} z, \alpha^{-1} z) = |\alpha|^{-2} h(z, z)$ para todo $\alpha \in \mathbb{C}$.) $\square$

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:

$$J_2 := \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

En adelante se verá que cualquier matriz antisimétrica es congruente con una suma directa de varias copias de esta J_2.

Una forma bilineal alternada *no degenerada* $s \colon V \times V \to \mathbb{F}$ se llama **forma simpléctica** sobre V.

Definición 4.31. 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 (4.3):

$$\begin{aligned} M^\perp &= \{\, \boldsymbol{x} \in V : s(\boldsymbol{x},\boldsymbol{y}) = 0 \text{ para todo } \boldsymbol{y} \in M \,\} \\ &= \{\, \boldsymbol{y} \in V : s(\boldsymbol{x},\boldsymbol{y}) = 0 \text{ para todo } \boldsymbol{x} \in M \,\}. \end{aligned}$$

Los complementos simplécticos presentan un fuerte contraste con los complementos ortogonales determinados por formas simétricas o hermíticas. Respecto de s, cualquier vector $\boldsymbol{x} \in V$ cumple $s(\boldsymbol{x},\boldsymbol{x}) = 0$ por antisimetría. En consecuencia, si $N := \operatorname{lin}\langle \boldsymbol{x} \rangle = \{\, c\boldsymbol{x} : c \in \mathbb{F} \,\}$ es el subespacio unidimensional generado por $\boldsymbol{x}$, entonces $N \subseteq N^\perp$.

Un subespacio $M \leq V$ tal que $\underline{M \subseteq M^\perp}$ se llama un **subespacio isotrópico** de V.

Proposición 4.32. *Sea $s \colon V \times V \to \mathbb{F}$ una forma bilineal alternada, de rango k. Entonces k es un número entero par; al escribir $k =: 2m$, hay una base $\mathcal{B} = \{\boldsymbol{x}_1,\boldsymbol{y}_1,\dots,\boldsymbol{x}_m,\boldsymbol{y}_m,\boldsymbol{z}_1,\dots,\boldsymbol{z}_{n-2m}\}$ para la cual la matriz de $\underline{s}$ es de la forma*

$$[s]^{\mathcal{B}}_{\mathcal{B}} = \begin{bmatrix} J_{2m} & O \\ O & O_{n-2m} \end{bmatrix}, \quad \textit{con} \quad J_{2m} := \begin{bmatrix} J_2 & O & \dots & O \\ O & J_2 & \dots & O \\ \vdots & \vdots & \ddots & \vdots \\ O & O & \dots & J_2 \end{bmatrix} = \underbrace{J_2 \oplus \dots \oplus J_2}_{m \text{ veces}}. \tag{4.7}$$

Demostración. Obsérvese primero que la matriz de s es del tipo indicado si y sólo si

$$s(\boldsymbol{x}_j,\boldsymbol{y}_j) = -s(\boldsymbol{y}_j,\boldsymbol{x}_j) = 1$$

y además $s(\boldsymbol{z},\boldsymbol{w}) = 0$ para cualquier otro par de vectores básicos $\boldsymbol{z},\boldsymbol{w} \in \mathcal{B}$. Si $s \equiv 0$, se puede usar una base arbitaria $\mathcal{B} = \{\boldsymbol{z}_1,\dots,\boldsymbol{z}_n\}$ de V, porque $[s]^{\mathcal{B}}_{\mathcal{B}} = O$. En cambio, si s no es idénticamente nula, hay dos vectores $\boldsymbol{x}_1,\boldsymbol{y}'_1 \in V$ con $s(\boldsymbol{x}_1,\boldsymbol{y}'_1) =: c_1 \neq 0$. Ahora $\boldsymbol{x}_1,\boldsymbol{y}'_1$ son linealmente independientes porque $s(\boldsymbol{x}_1, \boldsymbol{y}'_1)$ sería cero por antisimetría si $\boldsymbol{x}_1$, $\boldsymbol{y}'_1$ fueran proporcionales. Colóquese $\boldsymbol{y}_1 := c_1^{-1}\boldsymbol{y}'_1$, de modo que $s(\boldsymbol{x}_1, \boldsymbol{y}_1) = 1$ y $s(\boldsymbol{y}_1, \boldsymbol{x}_1) = -1$.

Supóngase, para argumentar por inducción, que ya se haya elegido $2r$ vectores linealmente independientes $\{\boldsymbol{x}_1, \boldsymbol{y}_1, \dots, \boldsymbol{x}_r, \boldsymbol{y}_r\}$ en V con

$$s(\boldsymbol{x}_i, \boldsymbol{y}_j) = -s(\boldsymbol{y}_j, \boldsymbol{x}_i) = [\![i=j]\!], \qquad s(\boldsymbol{x}_i, \boldsymbol{x}_j) = s(\boldsymbol{y}_i, \boldsymbol{y}_j) = 0,$$

para $i, j = 1, \dots, r$. 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, y además $M_{2r} \cap M_{2r}^{\perp} = \{\boldsymbol{0}\}$.

Si s es no degenerada (este el caso $k = n$), entonces $V = M_{2r} \oplus M_{2r}^{\perp}$ por la Proposición 4.11. (Fíjese que la demostración de esta Proposición sigue válida sin cambio alguno para formas bilineales alternadas en vez de simétricas). Resulta que esta relación es válida aun para $k < n$. Para $\boldsymbol{x} \in V$, 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 $j = 1, \dots, r$, vale

$$\begin{aligned} s(\boldsymbol{z}, \boldsymbol{x}_j) &= s(\boldsymbol{x}, \boldsymbol{x}_j) + s(\boldsymbol{x}, \boldsymbol{x}_j)s(\boldsymbol{y}_j, \boldsymbol{x}_j) = 0, \\ s(\boldsymbol{z}, \boldsymbol{y}_j) &= s(\boldsymbol{x}, \boldsymbol{y}_j) - s(\boldsymbol{x}, \boldsymbol{y}_j)s(\boldsymbol{x}_j, \boldsymbol{y}_j) = 0, \end{aligned}$$

así que $\boldsymbol{z} \in M_{2r}^{\perp}$. Por tanto,

$$\boldsymbol{x} = \sum_{i=1}^{r} s(\boldsymbol{x}, \boldsymbol{y}_i)\boldsymbol{x}_i - \sum_{i=1}^{r} s(\boldsymbol{x}, \boldsymbol{x}_i)\boldsymbol{y}_i + \boldsymbol{z} \in M_{2r} \oplus M_{2r}^{\perp}.$$

Como $\boldsymbol{x}$ es arbitrario, se concluye que $V = M_{2r} \oplus M_{2r}^{\perp}$.

La matriz de s con respecto a esta descomposición de V es evidentemente de la forma $\begin{bmatrix} J_{2r} & O \\ O & C_{n-2r} \end{bmatrix}$, donde $C_{n-2r} \in M_{n-2r}(\mathbb{F})$.

Si $C_{n-2r} = O$, entonces $2r = k$ y se puede elegir una base $\{\boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-2r}\}$ para $M_{2r}^{\perp}$ cuya unión con la base $\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_r, \boldsymbol{y}_1, \dots, \boldsymbol{y}_r\}$ de M_{2r} es la base $\mathcal{B}$ deseada.

Si $C_{n-2r} \neq 0$ y por ende $2r < k$, la restricción de s al subespacio $M_{2r}^{\perp}$ no es nula y existen dos vectores no proporcionales $\boldsymbol{x}_{r+1}, \boldsymbol{y}_{r+1} \in M_{2r}^{\perp}$ tales que $s(\boldsymbol{x}_{r+1}, \boldsymbol{y}_{r+1}) = -s(\boldsymbol{y}_{r+1}, \boldsymbol{x}_{r+1}) = 1$. Ahora $\{\boldsymbol{x}_1, \boldsymbol{y}_1, \dots, \boldsymbol{x}_r, \boldsymbol{y}_r, \boldsymbol{x}_{r+1}, \boldsymbol{y}_{r+1}\}$ es la base de un subespacio $M_{2r+2} \leq V$ y la restricción de s a este subespacio tiene la matriz J_{2r+2}.

Este proceso se repite hasta llegar a J_{2m}, donde $2m + 2 > k$. Si fuera $k = 2m + 1$, sería imposible elegir dos vectores no proporcionales en $M_{2m}^{\perp}$ en los cuales s no se anula. Luego es $k = 2m$ y $C_{n-2m} = O$. Al elegir una base cualquiera $\{\boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-2m}\}$ para $M_{2m}^{\perp}$, se obtiene la base deseada de V. □

A veces conviene permutar los vectores de la base $\mathcal{B}$ obtenida en la Proposición anterior para cambiarla a $\mathcal{B}' = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m, \boldsymbol{y}_1, \dots, \boldsymbol{y}_m, \boldsymbol{z}_1, \dots, \boldsymbol{z}_{n-2m}\}$, para la cual la matriz de s tiene el formato:

$$[s]_{\mathcal{B}'}^{\mathcal{B}'} = \begin{bmatrix} O & I_r & O \\ -I_r & O & O \\ O & O & O_{n-2m} \end{bmatrix}.$$

Corolario 4.33. *El rango de una matriz antisimétrica $A = -A^t \in M_n(\mathbb{F})$ es* par. *Dos matrices antisimétricas reales $A, B \in M_n(\mathbb{F})$ son congruentes si y sólo si tienen el mismo rango.*

Demostración. Defínase la forma bilineal s sobre $\mathbb{F}^n$ por $s_A(\boldsymbol{x},\boldsymbol{y}) := \boldsymbol{x}^t A\boldsymbol{y}$. Si $A = -A^t$, entonces s es una forma alternada. La Proposición anterior dice que hay una matriz de cambio de base $P = [I]^{\mathcal{B}}_{\mathcal{E}}$ tal que $P^t AP = J_{2m} \oplus O_{n-2m}$, para algún $m \in \mathbb{N}$ con $2m \leq n$. El rango de A es $r(A) = r(P^t AP) = 2m$.

Si $A = -A^t$, $B = -B^t$ y $r(A) = r(B) = 2m$, entonces la Proposición anterior, aplicada a las formas alternadas s_A y s_B, muestra que $A \doteq J_{2m} \oplus O_{n-2m} \doteq B$. $\square$

Corolario 4.34. *Una forma simpléctica sobre V existe sólo si* $\dim V$ *es par: $n = 2m$.* $\boxminus$

Sea $A \in M_n(\mathbb{F})$ una matriz antisimétrica *inversible*. Entonces $s_A(\boldsymbol{x},\boldsymbol{y}) := \boldsymbol{x}^t A\boldsymbol{y}$ es una forma simpléctica sobre $\mathbb{F}^n$, y $n = 2m$ es necesariamente par. La Proposición 4.32 dice que $A \doteq J_{2r}$: hay una matriz inversible R tal que $A = R^t J_{2r} R$. Obsérvese que $\det J_{2r} = (\det J_2)^r = 1$. En consecuencia, vale

$$\det A = (\det R)^2.$$

En particular, si $\mathbb{F} = \mathbb{R}$, esto implica que $\det A > 0$.

Es legítimo escribir $\det R = \pm\sqrt{\det A}$, porque $\det R$ es una raíz cuadrada del determinante de A. Lo que es menos evidente, pero cierto, es que $\det R$ 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 entero.) Además, esta "raíz cuadrada del determinante" resulta de la evaluación en las entradas de A de un polinomio "universal", muy análogo al polinomio (1.15) que define el determinante de A por la fórmula de Leibniz.

Es necesario, para entender su definición, hacer un inciso de la teoría de polinomios. Para $i, j = 1,\dots,n$ con $i < j$, tómese una "incógnita" t_{ij}, y sea $\boldsymbol{t} := (t_{12}, t_{13}, \dots, t_{n-1,n})$. Sea $\mathbb{Q}(\boldsymbol{t})$ el cuerpo de cocientes de polinomios en estas incógnitas, con coeficientes racionales. (Al multiplicar el numerador y denominador de un tal cociente por un número entero apropiado, se puede suponer que ese numerador y denominador tienen coeficientes enteros, es decir, pertenecen al anillo de polinomios $\mathbb{Z}[\boldsymbol{t}]$.) Ahora considérese la matriz antisimétrica T definido por

$$T := \begin{bmatrix} 0 & t_{12} & t_{13} & \dots & t_{1n} \\ -t_{12} & 0 & t_{23} & \dots & t_{2n} \\ -t_{13} & -t_{23} & 0 & \dots & t_{3n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -t_{1n} & -t_{2n} & -t_{3n} & \dots & 0 \end{bmatrix} \in M_n(\mathbb{Q}(\boldsymbol{t})).$$

En el cuerpo $\mathbb{Q}(\boldsymbol{t})$, vale $1 + 1 \neq 0$, así que las proposiciones ya vistas sobre formas bilineales alternadas siguen válidas para $\mathbb{F} = \mathbb{Q}(\boldsymbol{t})$. Se concluye que $\underline{\det T = (\det R)^2}$ para cierta matriz R con entradas en $\mathbb{Q}(\boldsymbol{t})$.

De la fórmula (1.15) se sabe que $\det R$ es un polinomio en las entradas de R, así que $\det R = q(\boldsymbol{t})/r(\boldsymbol{t})$ donde q, r son polinomios en $\mathbb{Z}[\boldsymbol{t}]$ sin factor común. También por (1.15), resulta que $\det T =: s(\boldsymbol{t})$ es otro polinomio en las incógnitas $\boldsymbol{t}$. La relación $\det T = (\det R)^2$ implica

$$s(\boldsymbol{t})r(\boldsymbol{t})^2 = q(\boldsymbol{t})^2. \tag{4.8}$$

Los polinomios con coeficientes enteros (en varias variables) admiten *factorización única*: al expresar los dos lados de (4.8) como producto de polinomios irreducibles, se ve que 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, se concluye que $r(\boldsymbol{t}) \equiv \pm 1$. Por lo tanto, $\underline{\det T = q(\boldsymbol{t})^2}$ *es un cuadrado perfecto* en $\mathbb{Z}[\boldsymbol{t}]$.

Definición 4.35. Sea $A = -A^t$ una matriz antisimétrica en $M_n(\mathbb{F})$ donde $n = 2m$ es par. La evaluación de polinomios $t_{ij} \mapsto a_{ij}$ lleva $\mathbb{Q}(\boldsymbol{t})$ en $\mathbb{F}$ y lleva la matriz T en A. El polinomio en las entradas a_{ij} que es la imagen de $\det R$ es el **pfaffiano** de A:

$$\operatorname{Pf} A := q(a_{12}, a_{13}, \dots, a_{n-1,n}).$$

Para resolver la ambigüedad de signo en $q(\boldsymbol{t})$, se requiere adicionalmente que $\underline{\operatorname{Pf}(J_{2r}) = +1}$.

La evaluación de polinomios preserva productos; en consecuencia, vale

$$\det A = (\operatorname{Pf} A)^2. \tag{4.9}$$

Ejemplo 4.36. En el caso $n = 2$, es $A = \begin{bmatrix} 0 & a_{12} \\ -a_{12} & 0 \end{bmatrix}$, así que $\det A = a_{12}^2$ así que $\operatorname{Pf} A = \pm a_{12}$. El signo queda determinado por la condición $\operatorname{Pf} J_2 = +1$. Luego, vale $\underline{\operatorname{Pf} A = a_{12}}$.

Ejemplo 4.37. En el caso $n = 4$, el pfaffiano es

$$A = \begin{bmatrix} 0 & a_{12} & a_{13} & a_{14} \\ -a_{12} & 0 & a_{23} & a_{24} \\ -a_{13} & -a_{23} & 0 & a_{34} \\ -a_{14} & -a_{24} & -a_{34} & 0 \end{bmatrix} \quad \Longrightarrow \quad \operatorname{Pf} A = a_{12}a_{34} - a_{13}a_{24} + a_{14}a_{23}.$$

Si $A = -A^t$ es una matriz antisimétrica en $M_n(\mathbb{F})$ donde $n = 2m+1$ es *impar*, entonces la Proposición 4.32 muestra que $A \simeq J_{2r} \oplus O_{2m-2r+1}$ para algún $r \leq m$, y en particular que A *no es inversible*, pues $\det A = 0$. En este caso conviene definir $\operatorname{Pf} A := 0$ también.

Ejemplo 4.38. La fórmula general para el pfaffiano de una matriz antisimétrica $A \in M_{2m}(\mathbb{F})$ es la siguiente:

$$\operatorname{Pf} A := \frac{1}{2^m m!} \sum_{\sigma} (-1)^{\sigma} a_{\sigma(1)\sigma(2)} a_{\sigma(3)\sigma(4)} \dots a_{\sigma(2m-1)\sigma(2m)}. \tag{4.10}$$

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 esta fórmula aparecerá en los Ejercicios.

Proposición 4.39. *Si $A = -A^t$ es una matriz antisimétrica en $M_n(\mathbb{F})$ y si $S \in M_n(\mathbb{F})$ es una matriz cualquiera, entonces*

$$\operatorname{Pf}(S^t A S) = (\det S)(\operatorname{Pf} A). \tag{4.11}$$

Demostración. Si n es impar, los dos lados de la ecuación valen 0. Supóngase, entonces, que n es par.

Obsérvese que $(S^tAS)^t = S^tA^tS = -S^tAS$, así que la matriz S^tAS es también antisimétrica y posee un pfaffiano. Ahora $\det(S^tAS) = (\det S)^2 \det A$ por las propiedades conocidas de determinantes. De (4.9) se obtiene enseguida:

$$\operatorname{Pf}(S^tAS) = \pm(\det S)(\operatorname{Pf} A).$$

Los dos lados de esta igualdad resultan de la evaluación de una identidad polinomial, con T en lugar de A y una matriz análoga (cuyas entradas son nuevas incógnitas s_{ij}) en lugar de S. Esto significa que el signo al lado derecho es el mismo, cualesquiera que sean A y S. Al tomar $S = I_n$, y al recordar que $\det I_n = 1$, se ve que este signo es positivo, y la fórmula deseada queda comprobada. □

► Hay un contexto importante en donde coexisten una forma bilineal simétrica y una forma bilineal alternada que juegan papeles complementarias. Ese es el caso de un un espacio vectorial *complejo* W de dimensión m sobre $\mathbb{C}$, dotado de un producto escalar $\langle\cdot,\cdot\rangle$. Se puede considerar W como un espacio vectorial *real* de dimensión $2m$ sobre $\mathbb{R}$, al tomar la multiplicación escalar $\boldsymbol{x} \mapsto c\boldsymbol{x}$ sólo para $c \in \mathbb{R}$, "olvidando" o despreciando las aplicaciones $\boldsymbol{x} \mapsto \pm i\boldsymbol{x}$ para $i = \sqrt{-1}$. Las *partes real e imaginaria* del producto escalar,

$$d(\boldsymbol{x},\boldsymbol{y}) := \Re\langle\boldsymbol{x},\boldsymbol{y}\rangle, \qquad s(\boldsymbol{x},\boldsymbol{y}) := \Im\langle\boldsymbol{x},\boldsymbol{y}\rangle,$$

definen dos formas $\mathbb{R}$-bilineales, d y s, sobre W. Es evidente que d es simétrica y que s es alternada. La positividad del producto escalar implica que d y s son formas no degeneradas, y que d sea *definida positiva*, es decir, su rango y signatura son máximos: $r(d) = s(d) = 2m$.

Considérese el problema inverso, el de transformar un espacio vectorial real V de dimensión par $n = 2m$, con un producto escalar real $d\colon V \times V \to \mathbb{R}$, en un espacio vectorial complejo de dimensión m, con un producto escalar complejo $\langle\cdot,\cdot\rangle$. Lo que hace falta es una manera de prescribir la multiplicación escalar de $\pm i$ sobre V.

Definición 4.40. Sea (V,d) un espacio euclidiano (esto es, un espacio vectorial real V con una forma bilineal simétrica d que es definida positiva) de dimensión par $n = 2m$. Una **estructura compleja ortogonal** sobre (V,d) es un operador $J \in \operatorname{End}_{\mathbb{R}}(V)$ tal que

(a) $\underline{J^2 = -I}$ en $\operatorname{End}_{\mathbb{R}}(V)$, y además

(b) $d(J\boldsymbol{x}, J\boldsymbol{y}) = d(\boldsymbol{x},\boldsymbol{y})$ para todo $\boldsymbol{x},\boldsymbol{y} \in V$.

Por el teorema de inercia de Sylvester, hay una base $\mathcal{E} = \{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$ de V tal que

$$d(\boldsymbol{x},\boldsymbol{y}) = x_1y_1 + \cdots + x_ny_n, \quad \text{cuando} \quad \boldsymbol{x} = \sum_{i=1}^n x_i\boldsymbol{e}_i, \quad \boldsymbol{y} = \sum_{j=1}^n y_j\boldsymbol{e}_j.$$

En otras palabras, la base $\mathcal{E}$ es una base ortonormal para el espacio euclidiano (V,d). Al identificar el vector $\boldsymbol{x} \in V$ con $(x_1,\dots,x_n) \in \mathbb{R}^n$, se obtiene $d(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^t\boldsymbol{y}$.

Con un leve abuso de notación, se puede usar la misma letra J para denotar la matriz $[J]^{\mathcal{E}}_{\mathcal{E}} \in M_n(\mathbb{R})$. Esta matriz J cumple dos propiedades:

(a) $\underline{J^2 = -I_n}$ en $M_n(\mathbb{R})$, y además

(b) $\underline{J^t J = I_n}$.

La propiedad (b) es consecuencia de la relación

$$\boldsymbol{x}^t J^t J\boldsymbol{y} = (J\boldsymbol{x})^t J\boldsymbol{y} = d(J\boldsymbol{x}, J\boldsymbol{y}) = d(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^t\boldsymbol{y} \quad \text{para todo} \quad \boldsymbol{x},\boldsymbol{y} \in \mathbb{R}^n.$$

En otras palabras, J es una matriz ortogonal. Además, las propiedades (a) y (b) implican que $\underline{J^t = -J} = J^{-1}$.

Ejemplo 4.41. Si $V = \mathbb{R}^{2m}$, se puede tomar $J := J_{2m}$ como en (4.7), la suma directa de m copias de la matriz $J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$. Es claro que $J_{2m}^2 = -I_2 \oplus \cdots \oplus -I_2 = -I_{2m}$.

Lema 4.42. *Sea (V,d) un espacio euclidiano de dimensión $2m$ sobre $\mathbb{R}$ y sea $J \in \operatorname{End}_{\mathbb{R}}(V)$ una estructura compleja ortogonal. Entonces $\underline{s(\boldsymbol{x},\boldsymbol{y}) := d(J\boldsymbol{x},\boldsymbol{y})}$ define una forma bilineal simpléctica sobre V tal que $s(J\boldsymbol{x}, J\boldsymbol{y}) \equiv s(\boldsymbol{x},\boldsymbol{y})$.*

Demostración. Es evidente que s es una forma bilineal sobre V. Si $\boldsymbol{x},\boldsymbol{y} \in V$, entonces

$$s(\boldsymbol{y},\boldsymbol{x}) = d(J\boldsymbol{y},\boldsymbol{x}) = d(J^2\boldsymbol{y}, J\boldsymbol{x}) = -d(\boldsymbol{y}, J\boldsymbol{x}) = -d(J\boldsymbol{x},\boldsymbol{y}) = -s(\boldsymbol{x},\boldsymbol{y}), \tag{4.12}$$

por las propiedades (a) y (b) de la Definición 4.40 y la simetría de d. Por 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 $\boldsymbol{y}$; de ahí, es $d(J\boldsymbol{x}, J\boldsymbol{x}) = 0$. Luego $J\boldsymbol{x} = \boldsymbol{0}$ porque d es definida positiva, y en consecuencia $\boldsymbol{x} = \boldsymbol{0}$ ya que J es inversible (con $J^{-1} = -J$). Esto comprueba que s es no degenerada.

También, es $s(J\boldsymbol{x}, J\boldsymbol{y}) = d(J^2\boldsymbol{x}, J\boldsymbol{y}) = s(\boldsymbol{x},\boldsymbol{y})$, a partir de la ecuación (4.12) con $\boldsymbol{x} \leftrightarrow \boldsymbol{y}$. $\square$

Proposición 4.43. *Sea (V,d) un espacio euclidiano de dimensión $2m$ sobre $\mathbb{R}$ y sea $J \in \operatorname{End}_{\mathbb{R}}(V)$ una estructura compleja ortogonal. Defínase una multiplicación escalar compleja sobre V por*

$$(a+ib)\boldsymbol{x} := a\boldsymbol{x} + bJ(\boldsymbol{x}), \quad \textit{para todo} \quad a,b \in \mathbb{R}. \tag{4.13a}$$

Denótese por V_J el espacio vectorial complejo formado por el conjunto V con su propia operación de suma y esta nueva multiplicación escalar. Entonces V_J es un espacio hilbertiano de dimensión m sobre $\mathbb{C}$, con el producto escalar

$$\langle \boldsymbol{x},\boldsymbol{y}\rangle_J := d(\boldsymbol{x},\boldsymbol{y}) + i\,s(\boldsymbol{x},\boldsymbol{y}) = d(\boldsymbol{x},\boldsymbol{y}) + i\,d(J\boldsymbol{x},\boldsymbol{y}). \tag{4.13b}$$

Demostración. Es fácil comprobar que la operación (4.13a) hace de V un espacio vectorial complejo: sólo hay que observar que para $a,b,p,q \in \mathbb{R}$, $\boldsymbol{x} \in V$, vale

$$\begin{aligned}(a+ib)(p+iq)\boldsymbol{x} &= (a+ib)(p\boldsymbol{x} + qJ(\boldsymbol{x})) = ap\boldsymbol{x} + (aq+bp)\,J(\boldsymbol{x}) + bqJ^2(\boldsymbol{x}) \\ &= (ap-bq)\,\boldsymbol{x} + (aq+bp)\,J(\boldsymbol{x}) = ((ap-bq) + i(aq+bp))\,\boldsymbol{x}.\end{aligned}$$

Por el Lema anterior, $s(\boldsymbol{x},\boldsymbol{y}) := d(J\boldsymbol{x},\boldsymbol{y})$ es una forma bilineal alternada, así que $d+is$ es una forma hermítica sobre V. Como $s(\boldsymbol{x},\boldsymbol{x}) = 0$ para todo $\boldsymbol{x}$, se ve que $\langle \boldsymbol{x},\boldsymbol{x}\rangle_J = d(\boldsymbol{x},\boldsymbol{x}) \geq 0$ para todo $\boldsymbol{x} \in V$, con igualdad sólo si $\boldsymbol{x} = \boldsymbol{0}$. Luego V_J es un espacio hilbertiano.

Fíjese que

$$s(\boldsymbol{x},\boldsymbol{y}) = d(J\boldsymbol{x},\boldsymbol{y}) \quad \text{es equivalente a} \quad d(\boldsymbol{x},\boldsymbol{y}) = s(\boldsymbol{x},J\boldsymbol{y}),$$

porque $s(\boldsymbol{x},J\boldsymbol{y}) = d(J\boldsymbol{x},J\boldsymbol{y}) = d(\boldsymbol{x},\boldsymbol{y})$.

Sea $\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_r\}$ una familia ortonormal en el espacio hilbertiano V_J y considérese el conjunto de vectores $\mathcal{B}_r = \{\boldsymbol{u}_1, J(\boldsymbol{u}_1),\dots,\boldsymbol{u}_r, J(\boldsymbol{u}_r)\}$ en V. Si $i \neq j$, entonces

$$d(\boldsymbol{u}_i,\boldsymbol{u}_j) + i\,d(J\boldsymbol{u}_i,\boldsymbol{u}_j) = \langle \boldsymbol{u}_i,\boldsymbol{u}_j\rangle = 0.$$

Además, $d(J\boldsymbol{u}_i,\boldsymbol{u}_i) = s(\boldsymbol{u}_i,\boldsymbol{u}_i) = 0$ para $i = 1,\dots,r$. Entonces $\mathcal{B}_r$ es una base ortonormal de un subespacio M_r de (V,d), de dimensión real $2r$. Si $r < m$, hay un vector $\boldsymbol{u}_{r+1} \in V$ 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 en (V,d). También, es $d(J\boldsymbol{u}_{r+1},J\boldsymbol{u}_{r+1}) = d(\boldsymbol{u}_{r+1},\boldsymbol{u}_{r+1}) = 1$ y $d(J\boldsymbol{u}_{r+1},\boldsymbol{u}_i) = d(J\boldsymbol{u}_{r+1},J\boldsymbol{u}_i) = 0$, así que $\{\boldsymbol{u}_1,\dots,\boldsymbol{u}_{r+1}\}$ es una familia ortonormal en V_J. Al llegar a $r = m$, se ha construido una base ortonormal $\mathcal{U} = \{\boldsymbol{u}_1,\dots,\boldsymbol{u}_m\}$ del espacio hilbertiano V_J y al mismo tiempo una base ortonormal $\mathcal{B}_r = \{\boldsymbol{u}_1, J(\boldsymbol{u}_1),\dots,\boldsymbol{u}_m, J(\boldsymbol{u}_m)\}$ del espacio euclidiano (V,d). En particular, es $\dim_{\mathbb{C}} V_J = m$. □

Definición 4.44. Sea (V,d) un espacio euclidiano con $\dim_{\mathbb{R}} V = n$. La **complexificación** de V es el espacio vectorial complejo

$$V_{\mathbb{C}} = V \oplus iV := \{\,\boldsymbol{x} + i\boldsymbol{y} : \boldsymbol{x},\boldsymbol{y} \in V\,\},$$

con multiplicación escalar $(a+ib)(\boldsymbol{x}+i\boldsymbol{y}) := (a\boldsymbol{x} - b\boldsymbol{y}) + i(b\boldsymbol{x} + a\boldsymbol{y})$ para $a,b \in \mathbb{R}$, $\boldsymbol{x},\boldsymbol{y} \in V$. Si $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y} \in V_{\mathbb{C}}$, escríbase $\bar{\boldsymbol{z}} := \boldsymbol{x} - i\boldsymbol{y}$. La forma $\mathbb{R}$-bilineal d sobre V se puede ampliar a una forma $\mathbb{C}$-bilineal sobre $V_{\mathbb{C}}$ al poner $d(\boldsymbol{x}+i\boldsymbol{y},\boldsymbol{x}'+i\boldsymbol{y}') := d(\boldsymbol{x},\boldsymbol{x}') + i\,d(\boldsymbol{x},\boldsymbol{y}') + i\,d(\boldsymbol{y},\boldsymbol{x}') - d(\boldsymbol{y},\boldsymbol{y}')$. Es posible dotar $V_{\mathbb{C}}$ de un producto escalar complejo al definir

$$\langle\!\langle \boldsymbol{z},\boldsymbol{w}\rangle\!\rangle := 2d(\bar{\boldsymbol{z}},\boldsymbol{w}) \quad \text{para} \quad \boldsymbol{z},\boldsymbol{w} \in V_{\mathbb{C}}. \tag{4.14}$$

Fíjese que $V_{\mathbb{C}}$ tiene dimensión n sobre $\mathbb{C}$.

Lema 4.45. *Sea (V,d) un espacio euclidiano con una estructura compleja ortogonal J. Entonces $W := \{\,\boldsymbol{x} - iJ(\boldsymbol{x}) : \boldsymbol{x} \in V\,\}$ es un subespacio complejo de $V_{\mathbb{C}}$, isomorfo a V_J como espacio hilbertiano.*

Demostración. Defínase $P_J := \frac{1}{2}(I - iJ) \in \operatorname{End}_{\mathbb{C}}(V_{\mathbb{C}})$. Entonces $P_J(V_{\mathbb{C}}) = P_J(V) = W$, mientras $P_J^2 = P_J$ y además $P_J^* = \frac{1}{2}(I + iJ^t) = \frac{1}{2}(I - iJ) = P_J$. Luego P_J es el proyector ortogonal sobre $V_{\mathbb{C}}$ con imagen W. Si $\boldsymbol{x},\boldsymbol{y} \in V$, entonces

$$\begin{aligned}
\langle\!\langle P_J(\boldsymbol{x}), P_J(\boldsymbol{y})\rangle\!\rangle &= \tfrac{1}{2}d(\boldsymbol{x} + iJ(\boldsymbol{x}), \boldsymbol{y} - iJ(\boldsymbol{y})) \\
&= \tfrac{1}{2}d(\boldsymbol{x},\boldsymbol{y}) + \tfrac{i}{2}s(\boldsymbol{x},\boldsymbol{y}) - \tfrac{i}{2}s(\boldsymbol{y},\boldsymbol{x}) + \tfrac{1}{2}d(J(\boldsymbol{x}),J(\boldsymbol{y})) \\
&= d(\boldsymbol{x},\boldsymbol{y}) + i\,s(\boldsymbol{x},\boldsymbol{y}) = \langle \boldsymbol{x},\boldsymbol{y}\rangle_J.
\end{aligned}$$

Como $\dim_{\mathbb{C}} W = \frac{1}{2}\dim_{\mathbb{R}} W = \frac{1}{2}\dim_{\mathbb{R}} V = \dim_{\mathbb{C}} V_J$, se ve que $P_J \colon V_J \to W$ es una biyección lineal que entrelaza los productos escalares complejos de V_J y W. □

4.5 Ejercicios sobre formas bilineales

Ejercicio 4.1. Sean f, g dos formas bilineales sobre V, con f no degenerada. Demostrar que hay un único operador lineal $T \in \operatorname{End}(V)$ 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 sólo si g también es no degenerada.
⟦ Indicación: Fijar una base $\mathcal{B}$ de V y expresar la matriz de T en términos de las matrices de f y g. ⟧

Ejercicio 4.2. 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 $D := \det[d(\boldsymbol{x}_i,\boldsymbol{x}_j)]$. Verificar que la forma d es no degenerada si y sólo si $D \neq 0$.

Ejercicio 4.3. Sea d una forma bilineal simétrica sobre V. Para cada subespacio $M \leq V$, denótese por $M^\perp$ el subespacio ortogonal a M con respecto a d. Si N es otro subespacio de V, demostrar que $(M+N)^\perp = M^\perp \cap N^\perp$.

Demostrar también que $(M \cap N)^\perp = M^\perp + N^\perp$ si d es no degenerada.

Ejercicio 4.4. Las formas bilineales simétricas aparecen en la teoría geométrica de *polos y polares*. 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 determinado punto, la *recta polar* de $\boldsymbol{y}$ con respecto a esta cónica[7] es la recta con ecuación $d(\boldsymbol{x},\boldsymbol{y}) = 1$.

Por ejemplo, si la cónica es la hipérbola $x_1^2 - 4x_1x_2 + 2x_2^2 = 1$, la recta polar del punto $(2,3)$ es la recta $-4x_1 + 2x_2 = 1$.

Verificar que un punto $\boldsymbol{y} \in \mathbb{R}^2$ queda sobre la curva $q(\boldsymbol{x}) = 1$ si y sólo si $\boldsymbol{y}$ queda sobre su propia recta polar. Concluir que esa recta polar es tangencial a la cónica en ese punto.

Ejercicio 4.5. Verificar que las formas lineales $f_1,\dots,f_k$, construidas en la Proposición 4.20 por el proceso de reducción de Lagrange, son linealmente independientes.

Ejercicio 4.6. 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) $q(x_1,x_2,x_3) = -2x_1^2 + 6x_1x_2 + 10x_1x_3 + x_2^2 - 2x_2x_3 + 4x_3^2$,

(c) $q(x_1,x_2,x_3) = x_1x_2 + x_1x_3 + x_2x_3$,

(d) $q(x_1,x_2,x_3,x_4) = x_1^2 + 2x_1x_2 - 2x_1x_3 + 2x_1x_4 + 4x_2^2 + 4x_2x_3 + 2x_2x_4 + 4x_3^2 - 2x_3x_4 - x_4^2$,

(e) $q(x_1,x_2,x_3,x_4) = 4x_1^2 - 4x_1x_2 - 4x_1x_3 + 4x_1x_4 + x_2^2 + 4x_2x_3 - 4x_2x_4 + x_3^2 + x_4^2$.

[7] Para más información sobre polos y polares, véase, por ejemplo, el Tema VI de: Joseph C. Várilly, *Elementos de Geometría Plana*, Editorial de la UCR, San José, 1988.

Ejercicio 4.7. ¿Cuáles son el rango y la signatura de cada una de las formas cuadráticas del Ejercicio anterior?

Ejercicio 4.8. 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) $q_{a,b}(\boldsymbol{x}) = a\sum_{k=1}^{n} x_k^2 + b\sum_{i<j} x_ix_j$, con $a, b \in \mathbb{R}$.

(Hay varios casos, según los valores de a y b).

Ejercicio 4.9. Determinar el rango y la signatura de la forma cuadrática sobre $\mathbb{R}^n$ cuya definición es $q(\boldsymbol{x}) := \sum_{i<j}(x_i - x_j)^2$.

Ejercicio 4.10. Gantmacher da la siguiente receta[8] para determinar la signatura de la forma cuadrática $q(\boldsymbol{x}) = \boldsymbol{x}^tA\boldsymbol{x}$ de rango k sobre $\mathbb{R}^n$. Sean $D_1, \dots, D_n$ los menores principales delanteras de la matriz A. [Esto es, $D_r := m_{J_r,J_r}$ es el determinante de la submatriz de A formado al borrar las últimas $(n-r)$ filas y columnas. Se sabe que $D_r = 0$ para $r > k$.]

Supóngase que $D_r \neq 0$ para $r = 1, \dots, k$ o bien que la lista $(D_1, \dots, D_k)$ tenga ceros no consecutivos. Sea q el *número de cambios de signo* en la lista $(1, D_1, \dots, D_k)$, después de suprimir los ceros no consecutivos si los hubiese. Entonces la signatura de la forma cuadrática es $s := k - 2q$.

Usar esta prescripción para calcular la signatura de las formas cuadráticas del Ejercicio 4.6, sin aplicar la reducción de Lagrange.

Ejercicio 4.11. Sean $\mu_1, \dots, \mu_r$ los autovalores distintos de la matriz simétrica $A \in M_n(\mathbb{R})$, en orden decreciente: $\mu_1 > \mu_2 > \cdots > \mu_r$. Demostrar que la forma cuadrática $q(\boldsymbol{x}) := \boldsymbol{x}^tA\boldsymbol{x}$ obedece

$$\lambda_r\, \boldsymbol{x}^t\boldsymbol{x} \leq \boldsymbol{x}^tA\boldsymbol{x} \leq \lambda_1\, \boldsymbol{x}^t\boldsymbol{x},$$

y que tiene los valores máximo y mínimo de $q(\boldsymbol{x})$ sobre la esfera $\boldsymbol{x}^t\boldsymbol{x} = 1$ son λ_1 y λ_r, respectivamente.

⟦ Indicación: Recordar que A puede ser diagonalizada por una matriz ortogonal. ⟧

Ejercicio 4.12. Encontrar una matriz ortogonal $Q \in M_3(\mathbb{R})$ tal que $Q^{-1}AQ$ sea diagonal, donde

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

Luego, hallar los valores máximo y mínimo de la función $q(x, y, z) := x^2 + xy + xz + y^2 + yz + z^2$ sobre la esfera $x^2 + y^2 + z^2 = 1$.

[8]Referencia: Feliks Gantmacher, *The Theory of Matrices*, tomo 1, Chelsea, New York, 1959; pp. 303–304.

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. Sea V un espacio vectorial de dimensión par $2m$ sobre $\mathbb{F}$ y sea s una forma simpléctica sobre V.

(a) Si N es un subespacio de V, y si $N^\perp$ denota su complemento simpléctico con respecto a s, demostrar que $(N^\perp)^\perp = N$.

(b) Demostrar que hay un subespacio $M \leq V$ de dimensión m que es isotrópico[9] respecto de s. Concluir que $M^\perp = M$ y que M es un subespacio isotrópico maximal.
[[Indicación: Usar la base construida en la Proposición 4.32.]]

Ejercicio 4.15. Para la siguiente matriz antisimétrica $A \in M_4(\mathbb{F})$,

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

demostrar que $A \simeq J_4$ y encontrar una matriz inversible $P \in M_4(\mathbb{F})$ tal que $P^tAP = J_4$.

Ejercicio 4.16. (a) Sea $B + C \in M_n(\mathbb{F})$ una matriz inversible, donde $B^t = B$ y $C^t = -C$ son sus partes simétrica y antisimétrica. Sea $P := (B+C)^{-1}(B-C)$. Verificar las relaciones

$$P^t(B+C)P = B+C, \qquad P^t(B-C)P = B-C.$$

(b) Si $A \in M_n(\mathbb{R})$ es una matriz antisimétrica, comprobar que 1 y (-1) no son autovalores de A; concluir que las matrices $I_n - A$ y $I_n + A$ son inversibles.

(c) Demostrar que la **transformada de Cayley** de A, dada por $Q := \underline{(I_n + A)^{-1}(I_n - A)}$, es una matriz ortogonal.

Ejercicio 4.17. (a) Si $f(t), g(t) \in \mathbb{F}[t]$ son dos polinomios y si $B \in M_n(\mathbb{F})$ es una matriz tal que $f(B)$ sea inversible en $M_n(\mathbb{F})$, demostrar que $f(B)^{-1}g(B) = g(B)f(B)^{-1}$.

(b) Si $Q \in M_n(\mathbb{R})$ es una matriz ortogonal tal que 1 y (-1) no sean autovalores de Q, demostrar que su transformada de Cayley $A := (I_n + Q)^{-1}(I_n - Q)$ es una matriz antisimétrica.

Ejercicio 4.18. Se dice que $R \in M_{2m}(\mathbb{R})$ es una **matriz simpléctica** si $s(R\boldsymbol{x}, R\boldsymbol{y}) = s(\boldsymbol{x}, \boldsymbol{y})$ para todo $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^{2m}$, donde

$$s(\boldsymbol{x}, \boldsymbol{y}) := \boldsymbol{x}^t J_{2m} \boldsymbol{y} = x_1y_2 - x_2y_1 + x_3y_4 - x_4y_3 + \cdots + x_{2m-1}y_{2m} - x_{2m}y_{2m-1}$$

es la forma simpléctica estándar sobre $\mathbb{R}^{2m}$. Fíjese que la matriz R es simpléctica si y sólo si $R^tJ_{2m}R = J_{2m}$.

(a) Mostrar que R es inversible, que R^{-1} también es simpléctica, y que el producto de dos matrices simplécticas es otra matriz simpléctica.

(b) Verificar que una matriz simpléctica R cumple $\det R = +1$.

[9] Recordar que N es *isotrópico* si $s(\boldsymbol{x}, \boldsymbol{x}) = 0$ para $\boldsymbol{x} \in N$, o equivalentemente, si $N \subseteq N^\perp$.

Ejercicio 4.19. Considérese el espacio vectorial $\mathbb{R}^{2m}$ con su producto escalar real estándar $d(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^t\boldsymbol{y}$ y su forma simpléctica estándar $s(\boldsymbol{x},\boldsymbol{y}) = \boldsymbol{x}^t J_{2m}\boldsymbol{y}$. Sean $A,B \in M_{2m}(\mathbb{R})$ tales que $AJ_{2m} = J_{2m}A$ y $BJ_{2m} = -J_{2m}B$. Verificar las siguientes reglas de transposición para s:

$$s(\boldsymbol{x},A^t\boldsymbol{y}) = s(A\boldsymbol{x},\boldsymbol{y}), \qquad s(\boldsymbol{x},B^t\boldsymbol{y}) = s(\boldsymbol{y},B\boldsymbol{x}).$$

Ejercicio 4.20. Si $A \in M_m(\mathbb{R})$, demostrar que

$$\operatorname{Pf}\begin{bmatrix} O & A \\ -A^t & O \end{bmatrix} = (-1)^{m(m-1)/2} \det A.$$

⟦ Indicación: Factorizar la matriz al lado izquierdo como un producto R^tJR para ciertas matrices apropiadas $R,J \in M_{2m}(\mathbb{R})$. ⟧

Ejercicio 4.21. Si A es una matriz antisimétrica con $\operatorname{Pf} A \neq 0$, mostrar que A es inversible y que A^{-1} es antisimétrica, con $\operatorname{Pf}(A^{-1}) = 1/\operatorname{Pf}(-A)$.

Ejercicio 4.22. Hay una fórmula inductiva que define el pfaffiano de una matriz antisimétrica A por expansión en filas *y columnas*. 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; escríbase $p_{ij} := \operatorname{Pf}(A_{ij,ij})$. La regla de expansión es:

$$\operatorname{Pf} A = a_{12}p_{12} - a_{13}p_{13} + \cdots + (-1)^n a_{1n}p_{1n} = \sum_{j=2}^{n} (-1)^j a_{1j}p_{1j}.$$

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 (4.10) 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, J es una estructura compleja ortogonal, s es la forma simpléctica sobre V definido por $s(\boldsymbol{x},\boldsymbol{y}) := d(J\boldsymbol{x},\boldsymbol{y})$, $V_{\mathbb{C}}$ denota la complexificación de V. Cada operador $\mathbb{R}$-lineal $T \in \operatorname{End}_{\mathbb{R}}(V)$ se puede ampliar a un operador $\mathbb{C}$-lineal, $T \in \operatorname{End}_{\mathbb{C}}(V_{\mathbb{C}})$, mediante la redefinición $T(\boldsymbol{x} + i\boldsymbol{y}) := T(\boldsymbol{x}) + iT(\boldsymbol{y})$ para $\boldsymbol{x},\boldsymbol{y} \in V$.

Ejercicio 4.23. Si $R \in \operatorname{End}_{\mathbb{R}}(V)$ es un operador ortogonal, es decir, $d(R(\boldsymbol{x}),R(\boldsymbol{y})) = d(\boldsymbol{x},\boldsymbol{y})$ para todo $\boldsymbol{x},\boldsymbol{y} \in V$, demostrar que R es inversible y ortogonal, y que RJR^{-1} es otra estructura compleja ortogonal sobre V.

Ejercicio 4.24. Para $V = \mathbb{R}^4$ con el producto escalar usual, sean α,β dos ángulos cualesquiera; demostrar que las siguientes dos matrices determinan estructuras complejas ortogonales:[10]

$$J_{\alpha,\beta} = \begin{bmatrix} 0 & -\cos\alpha & -\sin\alpha\cos\beta & -\sin\alpha\sin\beta \\ \cos\alpha & 0 & -\sin\alpha\sin\beta & \sin\alpha\cos\beta \\ \sin\alpha\cos\beta & \sin\alpha\sin\beta & 0 & -\cos\alpha \\ \sin\alpha\sin\beta & -\sin\alpha\cos\beta & \cos\alpha & 0 \end{bmatrix},$$

$$J'_{\alpha,\beta} = \begin{bmatrix} 0 & -\cos\alpha & -\sin\alpha\cos\beta & -\sin\alpha\sin\beta \\ \cos\alpha & 0 & \sin\alpha\sin\beta & -\sin\alpha\cos\beta \\ \sin\alpha\cos\beta & -\sin\alpha\sin\beta & 0 & \cos\alpha \\ \sin\alpha\sin\beta & \sin\alpha\cos\beta & -\cos\alpha & 0 \end{bmatrix}.$$

Ejercicio 4.25. Sea V_J el propio espacio vectorial V dotado del producto escalar complejo $\langle\cdot,\cdot\rangle_J = d + is$. Un operador $\mathbb{R}$-lineal $Q \in \operatorname{End}_{\mathbb{R}}(V)$ define un operador $\mathbb{C}$-*lineal* sobre V_J si y sólo si $QJ = JQ$; en cambio, Q define un operador $\mathbb{C}$-*semilineal* sobre V_J si y sólo si $QJ = -JQ$.

(a) Sean $R := \frac{1}{2}(Q - JQJ)$, $S := \frac{1}{2}(Q + JQJ)$. Fíjese que $Q = R + S$. Verificar que R es $\mathbb{C}$-lineal y que S es $\mathbb{C}$-semilineal como operadores sobre V_J.

(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

$$RR^t - SS^t = R^tR + S^tS = I, \qquad RS^t = -SR^t, \quad R^tS = -S^tR.$$

[[Indicación: Estudiar la relación $QQ^{-1} = Q^{-1}Q = I$.]]

Ejercicio 4.26. Si $W \leq V_{\mathbb{C}}$ es un subespacio, escríbase $\overline{W} := \{\boldsymbol{x} - i\boldsymbol{y} : \boldsymbol{x},\boldsymbol{y} \in V;\ \boldsymbol{x} + i\boldsymbol{y} \in W\}$. Una **polarización** de $V_{\mathbb{C}}$ es un subespacio complejo $W \leq V_{\mathbb{C}}$ que es d-isotrópico,[11] tal que $W \cap \overline{W} = \{\boldsymbol{0}\}$ y $W \oplus \overline{W} = V_{\mathbb{C}}$.

(a) Mostrar que $W_J := P_J(V) = \{\boldsymbol{x} - iJ\boldsymbol{x} : \boldsymbol{x} \in V\}$ es una polarización de $V_{\mathbb{C}}$, con $\overline{W_J} = W_{-J}$.

(b) Demostrar que W_J y W_{-J} son los subespacios de autovectores para el operador ampliado $J \in \operatorname{End}_{\mathbb{C}}(V_{\mathbb{C}})$ con los respectivos autovalores i y $-i$.

(c) Si $Q \in \operatorname{End}_{\mathbb{R}}(V)$ es 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 $V_{\mathbb{C}}$. Comprobar que $Q(W_J) = P_{QJQ^{-1}}(V)$.

(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 $\boldsymbol{x} + i\boldsymbol{y}_1 \in W$ y $\boldsymbol{x} + i\boldsymbol{y}_2 \in W$, mostrar que $\boldsymbol{y}_1 = \boldsymbol{y}_2$. Concluir que hay un operador $J_W \in \operatorname{End}_{\mathbb{R}}(V)$ determinado por $J_W(\boldsymbol{x}) := -\boldsymbol{y}$ cuando $\boldsymbol{x} + i\boldsymbol{y} \in W$.

(e) Dada una polarización W de $V_{\mathbb{C}}$, verificar que J_W es una estructura compleja ortogonal sobre V.

[10] Estas son *todas* las estructuras complejas ortogonales sobre $\mathbb{R}^4$. Geométricamente, forman dos copias disjuntas de la esfera bidimensional $\mathbb{S}^2$, en las cuales (α,β) son coordenadas esféricas.

[11] Se puede ampliar d a una forma bilineal simétrica sobre $V_{\mathbb{C}}$ mediante la fórmula evidente $d(\boldsymbol{x}+i\boldsymbol{y},\boldsymbol{x}'+i\boldsymbol{y}') := d(\boldsymbol{x},\boldsymbol{x}') + id(\boldsymbol{x},\boldsymbol{y}') + id(\boldsymbol{y},\boldsymbol{x}') - d(\boldsymbol{y},\boldsymbol{y}')$. Un subespacio W es isotrópico para d si $d(\boldsymbol{z},\boldsymbol{w}) = 0$ para todo $\boldsymbol{z},\boldsymbol{w} \in W$.

5 Algebras Exteriores y de Clifford

Hay varias maneras de enriquecer la teoría de espacios vectoriales al introducir una operación de producto de vectores, compatible con la operación de suma; pero en general el producto de dos vectores no es un vector. Dicho de otra forma, es posible extender un espacio vectorial V al incluirlo dentro de un álgebra más grande. En este capítulo se examinará algunas de estas posibilidades. El álgebra exterior $\Lambda^\bullet V$ extiende V mediante un producto anticonmutativo. En presencia de una forma cuadrática q sobre V, el álgebra de Clifford $C\ell(V,q)$ extiende V de otra manera, que depende esencialmente de la signatura de q.

5.1 Formas multilineales alternadas

Definición 5.1. Sea V un espacio vectorial sobre un cuerpo $\mathbb{F}$. Sea $\underline{V^k} := V \times V \times \cdots \times V$ el producto cartesiano de k copias de V, para $k = 1,2,3,\dots$. Una **forma k-lineal** sobre V es una aplicación $g\colon V^k \to \mathbb{F}$ tal que cada aplicación parcial $\boldsymbol{x}_j \mapsto g(\boldsymbol{x}_1,\dots,\boldsymbol{x}_j,\dots,\boldsymbol{x}_k)$ es lineal,[1] para $j = 1,\dots,k$.

Una **forma multilineal** sobre V es una forma k-lineal, para algún k. Este concepto incluye las formas lineales ($k = 1$) y las formas bilineales ($k = 2$).

Definición 5.2. Si $f\colon V^k \to \mathbb{F}$ y $g\colon V^r \to \mathbb{F}$ son dos formas multilineales, su **producto tensorial** es la forma $(k+r)$-lineal $\underline{f \otimes g}$ dada por

$$(f \otimes g)(\boldsymbol{x}_1,\dots,\boldsymbol{x}_{k+r}) := f(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k)\, g(\boldsymbol{x}_{k+1},\dots,\boldsymbol{x}_{k+r}).$$

En particular, el producto tensorial de dos formas lineales es una forma bilineal:

$$(f \otimes g)(\boldsymbol{x},\boldsymbol{y}) := f(\boldsymbol{x})\, g(\boldsymbol{y}). \tag{5.1}$$

Si $f\colon V^k \to \mathbb{F}$, $g\colon V^r \to \mathbb{F}$ y $h\colon V^s \to \mathbb{F}$ son tres formas multilineales, es claro que las formas $(k+r+s)$-lineales $(f \otimes g) \otimes h$ y $f \otimes (g \otimes h)$ coinciden, y se puede denotar esta forma por $\underline{f \otimes g \otimes h}$ simplemente.

En vista del isomorfismo $V \simeq V^{**}$ para un espacio vectorial finitodimensional V, que identifica V con el espacio dual de V^*, se puede convertir la fórmula (5.1) en una definición del producto tensorial de dos vectores, mediante la Definición siguiente.

Definición 5.3. Sean V y W dos espacios vectoriales finitodimensionales sobre $\mathbb{F}$. Sea $B(V,W)$ la totalidad de aplicaciones bilineales $h\colon V \times W \to \mathbb{F}$. Si $\boldsymbol{x} \in V$, $\boldsymbol{y} \in W$, entonces $\underline{\boldsymbol{x} \otimes \boldsymbol{y}} : h \mapsto h(\boldsymbol{x},\boldsymbol{y})$ es lineal, así que pertenece al espacio dual $B(V,W)^*$. El subespacio generado por estos elementos es el **producto tensorial** de V y W, denotado $\underline{V \otimes W}$. Cualquier elemento de $V \otimes W$ es una suma finita $\sum_{j=1}^r \boldsymbol{x}_j \otimes \boldsymbol{y}_j$ de estos "tensores simples", que cumplen las siguientes propiedades de combinación:

$$\begin{aligned} (\boldsymbol{x}_1 + \boldsymbol{x}_2) \otimes \boldsymbol{y} &= \boldsymbol{x}_1 \otimes \boldsymbol{y} + \boldsymbol{x}_2 \otimes \boldsymbol{y}, \\ \boldsymbol{x} \otimes (\boldsymbol{y}_1 + \boldsymbol{y}_2) &= \boldsymbol{x} \otimes \boldsymbol{y}_1 + \boldsymbol{x} \otimes \boldsymbol{y}_2, \\ c(\boldsymbol{x} \otimes \boldsymbol{y}) &= c\boldsymbol{x} \otimes \boldsymbol{y} = \boldsymbol{x} \otimes c\boldsymbol{y}, \end{aligned} \quad \text{para todo} \quad \begin{cases} \boldsymbol{x}, \boldsymbol{x}_1, \boldsymbol{x}_2 \in V, \\ \boldsymbol{y}, \boldsymbol{y}_1, \boldsymbol{y}_2 \in W, \\ c \in \mathbb{F}. \end{cases}$$

[1] Más generalmente, se puede definir una aplicación k-lineal $T\colon V_1 \times V_2 \times \cdots \times V_k \to W$, en donde $V_1,\dots,V_k$ y W son diversos espacios vectoriales sobre $\mathbb{F}$.

La expresión $\sum_{j=1}^{r} \boldsymbol{x}_j \otimes \boldsymbol{y}_j$ para un elemento de $V \otimes W$ no es única, pero se puede suponer que los $\boldsymbol{x}_1, \dots, \boldsymbol{x}_r$ son linealmente independientes en V y que los $\boldsymbol{y}_1, \dots, \boldsymbol{y}_r$ son linealmente independientes en W. En consecuencia, se ve que $\dim(V \otimes W) = (\dim V)(\dim W)$ y por tanto los espacios vectoriales $V \otimes W$ y $B(V,W)^*$ coinciden.

Se identifica $\mathbb{F} \otimes V$ y $V \otimes \mathbb{F}$ con V, al identificar $1 \otimes \boldsymbol{x} \leftrightarrow \boldsymbol{x} \otimes 1 \leftrightarrow \boldsymbol{x}$ para $\boldsymbol{x} \in V$. Si V, W, U son tres espacios vectoriales sobre $\mathbb{F}$, la evaluación $h \mapsto h(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z})$ de una forma trilineal h en tres vectores es el elemento $(\boldsymbol{x} \otimes \boldsymbol{y}) \otimes \boldsymbol{z} = \boldsymbol{x} \otimes (\boldsymbol{y} \otimes \boldsymbol{z}) =: \boldsymbol{x} \otimes \boldsymbol{y} \otimes \boldsymbol{z}$ del producto tensorial $(V \otimes W) \otimes U = V \otimes (W \otimes U) = V \otimes W \otimes U$. De este modo, se puede escribir el producto tensorial $V_1 \otimes V_2 \otimes \cdots \otimes V_r$ de varios espacios vectoriales sin emplear paréntesis.

Notación. El grupo S_k de permutaciones de k objetos actúa sobre las formas k-lineales por reordenación de sus argumentos. Si $\sigma \in S_k$ es una permutación y si $f \colon V^k \to \mathbb{F}$ es una forma k-lineal, escríbase

$$(\sigma \cdot f)(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) := f(\boldsymbol{x}_{\sigma^{-1}(1)}, \dots, \boldsymbol{x}_{\sigma^{-1}(k)}).$$

En los argumentos se usa la permutación inversa σ^{-1} para que valga la identidad $\tau \cdot (\sigma \cdot f) = (\tau\sigma) \cdot f$ para todo $\tau, \sigma \in S_k$. Es claro que $\mathrm{Id} \cdot f = f$, es decir, la permutación trivial Id actúa trivialmente.[2]

Definición 5.4. Una forma k-lineal $f \colon V^k \to \mathbb{F}$ es **simétrica** si

$$f(\boldsymbol{x}_{\sigma(1)}, \dots, \boldsymbol{x}_{\sigma(k)}) = f(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \quad \text{para todo} \quad \boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in V,\ \sigma \in S_k,$$

o equivalentemente, si $\sigma \cdot f = f$ para todo $\sigma \in S_k$.

Una forma k-lineal $g \colon V^k \to \mathbb{F}$ es **alternada** si

$$g(\boldsymbol{x}_{\sigma(1)}, \dots, \boldsymbol{x}_{\sigma(k)}) = (-1)^\sigma g(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) \quad \text{para todo} \quad \boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in V,\ \sigma \in S_k, \tag{5.2}$$

o equivalentemente, si $\sigma \cdot g = (-1)^\sigma g$ para todo $\sigma \in S_k$.

Definición 5.5. Si $h \colon V^k \to \mathbb{F}$ es una forma k-lineal cualquiera, se obtiene una forma simétrica por **simetrización**:[3]

$$(\mathbb{S}h)(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) := \frac{1}{k!} \sum_{\sigma \in S_k} h(\boldsymbol{x}_{\sigma(1)}, \dots, \boldsymbol{x}_{\sigma(k)}).$$

Al poner $\tau = \sigma^{-1}$, esto es $\mathbb{S}h := (1/k!) \sum_\tau \tau \cdot h$. Es fácil verificar que $\mathbb{S}h$ es k-lineal y simétrica, y que una forma k-lineal f es simétrica si y sólo si $\mathbb{S}f = f$.

[2]En general, una *acción de un grupo G* sobre un conjunto X es una función $G \times X \to X : (g, x) \mapsto g \cdot x$ que cumple las dos reglas (a) $1 \cdot x = x$; (b) $g \cdot (h \cdot x) = (gh) \cdot x$ para $g, h \in G$, $x \in X$. La asignación $(\sigma, f) \mapsto \sigma \cdot f$ es una acción del grupo S_n en este sentido.

[3]El coeficiente $1/k!$ en estas fórmulas es convencional. Sin embargo, es importante notar que algunos autores lo omiten, en cuyo caso los coeficientes factoriales en las fórmulas que siguen no son iguales que los nuestros. Véase, por ejemplo: Jean Dieudonné, *Éléments d'Analyse*, tomo 3, inciso A.12.

También se puede fabricar una forma alternada por **antisimetrización**:

$$(\mathbb{A}h)(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k) := \frac{1}{k!}\sum_{\sigma\in S_k}(-1)^\sigma\, h(\boldsymbol{x}_{\sigma(1)},\dots,\boldsymbol{x}_{\sigma(k)}), \tag{5.3}$$

o bien $\mathbb{A}h := (1/k!)\sum_\tau(-1)^\tau\tau\cdot h$. Es fácil verificar que $\mathbb{A}h$ es k-lineal y alternada, y que una forma k-lineal g es alternada si y sólo si $\mathbb{A}g = g$.

Ejemplo 5.6. Considérese el *determinante* de una matriz $A\in M_n(\mathbb{F})$ en función de sus columnas $\boldsymbol{a}_1,\dots,\boldsymbol{a}_n\in\mathbb{F}^n$. La fórmula de Leibniz (1.15), hace evidente que $\underline{\det A}$ depende linealmente de cada columna, y que es *una n-forma alternada* sobre $\mathbb{F}^n$.

Al omitir el signo $(-1)^\sigma$ en la fórmula de Leibniz, se obtiene el **permanente** de A,

$$\operatorname{per}A := \sum_{\sigma\in S_n} a_{1j_1}a_{2j_2}\dots a_{nj_n}.$$

Esta función es una n-forma simétrica sobre el espacio de columnas $\mathbb{F}^n$.

Definición 5.7. Si $f\colon V^k\to\mathbb{F}$ y $g\colon V^r\to\mathbb{F}$ son dos formas multilineales alternadas, su **producto exterior** es la forma $(k+r)$-lineal alternada definido por

$$f\wedge g := \frac{(k+r)!}{k!\,r!}\mathbb{A}(f\otimes g).$$

En particular, el producto exterior de dos formas *lineales* $f,g\in V^*$ es la forma *bilineal* $f\wedge g = 2\mathbb{A}(f\otimes g) = f\otimes g - g\otimes f$, es decir,

$$(f\wedge g)(\boldsymbol{x},\boldsymbol{y}) := f(\boldsymbol{x})\,g(\boldsymbol{y}) - f(\boldsymbol{y})\,g(\boldsymbol{x}). \tag{5.4}$$

Nótese la *anticonmutatividad*: $g\wedge f = -f\wedge g$ para $f,g\in V^*$.

Proposición 5.8. *El producto exterior de formas multilineales alternadas es* ***asociativa****: si $f\colon E^k\to\mathbb{F}$, $g\colon E^r\to\mathbb{F}$, $h\colon E^s\to\mathbb{F}$ son alternadas, entonces*

$$(f\wedge g)\wedge h = f\wedge(g\wedge h).$$

Demostración. La asociatividad del producto tensorial permite calcular:

$$\begin{aligned}
(f\wedge g)\wedge h &= \frac{(k+r+s)!}{(k+r)!\,s!}\mathbb{A}((f\wedge g)\otimes h) = \frac{(k+r+s)!}{(k+r)!\,s!}\frac{(k+r)!}{k!\,r!}\mathbb{A}(\mathbb{A}(f\otimes g)\otimes h) \\
&= \frac{(k+r+s)!}{k!\,r!\,s!}\frac{1}{(k+r+s)!}\frac{1}{(k+r)!}\sum_{\sigma,\tau}(-1)^\sigma(-1)^\tau\sigma\cdot(\tau\cdot(f\otimes g)\otimes h) \\
&= \frac{1}{k!\,r!\,s!\,(k+r)!}\sum_{\sigma,\tau}(-1)^{\sigma\tau}(\sigma\tau)\cdot(f\otimes g\otimes h) \\
&= \frac{1}{k!\,r!\,s!}\sum_{\rho\in S_{k+r+s}}(-1)^\rho\rho\cdot(f\otimes g\otimes h).
\end{aligned}$$

En la penúltima igualdad, se identifica $\tau \in S_{k+r}$ con la permutación correspondiente en S_{k+r+s} que deja fijos los últimos s objetos. Colóquese $\rho := \sigma\tau$ y fíjese que para cada $\tau \in S_{k+r}$, la suma sobre $\sigma \in S_{k+r+s}$ da la misma contribución al lado derecho; luego la sumatoria sobre ρ aparece repetida $(k+r)!$ veces, cancelado así el factor $(k+r)!$ en el denominador.

De igual manera, se calcula que

$$f \wedge (g \wedge h) = \frac{1}{k!\,r!\,s!} \sum_{\rho \in S_{k+r+s}} (-1)^{\rho} \rho \cdot (f \otimes g \otimes h). \qquad \square$$

Ejemplo 5.9. Si $f_1, \dots, f_k \in V^*$, entonces $f_1 \otimes \cdots \otimes f_k$ es una forma k-lineal sobre V. Como el producto exterior es asociativo, también se escribe $\underline{f_1 \wedge \cdots \wedge f_k}$ sin paréntesis. Por inducción sobre k, se verifica la fórmula

$$f_1 \wedge \cdots \wedge f_k = k!\, \mathbb{A}(f_1 \otimes \cdots \otimes f_k) \quad \text{para} \quad f_1, \dots, f_k \in V^*. \tag{5.5}$$

De ahí se ve que $f_{\sigma(1)} \wedge \cdots \wedge f_{\sigma(k)} = (-1)^{\sigma} f_1 \wedge \cdots \wedge f_k$ para $\sigma \in S_k$. Si se toman los f_j de entre una base $\{f_1, \dots, f_n\}$ de V^*, donde $n = \dim V$, el número de productos $f_1 \wedge \cdots \wedge f_k$ que son linealmente independientes es entonces el número de maneras de elegir k vectores de la base, sin repetición pero olvidando su orden; es decir, es el *coeficiente binomial* $\binom{n}{r}$.

Lema 5.10. *El producto exterior de formas alternadas es* ***superconmutativa****:*[4] *si* $f : E^k \to \mathbb{F}$, $g : E^r \to \mathbb{F}$ *son formas alternadas, entonces*

$$g \wedge f = (-1)^{kr} f \wedge g. \tag{5.6}$$

Demostración. Sea σ la permutación de baraje que intercambia $\{1, \dots, k\}$ con $\{k+1, \dots, k+r\}$; es decir, $\sigma(i) := i+k$ si $i \leq k$, $\sigma(i) := i-k$ si $i > k$. Ahora σ es el producto de kr transposiciones, porque se necesitan r transposiciones para llevar cada uno de los k elementos iniciales a su posición final, luego $(-1)^{\sigma} = (-1)^{kr}$.

Si $f = f_1 \wedge \cdots \wedge f_k$, $g = g_1 \wedge \cdots \wedge g_r$ con $f_i, g_j \in V^*$, la relación (5.6) es una consecuencia directa de la anticonmutatividad $f_i \wedge g_j = -g_j \wedge f_i$ de formas lineales. El caso general sigue por linealidad, porque tales productos exteriores de formas lineales generan los espacios vectoriales de k-formas y r-formas alternadas, respectivamente. $\square$

Proposición 5.11. *Si* $f_1, \dots, f_k \in V^*$ *y* $\boldsymbol{x}_1, \dots, \boldsymbol{x}_k \in V$, *se verifica*

$$(f_1 \wedge \cdots \wedge f_k)(\boldsymbol{x}_1, \dots, \boldsymbol{x}_k) = \det[f_i(\boldsymbol{x}_j)], \tag{5.7}$$

donde el lado derecho es el determinante de la matriz cuya entrada (i, j) *es* $f_i(\boldsymbol{x}_j)$.

[4] Esta terminología inelegante se debe a Berezin. Una *superálgebra* es un álgebra $A = A_+ \oplus A_-$ en donde cada elemento a es la suma de una "parte par" $a_+ \in A_+$ y una "parte impar" $a_- \in A_-$, con la estipulación de que los elementos pares conmutan entre sí y conmutan con los elementos impares, mientras los elementos impares anticonmutan entre sí. La terminología más correcta, "álgebra $\mathbb{Z}_2$-graduada", es inmanejable. La moda del prefijo *super-* fue establecida en: Feliks Aleksandrovich Berezin, *The Method of Second Quantization*, Academic Press, New York, 1966.

Demostración. De la fórmula (5.5) se obtiene

$$\begin{aligned}(f_1 \wedge \cdots \wedge f_k)(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k) &= k!\,\mathbb{A}(f_1 \otimes \cdots \otimes f_k)(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k) \\ &= \sum_\sigma (-1)^\sigma (f_{\sigma(1)} \otimes \cdots \otimes f_{\sigma(k)})(\boldsymbol{x}_1,\dots,\boldsymbol{x}_k) \\ &= \sum_\sigma (-1)^\sigma f_{\sigma(1)}(\boldsymbol{x}_1)\dots f_{\sigma(k)}(\boldsymbol{x}_k) \\ &= \det[f_i(\boldsymbol{x}_j)].\end{aligned}$$

La última igualdad no es más que la fórmula de Leibniz para el determinante. □

Definición 5.12. Denótese por $\underline{\Lambda^k V^*}$ el espacio vectorial de formas k-lineales alternadas $g\colon V^k \to \mathbb{F}$.

Si $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_k\}$ es una base de V, la forma $g \in \Lambda^k(V^*)$ depende solamente de los valores $g(\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_k})$, donde los argumentos son k elementos de la base dada. Por la antisimetría de g, es suficiente tomar elementos distintos, con sus índices en orden creciente:

$$1 \le i_1 < i_2 < \cdots < i_k \le n.$$

Hay $\binom{n}{k}$ maneras de elegir una parte $I = \{i_1,\dots,i_k\} \subseteq \{1,\dots,n\}$ con $|I| = k$. Luego la dimensión de $\Lambda^k V^*$ es $\binom{n}{k}$.

Además, si $\{f_1,\dots,f_n\}$ es la base dual de V^*, los elementos $\underline{f_I := f_{i_1} \wedge \cdots \wedge f_{i_k}}$ son linealmente independientes en vista de (5.7). Por lo tanto, forman una base de $\Lambda^k V^*$.

De nuevo, se puede aprovechar la dualidad entre V y V^* para definir el espacio vectorial $\underline{\Lambda^k V}$ como la totalidad de formas k-lineales alternadas $(V^*)^k \to \mathbb{F}$. Si $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ es una base de V, los productos exteriores $\underline{\boldsymbol{x}_I := \boldsymbol{x}_{i_1} \wedge \cdots \wedge \boldsymbol{x}_{i_k}}$ forman una base de $\Lambda^k V$.

Los elementos de $\Lambda^k V$ se llaman **k-vectores**. En particular, los elementos $\boldsymbol{x} \wedge \boldsymbol{y} \in \Lambda^2 V$ se llaman **bivectores**.

Proposición 5.13. *Para cada forma k-lineal alternante $g\colon E^k \to \mathbb{F}$, hay una única aplicación* lineal $\tilde{g}\colon \Lambda^k V \to \mathbb{F}$ *tal que*

$$\tilde{g}(\boldsymbol{y}_1 \wedge \cdots \wedge \boldsymbol{y}_k) = g(\boldsymbol{y}_1,\dots,\boldsymbol{y}_k) \quad \textit{para todo} \quad \boldsymbol{y}_1,\dots,\boldsymbol{y}_k \in V.$$

Demostración. La forma g queda determinada por los coeficientes $c_I := g(\boldsymbol{x}_{i_1},\dots,\boldsymbol{x}_{i_k})$, donde $I \subseteq \{1,\dots,n\}$ con $|I| = k$ y $\{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ es una base de V. Entonces $\tilde{g}(\boldsymbol{x}_I) := c_I$ necesariamente. Pero una forma lineal sobre $\Lambda^k V$ queda determinada por sus valores en una base, así que esta asignación de valores define la forma lineal $\tilde{g}$ deseada. □

De este modo, se identifica el espacio vectorial $\Lambda^k V^*$ de formas k-lineales alternantes con el espacio dual $(\Lambda^k V)^*$. Bajo esta identificación, la fórmula (5.7) dice simplemente que la base $\{f_I : |I| = k\}$ de $\Lambda^k V^*$ es la *base dual* a la base $\{\boldsymbol{x}_I : |I| = k\}$ de $\Lambda^k V$.

5.2 El álgebra exterior de un espacio vectorial

Definición 5.14. Sea V un espacio vectorial de dimensión n sobre $\mathbb{F}$. El **álgebra exterior** sobre V es el espacio vectorial

$$\Lambda^\bullet V := \bigoplus_{k=0}^{n} \Lambda^k V = \mathbb{F} \oplus V \oplus \Lambda^2 V \oplus \cdots \oplus \Lambda^n V, \tag{5.8}$$

que es la suma directa de las *potencias exteriores* $\Lambda^k V$ del espacio vectorial V, dotado con el producto exterior de multivectores. Un escalar $c \in \mathbb{F}$ se considera como elemento del espacio vectorial unidimensional $\Lambda^0 V$, con $c \wedge z := cz$ para $c \in \mathbb{F}$, $z \in \Lambda^\bullet V$.

Si $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V y si $\boldsymbol{x}_{i_1}, \dots, \boldsymbol{x}_{i_k} \in \mathcal{B}$, entonces $\boldsymbol{x}_{i_1} \wedge \cdots \wedge \boldsymbol{x}_{i_k} = \mathbf{0}$ en $\Lambda^k V$ cuando $k > n$. En efecto, si $k > n$, entonces algún índice $j \in \{1, \dots, n\}$ ocurre dos veces[5] en la lista $i_1, \dots, i_k$; como $\boldsymbol{x}_j \wedge \boldsymbol{x}_j = \mathbf{0}$ por antisimetría, el producto exterior de estos k vectores básicos se anula. Por linealidad en cada entrada, un producto arbitrario $\boldsymbol{y}_1 \wedge \cdots \wedge \boldsymbol{y}_k$ se anula cuando $\boldsymbol{y}_1, \dots, \boldsymbol{y}_k \in V$ con $k > n$. Por lo tanto, es $\Lambda^k V = \{\mathbf{0}\}$ para $k > n$. En la suma directa (5.8), aparecen todas las potencias exteriores no triviales de V.

Lema 5.15. *La dimensión de $\Lambda^\bullet V$ es $2^{\dim V}$.*

Demostración. Basta observar que, si $\dim V = n$, entonces

$$\dim \Lambda^\bullet V := \sum_{k=0}^{n} \dim \Lambda^k V = \sum_{k=0}^{n} \binom{n}{k} = (1+1)^n = 2^n. \qquad \square$$

Definición 5.16. Un **álgebra graduada** sobre $\mathbb{F}$ es un espacio vectorial A sobre $\mathbb{F}$, que posee un producto asociativo $\odot$ y una **graduación** $A =: \bigoplus_{k \in \mathbb{Z}} A_k$ tal que

$$\boldsymbol{x} \in A_k,\ \boldsymbol{y} \in A_r \implies \boldsymbol{x} \odot \boldsymbol{y} \in A_{k+r}, \quad \text{para todo} \quad k, r \in \mathbb{Z}.$$

Si algún sumando es trivial (es decir, si $A_k = \{\mathbf{0}\}$ para algún k), se omite ese índice en la suma directa.[6]

Fíjese que si $\boldsymbol{x} \in \Lambda^k V$, $\boldsymbol{y} \in \Lambda^r V$, entonces $\boldsymbol{x} \wedge \boldsymbol{y} \in \Lambda^{k+r} V$. En efecto, las Definiciones 5.7 y 5.2 muestran que $\boldsymbol{x} \wedge \boldsymbol{y} = \binom{k+r}{k} \boldsymbol{x} \otimes \boldsymbol{y}$ es una forma $(k+r)$-lineal alternada sobre V^*. Por lo tanto, el álgebra exterior $\Lambda^\bullet V$ es un álgebra graduada con un número finito de niveles no triviales.

[5]Esta es una instancia del *Schubfachsprinzip* de Dirichlet, o bien el "principio de las palomas y los palomares": si $k+1$ palomas se distribuyen entre k palomares, debe haber al menos un palomar que albergue al menos dos palomas.

[6]En la mayoría de los ejemplos conocidos, es $A_k = \{\mathbf{0}\}$ para $k < 0$. Se dice que el álgebra es "$\mathbb{N}$-graduado" en estos casos.

Definición 5.17. Si V es un espacio vectorial sobre $\mathbb{F}$, sea $V^{\otimes 2} := V \otimes V$, $V^{\otimes 3} := V \otimes V \otimes V$; en general, denótese por $V^{\otimes k}$ el producto tensorial de k copias de V, para $k = 2, 3, \dots$; además, sea $V^{\otimes 1} := V$ y $V^{\otimes 0} := \mathbb{F}$. Por la Definición 5.2 (dualizada al cambiar V^* por V), se ve que $\boldsymbol{x} \otimes \boldsymbol{y} \in V^{\otimes(k+r)}$ toda vez que $\boldsymbol{x} \in V^{\otimes k}$, $\boldsymbol{y} \in V^{\otimes r}$. (Si $c \in \mathbb{F} = V^{\otimes 0}$ y $\boldsymbol{x} \in V^{\otimes k}$, se toma $c \otimes \boldsymbol{x} \equiv c\boldsymbol{x} \in V^{\otimes k}$.

El álgebra graduada $\underline{\mathcal{T}(V)} := \bigoplus_{k=0}^{\infty} V^{\otimes k}$, cuyo producto es $\otimes$, se llama el **álgebra tensorial** generado por el espacio vectorial V. Obsérvese que esta álgebra es infinitodimensional si $V \neq \{\mathbf{0}\}$.

Ejemplo 5.18. Otra álgebra infinitodimensional es el **álgebra simétrica** $\underline{S^\bullet V} := \bigoplus_{k=0}^{\infty} S^k V$. Se define $S^k V^*$ como el espacio vectorial de formas k-lineales *simétricas* $f : V^k \to \mathbb{F}$; por dualidad, $S^k V$ es el espacio vectorial de formas k-lineales *simétricas* sobre V^*. Si $f : V^k \to \mathbb{F}$, $g : V^r \to \mathbb{F}$ son formas simétricas, defínase

$$f \vee g := \frac{(k+r)!}{k!\, r!}\, \mathbb{S}(f \otimes g),$$

la cual es una forma $(k+r)$-lineal simétrica. Si $h : V^s \to \mathbb{F}$ es una forma s-lineal simétrica, entonces $(f \vee g) \vee h = f \vee (g \vee h)$; para comprobarlo, en la demostración de la Proposición 5.8 se omite los signos de todas las permutaciones allí presentes. Se concluye que $S^\bullet V$ es un álgebra graduada. Tiene dimensión infinita si $V \neq \{\mathbf{0}\}$, debido al Lema siguiente.

Lema 5.19. *Si* $\dim V = n$ *y si* $k \in \mathbb{N}$, *entonces* $\dim S^k V = \dim S^k V^* = \binom{n+k-1}{k}$.

Demostración. Si $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base de V y si $\mathcal{B}^* := \{f_1, \dots, f_n\}$ es la base dual de V^*, entonces cada elemento de $S^k V^*$ es una combinación lineal de las formas k-lineales simétricas

$$f_{j_1} \vee f_{j_2} \vee \cdots \vee f_{j_k}, \quad \text{con} \quad 1 \leq j_1 \leq j_2 \leq \cdots \leq j_k \leq n. \tag{5.9}$$

Fíjese que en este caso, al contrario de lo que sucede con el productos exteriores, se admiten índices iguales en los productos simétricos de las formas lineales básicas, porque ningún $f_j \vee f_j$ se anula. Este producto simétrico es *conmutativo*, en contraste con la superconmutatividad del Lema 5.10. Por tanto, se puede abreviar $f^{\vee r} := f \vee f \vee \cdots \vee f$ (r veces). Con este convenio, y con $f^{\vee 0} := 1 \in \mathbb{F}$, se puede reorganizar los productos simétricos (5.9) de formas lineales básicas así:

$$f_1^{\vee r_1} \vee f_2^{\vee r_2} \vee \cdots \vee f_n^{\vee r_n}, \quad \text{con} \quad r_1 + r_2 + \cdots + r_n = k.$$

En otras palabras, $\dim S^k V^*$ es la cantidad total de *particiones* del número natural $k \in \mathbb{N}$ en n sumandos. Para contarlas, es cuestión de “colocar k bolas indistinguibles en n urnas”,[7] separadas por $(n-1)$ paredes:

$$[\bullet\bullet\bullet|\bullet\bullet||\bullet\bullet\bullet\bullet|\bullet|\bullet\bullet\bullet|||\bullet\bullet\bullet\bullet\bullet\bullet|\bullet\bullet]$$

Alternativamente, se puede contar el número de maneras de desplegar una fila de $k + (n-1)$ objetos y marcar k objetos de entre ellos como “bolas”; los objetos no marcados serán las

[7] Una buena cantidad de cálculos combinatoriales se reducen a problemas de colocar varios objetos (bolas) en ciertos receptáculos (urnas). Este conteo es un ejemplo clásico de un cálculo de esa naturaleza.

paredes. Esto es, se debe elegir k objetos de entre una lista de $n+k-1$ objetos dados; hay $\binom{n+k-1}{k}$ maneras de hacer esa elección.

Para obtener $\dim S^k V = \binom{n+k-1}{k}$, se intercambian los papeles de V y de V^*, que tienen la misma dimensión n. $\square$

Definición 5.20. Si A es un álgebra sobre $\mathbb{F}$, un **ideal** de A es un subespacio J tal que $a \in A$, $j \in J$ implican $aj \in J$ y $ja \in J$. El espacio vectorial cociente $A/J := \{a + J : a \in A\}$ es un álgebra también, porque la relación $(a_1 + j_1)(a_2 + j_2) = a_1a_2 + (a_1 j_2 + j_1a_2 + j_1 j_2)$ muestra que $(a_1 + J)(a_2 + J) = a_1a_2 + J$ en A/J.

Ejemplo 5.21. Sea J_S el ideal del álgebra $\mathcal{T}(V)$ generado por todos los elementos de la forma $\underline{\boldsymbol{x}\otimes\boldsymbol{y}-\boldsymbol{y}\otimes\boldsymbol{x}}$, para $\boldsymbol{x},\boldsymbol{y} \in V$. El álgebra cociente $\mathcal{T}(V)/J_S$ es isomorfo a $S^\bullet V$, porque la simetrización $\mathbb{S}\colon V^{\otimes k} \to S^k V$ tiene como núcleo el subespacio $V^{\otimes k} \cap J_S$. Es claro que $V \cap J_S = \{\boldsymbol{0}\}$, es decir, J_S no contiene elementos de nivel $k = 1$, así que $\mathbb{S}\colon V^{\otimes 1} \to S^1 V$ es simplemente la identidad $I\colon V \to V$.

Ejemplo 5.22. Sea J_Λ el ideal del álgebra $\mathcal{T}(V)$ generado por todos los elementos de la forma $\underline{\boldsymbol{x}\otimes\boldsymbol{y}+\boldsymbol{y}\otimes\boldsymbol{x}}$, para $\boldsymbol{x},\boldsymbol{y} \in V$. El álgebra cociente $\mathcal{T}(V)/J_\Lambda$ es isomorfo a $\Lambda^\bullet V$, porque la simetrización $\mathbb{A}\colon V^{\otimes k} \to \Lambda^k V$ tiene como núcleo el subespacio $V^{\otimes k} \cap J_\Lambda$. Es claro que $V \cap J_\Lambda = \{\boldsymbol{0}\}$, es decir, J_λ no contiene elementos de nivel $k = 1$, así que $\mathbb{S}\colon V^{\otimes 1} \to \Lambda^1 V$ es simplemente la identidad $I\colon V \to V$.

► Si V es un espacio vectorial de dimensión n, entonces el espacio vectorial $\Lambda^n V$ tiene dimensión $\binom{n}{n} = 1$. Si $\mathcal{B} = \{\boldsymbol{x}_1,\dots,\boldsymbol{x}_n\}$ es una base de V, una base de $\Lambda^n V$ tiene un solo elemento[8]

$$\underline{\mathrm{vol}}_{\mathcal{B}} := \boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \wedge \cdots \wedge \boldsymbol{x}_n.$$

Si $\mathcal{B}' = \{\boldsymbol{x}'_1,\dots,\boldsymbol{x}'_n\}$ es otra base de V y si $P = [I]^{\mathcal{B}}_{\mathcal{B}'}$ es la matriz de cambio de base, entonces $\boldsymbol{x}'_s = \sum_{j=1}^n p_{js}\boldsymbol{x}_j$ por (1.9), luego

$$\begin{aligned} \boldsymbol{x}'_1 \wedge \boldsymbol{x}'_2 \wedge \cdots \wedge \boldsymbol{x}'_n &= \sum_{j_1,\dots,j_n=1}^{n} p_{j_1,1}p_{j_2,2}\dots p_{j_n,n}\, \boldsymbol{x}_{j_1} \wedge \boldsymbol{x}_{j_2} \wedge \cdots \wedge \boldsymbol{x}_{j_n} \\ &= \sum_{\sigma\in S_n} (-1)^\sigma p_{\sigma(1),1}p_{\sigma(2),2}\dots p_{\sigma(n),n}\, \boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \wedge \cdots \wedge \boldsymbol{x}_n \\ &= (\det P^t)\,\boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \wedge \cdots \wedge \boldsymbol{x}_n = (\det P)\,\boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \wedge \cdots \wedge \boldsymbol{x}_n. \end{aligned}$$

En la primera sumatoria, los términos $\boldsymbol{x}_{j_1} \wedge \boldsymbol{x}_{j_2} \wedge \cdots \wedge \boldsymbol{x}_{j_n}$ con un índice repetido son ceros, por antisimetría: la suma se reduce a los multiíndices $\sigma = (j_1,\dots,j_n)$ que son permutaciones en S_n, en cuyo caso $\boldsymbol{x}_{j_1} \wedge \cdots \wedge \boldsymbol{x}_{j_n} = (-1)^\sigma \boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_n$ también por antisimetría. Por tanto, es

$$\underline{\mathrm{vol}}_{\mathcal{B}'} = (\det P)\,\underline{\mathrm{vol}}_{\mathcal{B}} = \det[I]^{\mathcal{B}}_{\mathcal{B}'}\,\underline{\mathrm{vol}}_{\mathcal{B}}.$$

[8] La notación $\underline{\mathrm{vol}}_{\mathcal{B}}$ indica que este elemento representa el volumen de un paralelepípedo cuyas aristas son los vectores de la base $\mathcal{B}$.

Definición 5.23. Dada una base $\mathcal{B}$ de V, cada elemento de $\Lambda^\bullet V$ es de la forma

$$\boldsymbol{c} = \sum_{k=0}^{n} \sum_{|I|=k} c_I \boldsymbol{x}_I,$$

en donde $\boldsymbol{x}_\emptyset := 1 \in \mathbb{F}$ y $\boldsymbol{x}_{1,\dots,n} = \underline{\mathrm{vol}}_{\mathcal{B}}$. El componente escalar de $\boldsymbol{c}$ es $c_\emptyset$ y el componente en $\Lambda^n V$ es $c_{1,\dots,n}$. La asignación $\boldsymbol{c} \mapsto c_{1,\dots,n}$ es una forma lineal sobre $\Lambda^\bullet V$, llamada la **integral de Berezin**[9] con respecto a la base $\mathcal{B}$:

$$\int_{\mathcal{B}}^{\wedge} \boldsymbol{c} := c_{1,\dots,n}.$$

Ejemplo 5.24. Sea V un espacio vectorial de dimensión 4, con base $\mathcal{B} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \boldsymbol{x}_3, \boldsymbol{x}_4\}$, y sea $b \in \Lambda^2 V$ un bivector. Entonces b es de la forma

$$b = b_{12}\, \boldsymbol{x}_1 \wedge \boldsymbol{x}_2 + b_{13}\, \boldsymbol{x}_1 \wedge \boldsymbol{x}_3 + b_{14}\, \boldsymbol{x}_1 \wedge \boldsymbol{x}_4 + b_{23}\, \boldsymbol{x}_2 \wedge \boldsymbol{x}_3 + b_{24}\, \boldsymbol{x}_2 \wedge \boldsymbol{x}_4 + b_{34}\, \boldsymbol{x}_3 \wedge \boldsymbol{x}_4.$$

Para $i < j$, sea $b_{ji} := -b_{ij}$, de manera que $b_{ij}\boldsymbol{x}_i \wedge \boldsymbol{x}_j = b_{ji}\boldsymbol{x}_j \wedge \boldsymbol{x}_i$; además, sea $b_{ii} := 0$ para $i = 1,2,3,4$. Entonces $B = [b_{ij}]$ es una matriz antisimétrica 4×4. El "cuadrado" $b \wedge b$ es un 4-vector en $\Lambda^4 V$ y por ende es proporcional a $\underline{\mathrm{vol}}_{\mathcal{B}}$. En efecto, se calcula que

$$b \wedge b = 2(b_{12}b_{34} - b_{13}b_{24} + b_{14}b_{23})\, \boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \wedge \boldsymbol{x}_3 \wedge \boldsymbol{x}_4 = 2(\operatorname{Pf} B)\, \underline{\mathrm{vol}}_{\mathcal{B}}.$$

Sea V un espacio vectorial de dimensión par $2m$, con una base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$, un bivector $b \in \Lambda^2 V$ puede escribirse en la forma

$$b = \sum_{i<j} b_{ij}\, \boldsymbol{x}_i \wedge \boldsymbol{x}_j = \frac{1}{2} \sum_{i,j=1}^{n} b_{ij}\, \boldsymbol{x}_i \wedge \boldsymbol{x}_j$$

donde la matriz de coeficientes $B := [b_{ij}]$ es antisimétrica. Fíjese que en la segunda sumatoria entran *todos* los pares de índices i, j y se requiere el factor $\frac{1}{2}$ para compensar la duplicación de términos $b_{ij}\boldsymbol{x}_i \wedge \boldsymbol{x}_j = b_{ji}\boldsymbol{x}_j \wedge \boldsymbol{x}_i$ (los términos diagonales en la segunda sumatoria son nulos). La **exponencial** de b es una suma finita:

$$\exp(b) := \sum_{k=0}^{m} \frac{1}{k!} b^{\wedge k} = \sum_{k=0}^{m} \frac{1}{k!} \underbrace{b \wedge b \wedge \cdots \wedge b}_{k \text{ veces}}, \tag{5.10}$$

porque $b^k = \mathbf{0}$ en $\Lambda^{2k} V$ si $k > m$. En efecto, cualquier función $f(t)$ definido por una serie de potencias con coeficientes racionales —aplicable al caso actual de un cuerpo $\mathbb{F}$ cualquiera— conduce, por la sustitución $t \mapsto b$, a un elemento $f(b) \in \Lambda^\bullet V$ definido por una suma finita de elementos en la **subálgebra par** $\underline{\Lambda^+ V} := \bigoplus_{k=0}^{m} \Lambda^{2k} V$.

[9] Esta terminología curiosa se debe a una marcada analogía, enfatizada por Berezin, entre esta forma lineal y una cierta integral (con peso $e^{-t^2/2}$) sobre polinomios. De hecho, hay un isomorfismo evidente entre el álgebra conmutativa $S^\bullet V^*$ y el álgebra de polinomios $\mathbb{F}[t_1, \dots, t_n]$ en n variables. El punto de vista de Berezin es que $\Lambda^\bullet V$ debe considerarse como un "álgebra de polinomios en n variables que anticonmutan". Véase, por ejemplo: Victor Guillemin y Shlomo Sternberg, *Supersymmetry and Equivariant de Rham Theory*, Springer, Berlin, 1999, capítulo 7.

Proposición 5.25. *Si V tiene dimensión par y si $b \in \Lambda^2 V$ es un bivector con matriz antisimétrica B con respecto a una base $\mathcal{B}$ de V, entonces vale*[10]

$$\int_{\mathcal{B}}^{\wedge} \exp(b) = \operatorname{Pf} B.$$

Demostración. Sea $\dim V = 2m$. Entonces

$$\int_{\mathcal{B}}^{\wedge} \exp(b) = \int_{\mathcal{B}}^{\wedge} \sum_{k=0}^{m} \frac{1}{k!} b^{\wedge k} = \frac{1}{m!} \int_{\mathcal{B}}^{\wedge} b^{\wedge m},$$

porque la forma lineal $\int_{\mathcal{B}}^{\wedge}$ se anula sobre los subespacios $\Lambda^{2k} V$ para $k < m$. Ahora, vale

$$\begin{aligned}
b^{\wedge m} &= \Bigl(\frac{1}{2} \sum_{i,j} b_{ij}\, \boldsymbol{x}_i \wedge \boldsymbol{x}_j\Bigr)^{\wedge m} \\
&= \frac{1}{2^m} \sum_{i_1, j_1, \dots, i_m, j_m} b_{i_1, j_1} \dots b_{i_m, j_m}\, \boldsymbol{x}_{i_1} \wedge \boldsymbol{x}_{j_1} \wedge \cdots \wedge \boldsymbol{x}_{i_m} \wedge \boldsymbol{x}_{j_m} \\
&= \frac{1}{2^m} \sum_{\sigma \in S_{2m}} (-1)^{\sigma} b_{\sigma(1),\sigma(2)} \dots b_{\sigma(2m-1),\sigma(2m)}\, \boldsymbol{x}_1 \wedge \boldsymbol{x}_2 \wedge \cdots \wedge \boldsymbol{x}_{2m}.
\end{aligned}$$

Al aplicar la integral de Berezin a $(1/m!)\, b^{\wedge m}$, la fórmula (4.10) muestra que

$$\int_{\mathcal{B}}^{\wedge} \exp(b) = \frac{1}{2^m m!} \sum_{\sigma} (-1)^{\sigma} b_{\sigma(1)\sigma(2)} \dots b_{\sigma(2m-1)\sigma(2m)} = \operatorname{Pf} B. \qquad \square$$

Notación. Si $T : V \to W$ es una aplicación lineal entre dos espacios vectoriales sobre $\mathbb{F}$, se escribe $\Lambda T : \Lambda^{\bullet} V \to \Lambda^{\bullet} W$ para denotar la aplicación lineal (y multiplicativa) determinada por[11]

$$\Lambda T(\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k) := T(\boldsymbol{x}_1) \wedge \cdots \wedge T(\boldsymbol{x}_k).$$

Fíjese que ΛT lleva el subespacio $\Lambda^k V$ en $\Lambda^k W$.

Corolario 5.26. *Si V es un espacio vectorial con base $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ y si $b = \sum_{i<j} b_{ij}\, \boldsymbol{x}_i \wedge \boldsymbol{x}_j$ es un bivector en $\Lambda^2 V$, el desarrollo de* $\exp(b)$ *en la base $\{\boldsymbol{x}_I : I \subseteq \{1, \dots, n\}\}$ de $\Lambda^{\bullet} V$ es*

$$\exp(b) = \sum_{|I| \text{ par}} (\operatorname{Pf} B_{II})\, \boldsymbol{x}_I, \tag{5.11}$$

con el convenio de que $\operatorname{Pf} B_{\emptyset\emptyset} := 1$.

[10] Esta fórmula es también válida si $\dim V$ es impar, de manera trivial, porque ambos lados de la igualdad son nulos.

[11] Si $S : W \to Z$ es otra aplicación lineal, es evidente que $\Lambda(ST) = (\Lambda S)(\Lambda T)$. La notación indica que las correspondencias $V \mapsto \Lambda^{\bullet} V$ y $T \mapsto \Lambda T$ forman un *funtor* de la categoría de espacios vectoriales con aplicaciones lineales, en la categoría de álgebras graduadas con homomorfismos de álgebras.

Demostración. Obsérvese que la matriz B del bivector b es antisimétrica y que cada submatriz principal B_{II} es también antisimétrica. Además, $\operatorname{Pf} B_{II} = 0$ cuando $|I|$ es impar: la sumatoria al lado derecho de (5.11) extiende sobre toda parte $I \subseteq \{1,\dots,n\}$ pero se ha omitido los términos nulos.

Para un determinado conjunto I de índices, con $|I|$ par, sea W el subespacio de V generado por los vectores básicos $\{\boldsymbol{x}_i : i \in I\}$. Defínase la proyección lineal $P\colon V \to W$ por $P(\boldsymbol{x}_i) := \boldsymbol{x}_i$ si $i \in I$, $P(\boldsymbol{x}_j) := \boldsymbol{0}$ si $j \notin I$. Fíjese que $\Lambda P(\boldsymbol{x}_J) = \boldsymbol{x}_J$ si $J \subseteq I$ y que $\Lambda P(\boldsymbol{x}_J) = \boldsymbol{0}$ para otros J.

Ahora, es $\Lambda P(b) \in \Lambda^2 W$ y además

$$\Lambda P(\exp b) = \sum_{0 \le 2k \le n} \frac{1}{k!} (\Lambda P(b))^{\wedge k} = \exp(\Lambda P(b)).$$

La matriz de $\Lambda P(b)$ es la submatriz B_{II} de B. La Proposición 5.25 entonces muestra que $\int^{\wedge}_{P(\mathcal{B})} \exp(\Lambda P(b)) = \operatorname{Pf} B_{II}$. Esta integral de Berezin selecciona el coeficiente del elemento básico $\underline{\mathrm{vol}}_{P(\mathcal{B})} = \boldsymbol{x}_I$ en $\Lambda^\bullet W$, lo cual coincide con el coeficiente de $\boldsymbol{x}_I$ en el desarrollo de $\exp(b)$ en $\Lambda^\bullet V$, debido a que $\Lambda P(\boldsymbol{x}_I) = \boldsymbol{x}_I$. □

5.3 Algebras de Clifford

Definición 5.27. Si V es un espacio vectorial sobre un cuerpo $\mathbb{F}$ cualquiera, para cada vector $\boldsymbol{y} \in V$ se define un operador $\varepsilon(\boldsymbol{y}) \in \operatorname{End}(\Lambda^\bullet V)$, llamado **multiplicación exterior** por el vector $\boldsymbol{y}$, mediante la fórmula

$$\varepsilon(\boldsymbol{y})(\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k) := \boldsymbol{y} \wedge \boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k \quad \text{para todo} \quad \boldsymbol{x}_1,\dots,\boldsymbol{x}_k \in V.$$

Fíjese que $\varepsilon(\boldsymbol{y})^2(\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k) = \boldsymbol{y} \wedge \boldsymbol{y} \wedge \boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k = 0$, luego $\underline{\varepsilon(\boldsymbol{y})^2 = 0}$ en $\operatorname{End}(\Lambda^\bullet V)$. Nótese también que $\varepsilon(\boldsymbol{y})$ es un *operador de grado* $+1$ sobre el álgebra graduada $\Lambda^\bullet V$, porque lleva el subespacio $\Lambda^k V$ en $\Lambda^{k+1} V$.

► En adelante se considera el caso $\mathbb{F} = \mathbb{R}$. Sobre el espacio vectorial real V se elige una forma bilineal simétrica d, no degenerada. Entonces el rango de d es n, pero se admite cualquier signatura $s(d) \in \{-n, -n+2, \dots, n-2, n\}$. Por el Teorema 4.21(b), se sabe que existe una base $\mathcal{E} := \{\boldsymbol{e}_1,\dots,\boldsymbol{e}_p,\boldsymbol{e}_{p+1},\dots,\boldsymbol{e}_{p+q}\}$ donde $p+q = n$ y $p-q = s(d)$, tal que

$$d(\boldsymbol{e}_i,\boldsymbol{e}_j) = \begin{cases} 0, & \text{si } i \neq j, \\ +1. & \text{si } i = j \in \{1,\dots,p\}, \\ -1. & \text{si } i = j \in \{p+1,\dots,p+q\}. \end{cases}$$

Se dice que $\mathcal{E}$ es una **base ortonormal**[12] para (V,d).

[12] Obsérvese que d es un producto escalar si y sólo si d es definida positiva, si y sólo si $s(d) = n$; en cuyo caso $\mathcal{E}$ es una base ortonormal para el espacio euclidiano (V,d).

Definición 5.28. Si (V,d) es un espacio vectorial real con una forma bilineal simétrica no degenerada, para cada vector $\boldsymbol{y} \in V$ se define un operador $\iota(\boldsymbol{y}) \in \operatorname{End}(\Lambda^\bullet V)$, llamado **contracción** con el vector $\boldsymbol{y}$, mediante la fórmula

$$\iota(\boldsymbol{y})(\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k) := \sum_{j=1}^{k} (-1)^{j-1} d(\boldsymbol{y},\boldsymbol{x}_j)\, \boldsymbol{x}_1 \wedge \cdots \wedge \widehat{\boldsymbol{x}_j} \wedge \cdots \wedge \boldsymbol{x}_k,$$

para todo $\boldsymbol{x}_1,\dots,\boldsymbol{x}_k \in V$, donde el circunflejo en $\widehat{\boldsymbol{x}_j}$ significa que *el término* $\boldsymbol{x}_j$ *se omite* del producto exterior $\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k$. Obsérvese que $\iota(\boldsymbol{y})$ es un *operador de grado* -1 sobre el álgebra graduada $\Lambda^\bullet V$, porque lleva el subespacio $\Lambda^k V$ en $\Lambda^{k-1} V$. Sobre los escalares en $\Lambda^0 V$, el operador $\iota(\boldsymbol{y})$ se anula: se define $\iota(\boldsymbol{y})(1) := 0$ por convención.

Lema 5.29. *Si* $\boldsymbol{y} \in V$*, entonces* $\underline{\iota(\boldsymbol{y})^2 = 0}$ *en* $\operatorname{End}(\Lambda^\bullet V)$.

Demostración. Es obvio que $\iota(\boldsymbol{y})^2(1) = 0$ y $\iota(\boldsymbol{y})^2(\boldsymbol{x}) = \iota(\boldsymbol{y})(d(\boldsymbol{y},\boldsymbol{x})) = 0$ para todo $\boldsymbol{x} \in V$. Para $k = 2,\dots,n$, se calcula que

$$\begin{aligned}
\iota(\boldsymbol{y})^2(\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k) &= \sum_{i<j} (-1)^{i-1}(-1)^{j-1} d(\boldsymbol{y},\boldsymbol{x}_i)\, d(\boldsymbol{y},\boldsymbol{x}_j)\, \boldsymbol{x}_1 \wedge \cdots \wedge \widehat{\boldsymbol{x}_i} \wedge \cdots \wedge \widehat{\boldsymbol{x}_j} \wedge \cdots \wedge \boldsymbol{x}_k \\
&\quad + \sum_{i>j} (-1)^{i-2}(-1)^{j-1} d(\boldsymbol{y},\boldsymbol{x}_i)\, d(\boldsymbol{y},\boldsymbol{x}_j)\, \boldsymbol{x}_1 \wedge \cdots \wedge \widehat{\boldsymbol{x}_j} \wedge \cdots \wedge \widehat{\boldsymbol{x}_i} \wedge \cdots \wedge \boldsymbol{x}_k \\
&= \sum_{i<j} \bigl((-1)^{i+j-2} + (-1)^{i+j-3}\bigr) d(\boldsymbol{y},\boldsymbol{x}_i)\, d(\boldsymbol{y},\boldsymbol{x}_j)\, \boldsymbol{x}_1 \wedge \cdots \wedge \widehat{\boldsymbol{x}_i} \wedge \cdots \wedge \widehat{\boldsymbol{x}_j} \wedge \cdots \wedge \boldsymbol{x}_k.
\end{aligned}$$

Se pasa de la segunda a la tercera sumatoria al intercambiar los índices $i \leftrightarrow j$. Luego se obtiene $\iota(\boldsymbol{y})^2(\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k) = 0$ por cancelación de signos. ◻

Lema 5.30. *Si* $\boldsymbol{y}, z \in V$*, entonces* $\underline{\varepsilon(\boldsymbol{y})\iota(z) + \iota(z)\varepsilon(\boldsymbol{y}) = d(\boldsymbol{y},z)\, I}$ *en* $\operatorname{End}(\Lambda^\bullet V)$.

Demostración. La evaluación del operador $\varepsilon(\boldsymbol{y})\iota(z) + \iota(z)\varepsilon(\boldsymbol{y})$ en el escalar $1 \in \mathbb{R} = \Lambda^0 V$ da

$$\varepsilon(\boldsymbol{y})\iota(z)(1) + \iota(z)\varepsilon(\boldsymbol{y})(1) = \varepsilon(\boldsymbol{y})(0) + \iota(z)(\boldsymbol{y}) = d(\boldsymbol{y},z) \in \Lambda^0 V.$$

Para $k = 1,2,\dots,n$, escríbase $\boldsymbol{y} =: \boldsymbol{x}_0$, de modo que

$$\begin{aligned}
(\varepsilon(\boldsymbol{x}_0)\iota(z) + \iota(z)\varepsilon(\boldsymbol{x}_0))(\boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k) &= \sum_{j=1}^{k} (-1)^{j-1} d(\boldsymbol{x}_j,z)\; \boldsymbol{x}_0 \wedge \boldsymbol{x}_1 \wedge \cdots \wedge \widehat{\boldsymbol{x}_j} \wedge \cdots \wedge \boldsymbol{x}_k \\
&\quad + \sum_{j=0}^{k} (-1)^{j} d(\boldsymbol{x}_j,z)\; \boldsymbol{x}_0 \wedge \boldsymbol{x}_1 \wedge \cdots \wedge \widehat{\boldsymbol{x}_j} \wedge \cdots \wedge \boldsymbol{x}_k \\
&= d(\boldsymbol{x}_0,z)\; \boldsymbol{x}_1 \wedge \cdots \wedge \boldsymbol{x}_k,
\end{aligned}$$

porque los términos de las dos sumatorias se anulan por cancelación de signos, con la excepción del primer término de la segunda sumatoria. ◻

Considérese, para $\boldsymbol{y} \in V$, el operador en $\operatorname{End}(\Lambda^\bullet V)$ dado por

$$c(\boldsymbol{y}) := \varepsilon(\boldsymbol{y}) + \iota(\boldsymbol{y}).$$

Los Lemas 5.29 y 5.30 muestran que

$$c(\boldsymbol{y})^2 = \varepsilon(\boldsymbol{y})^2 + \varepsilon(\boldsymbol{y})\iota(\boldsymbol{y}) + \iota(\boldsymbol{y})\varepsilon(\boldsymbol{y}) + \iota(\boldsymbol{y})^2 = d(\boldsymbol{y},\boldsymbol{y})\,I \quad \text{en} \quad \operatorname{End}(\Lambda^\bullet V).$$

La correspondencia $\boldsymbol{y} \mapsto c(\boldsymbol{y}) : V \to \operatorname{End}(\Lambda^\bullet V)$ es lineal. Luego, si $\boldsymbol{y}, \boldsymbol{z} \in V$, entonces

$$\begin{aligned} c(\boldsymbol{y})c(\boldsymbol{z}) + c(\boldsymbol{z})c(\boldsymbol{y}) &= c(\boldsymbol{y}+\boldsymbol{z})^2 - c(\boldsymbol{y})^2 - c(\boldsymbol{z})^2 \\ &= \bigl(d(\boldsymbol{y}+\boldsymbol{z},\boldsymbol{y}+\boldsymbol{z}) - d(\boldsymbol{y},\boldsymbol{y}) - d(\boldsymbol{z},\boldsymbol{z})\bigr) I \\ &= 2\,d(\boldsymbol{y},\boldsymbol{z})\,I. \end{aligned} \tag{5.12a}$$

Si $c(\boldsymbol{y}) = 0$ en $\operatorname{End}(\Lambda^\bullet V)$, entonces $d(\boldsymbol{y},\boldsymbol{z}) = 0$ en $\mathbb{R}$ para todo $\boldsymbol{z} \in V$, luego $\boldsymbol{y} = \boldsymbol{0}$ porque d es no degenerada. Por lo tanto, la correspondencia $\boldsymbol{y} \mapsto c(\boldsymbol{y})$ es *lineal e inyectiva*.

Definición 5.31. Si (V,d) es un espacio vectorial real con una forma bilineal simétrica no degenerada, el **álgebra de Clifford** $\underline{C\ell(V,d)}$ es la subálgebra de $\operatorname{End}(\Lambda^\bullet V)$ generado por los operadores $c(\boldsymbol{y}) := \varepsilon(\boldsymbol{y}) + \iota(\boldsymbol{y})$, para todo $\boldsymbol{y} \in V$.

Conviene usar una *notación simplificada*, escribiendo $\boldsymbol{y}$ por $c(\boldsymbol{y})$ para $\boldsymbol{y} \in V$ y un **producto de Clifford** $\underline{\boldsymbol{yz}} \in C\ell(V,d)$ en vez de la composición de operadores $c(\boldsymbol{y})c(\boldsymbol{z})$. Se identifica el escalar $\lambda \in \mathbb{R}$ con el operador escalar λI. La relación (5.12) se traduce en

$$\underline{\boldsymbol{yz} + \boldsymbol{zy} = 2\,d(\boldsymbol{y},\boldsymbol{z})}, \quad \text{para todo} \quad \boldsymbol{y},\boldsymbol{z} \in V. \tag{5.12b}$$

Proposición 5.32. *Las álgebras $C\ell(V,d)$ y $\Lambda^\bullet V$ son isomorfos como espacios vectoriales reales, pero no como álgebras sobre $\mathbb{R}$.*

Demostración. Si $\mathcal{E} = \{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$ es una base ortonormal para (V,d), sus elementos *anticonmutan* en $C\ell(V,d)$, ya que $\underline{\boldsymbol{e}_i\boldsymbol{e}_j + \boldsymbol{e}_j\boldsymbol{e}_i = 0}$ para $i \neq j$, en vista de (5.12b). Por lo tanto, el álgebra $C\ell(V,d)$ es generada por los *productos ordenados*

$$\{\boldsymbol{e}_{i_1}\boldsymbol{e}_{i_2}\dots\boldsymbol{e}_{i_k} : 1 \leq i_1 < i_2 < \cdots < i_k \leq n\}.$$

Estos elementos están etiquetadas por las partes $I = \{i_1,\dots,i_k\} \subseteq \{1,\dots,n\}$. (El caso $I = \emptyset$ corresponde al escalar $1 \in \mathbb{R}$; fíjese que $\underline{\boldsymbol{e}_i^2 = \pm 1}$ en $C\ell(V,d)$ para $i = 1,\dots,n$.) En consecuencia, es $\dim C\ell(V,d) \leq 2^n = \dim \Lambda^\bullet V$.

Defínase una aplicación lineal $\sigma\colon C\ell(V,d) \to \Lambda^\bullet V$ por

$$\sigma(a) := a(1).$$

Es decir, se evalúa el operador $a \in \operatorname{End}(\Lambda^\bullet V)$ en el elemento escalar $1 \in \Lambda^\bullet V$, de modo que $a(1)$ es un elemento de $\Lambda^\bullet V$. Por ejemplo, si $\boldsymbol{y} \in V$, se obtiene

$$\sigma(1) = I(1) = 1, \qquad \sigma(\boldsymbol{y}) = c(\boldsymbol{y})(1) = \boldsymbol{y} \in \Lambda^1 V,$$

y si $\boldsymbol{y},\boldsymbol{z} \in V$, entonces[13]

$$\sigma(\boldsymbol{y}\boldsymbol{z}) = c(\boldsymbol{y})c(\boldsymbol{z})(1) = \boldsymbol{y}\wedge\boldsymbol{z} + d(\boldsymbol{y},\boldsymbol{z}) \in \Lambda^2 V \oplus \Lambda^0 V.$$

Entonces $\sigma(\boldsymbol{e}_{i_1}\boldsymbol{e}_{i_2}\dots\boldsymbol{e}_{i_k}) = \boldsymbol{e}_{i_1}\wedge\boldsymbol{e}_{i_2}\wedge\cdots\wedge\boldsymbol{e}_{i_k}$ para todo $I \subset \{1,\dots,n\}$. Luego la aplicación σ es sobreyectiva. Si $\sum_I a_I \boldsymbol{e}_{i_1}\dots\boldsymbol{e}_{i_k} = 0$ en $C\ell(V,d)$ para algunos coeficientes $a_I \in \mathbb{R}$, al aplicar σ se obtiene la relación $\sum_I a_I \boldsymbol{e}_{i_1}\wedge\cdots\wedge\boldsymbol{e}_{i_k} = 0$ en $\Lambda^\bullet V$ y se concluye que cada $a_I = 0$. Por lo tanto, los elementos $\boldsymbol{e}_{i_1}\dots\boldsymbol{e}_{i_k}$ son linealmente independientes en $C\ell(V,d)$, luego $\dim C\ell(V,d) = 2^n$ y la aplicación lineal σ es biyectiva.

Por otro lado, si $\boldsymbol{y} \in V$ con $d(\boldsymbol{y},\boldsymbol{y}) \neq 0$, entonces $\boldsymbol{y}^2 = d(\boldsymbol{y},\boldsymbol{y}) \neq 0$ en $C\ell(V,d)$ pero $\sigma(\boldsymbol{y})^{\wedge 2} = \boldsymbol{y}\wedge\boldsymbol{y} = 0$ en $\Lambda^2 V$ y por ende $\sigma(\boldsymbol{y}^2) \neq \sigma(\boldsymbol{y})^{\wedge 2}$. Esto comprueba que σ no es un homomorfismo de álgebras. □

Teorema 5.33 (Chevalley)**.** *Si A es un álgebra sobre $\mathbb{R}$, con elemento identidad 1_A, y si $f\colon V \to A$ es una aplicación $\mathbb{R}$-lineal tal que*

$$f(\boldsymbol{x})^2 = d(\boldsymbol{x},\boldsymbol{x})\,1_A \quad \textit{para todo} \quad \boldsymbol{x} \in V, \tag{5.13}$$

entonces hay un único homomorfismo de álgebras $\tilde{f}\colon C\ell(V,d) \to A$ que extiende f, es decir, tal que $\tilde{f}(\boldsymbol{x}) = f(\boldsymbol{x})$ cuando $\boldsymbol{x} \in V$.

Demostración. La unicidad de la aplicación lineal y multiplicativa $\tilde{f}$ es consecuencia de la fórmula

$$\tilde{f}(\boldsymbol{e}_{i_1}\boldsymbol{e}_{i_2}\dots\boldsymbol{e}_{i_k}) = \tilde{f}(\boldsymbol{e}_{i_1})\tilde{f}(\boldsymbol{e}_{i_2})\dots\tilde{f}(\boldsymbol{e}_{i_k}) = f(\boldsymbol{e}_{i_1})f(\boldsymbol{e}_{i_2})\dots f(\boldsymbol{e}_{i_k}), \tag{5.14}$$

toda vez que $\mathcal{E} = \{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$ es una base ortonormal para (V,d) e $I = \{i_1,\dots,i_k\} \subset \{1,\dots,n\}$.

Esta fórmula también sirve para definir $\tilde{f}$ por linealidad, porque prescribe los valores de $\tilde{f}$ en una base del espacio vectorial $C\ell(V,d)$. Fíjese que si $e_i^2 = \pm 1$, entonces $\tilde{f}(1) = \pm\tilde{f}(e_i^2) = \pm f(e_i)^2 = \pm d(\boldsymbol{e}_i,\boldsymbol{e}_i)\,1_A = 1_A$ por (5.13), lo cual cubre el caso $I = \emptyset$.

Sin embargo, para asegurar que $\tilde{f}$ esté *bien definida*, hay que verificar que la relación (5.12b) en $C\ell(V,d)$ conlleva la relación correspondiente $\tilde{f}(\boldsymbol{y}\boldsymbol{z}+\boldsymbol{z}\boldsymbol{y}) = 2\,d(\boldsymbol{y},\boldsymbol{z})\,1_A$ en A. En efecto, esta condición de buena definición está garantizada por (5.13), ya que

$$\begin{aligned}
\tilde{f}(\boldsymbol{y}\boldsymbol{z}+\boldsymbol{z}\boldsymbol{y}) &= \tilde{f}((\boldsymbol{y}+\boldsymbol{z})^2 - \boldsymbol{y}^2 - \boldsymbol{z}^2) \\
&= (f(\boldsymbol{y}+\boldsymbol{z}))^2 - (f(\boldsymbol{y}))^2 - (f(\boldsymbol{z}))^2 \\
&= d(\boldsymbol{y}+\boldsymbol{z},\boldsymbol{y}+\boldsymbol{z})\,1_A - d(\boldsymbol{y},\boldsymbol{y})\,1_A - d(\boldsymbol{z},\boldsymbol{z})\,1_A \\
&= 2\,d(\boldsymbol{y},\boldsymbol{z})\,1_A
\end{aligned}$$

porque d es bilineal y antisimétrica. □

[13] El elemento $\sigma(a) \in \Lambda^\bullet V$ se llama el **símbolo** de $a \in C\ell(V,d)$. A la inversa, el elemento $Q(\boldsymbol{c}) \in C\ell(V,d)$ cuyo símbolo es $\boldsymbol{c} \in \Lambda^\bullet V$ se llama la **cuantización** de $\boldsymbol{c}$. Para las razones detrás de esta terminología, véase el Capítulo 3 de: Nicole Berline, Ezra Getzler y Michèle Vergne, *Heat Kernels and Dirac Operators*, Springer, Berlin, 1992.

Corolario 5.34. *Sea V un espacio vectorial real con una forma bilineal simétrica no degenerada d con rango $p+q$, signatura $p-q$. Un álgebra real A es isomorfo a* $C\ell(V,d)$ *si y sólo si* $\dim_{\mathbb{R}} A = 2^{\dim V}$ *y hay elementos generadores*[14] $a_1,\dots,a_p,b_1,\dots,b_q \in A$ *tales que para* $i,j \in \{1,\dots,p\}$ *distintos,* $r,s \in \{1,\dots,q\}$ *distintos, valen*

$$a_i^2 = 1_A, \qquad b_r^2 = -1_A, \qquad a_i a_j = -a_j a_i, \qquad b_r b_s = -b_j b_s, \qquad a_i b_r = -b_r a_i. \tag{5.15}$$

Demostración. Con $n = p+q = \dim V$, sea $\mathcal{E} = \{\boldsymbol{e}_1,\dots,\boldsymbol{e}_n\}$ una base ortonormal para (V,d) tal que $d(\boldsymbol{e}_i,\boldsymbol{e}_i) = +1$ para $i = 1,\dots,p$ y $d(\boldsymbol{e}_{p+r},\boldsymbol{e}_{p+r}) = -1$ para $r = 1,\dots,q$. Defínase una aplicación lineal $f\colon V \to A$ por $f(\boldsymbol{e}_i) := a_i$ si $i = 1,\dots,p$ y $f(\boldsymbol{e}_{p+r}) := b_r$ si $r = 1,\dots,q$.

Si $\boldsymbol{x} = \lambda_1\boldsymbol{e}_1 + \cdots + \lambda_n\boldsymbol{e}_n \in V$, las relaciones (5.15) muestran que

$$f(\boldsymbol{x})^2 = \Bigl(\sum_{i=1}^{p} \lambda_i a_i + \sum_{r=1}^{q} \lambda_{p+r} b_r\Bigr)^2 = (\lambda_1^2 + \ldots + \lambda_p^2 - \lambda_{p+1}^2 - \ldots - \lambda_{p+q}^2)\,1_A = d(\boldsymbol{x},\boldsymbol{x})\,1_A.$$

El Teorema 5.33 garantiza que f se extiende, mediante la fórmula (5.14), en un homomorfismo $\tilde{f}\colon C\ell(V,d) \to A$, el cual es sobreyectivo porque los a_i, b_r son generadores de A. La igualdad de dimensiones $\dim_{\mathbb{R}} A = \dim_{\mathbb{R}} C\ell(V,d)$ garantiza que $\tilde{f}$ también es inyectivo. □

El Teorema 5.33 conduce fácilmente a una propiedad estructural importante de $C\ell(V,d)$.

Definición 5.35. Si $C\ell(V,d)$ es un álgebra de Clifford real, sea $C\ell^+(V,d)$ la *subálgebra* generada por productos de dos vectores $\boldsymbol{yz}$, para todo $\boldsymbol{y},\boldsymbol{z} \in V$. Si $\theta\colon V \to C\ell(V,d)$ es la aplicación definido por $\theta(\boldsymbol{x}) := -\boldsymbol{x}$, es evidente que θ cumple (5.13) y por tanto extiende un *automorfismo* $\tilde{\theta}$ de $C\ell(V,d)$. Es evidente que $\tilde{\theta}^2 = I$ en $\operatorname{End}(C\ell(V,d))$, así que los autovalores de $\tilde{\theta}$ son $+1$ y -1. También es claro que la **subálgebra par** $C\ell^+(V,d)$ consiste de autovectores para el autovalor $+1$. Si $C\ell^-(V,d)$ denota el **subespacio impar** de autovectores para el autovalor -1, entonces

$$C\ell(V,d) = C\ell^+(V,d) \oplus C\ell^-(V,d)$$

como espacios vectoriales. Se ve que $\mathbb{R} \subset C\ell^+(V,d)$ y que $V \subset C\ell^-(V,d)$.

► El Corolario 5.34 permite una descripción explícita de las álgebras de Clifford reales de baja dimensión.

Notación. Escríbase $C\ell_{p,q} := C\ell(\mathbb{R}^{p+q}, d_{p,q})$ donde $d_{p,q}$ denota la forma bilineal simétrica sobre $\mathbb{R}^{p+q}$ dada por

$$d_{p,q}(\boldsymbol{x},\boldsymbol{y}) := x_1y_1 + \cdots + x_py_p - x_{p+1}y_{p+1} - \cdots - x_{p+q}y_{p+q}.$$

Para la base ortonormal estándar de R^{p+q} se escribe $\mathcal{E} = \{\boldsymbol{e}_1,\dots,\boldsymbol{e}_p,\boldsymbol{\varepsilon}_1,\dots,\boldsymbol{\varepsilon}_q\}$. Con esta notación, es $\boldsymbol{e}_i^2 = +1$ en $C\ell(V,d)$ para $i = 1,\dots,p$ y $\boldsymbol{\varepsilon}_r^2 = -1$ en $C\ell(V,d)$ para $r = 1,\dots,q$.

Como caso trivial, se designa $C\ell_{0,0} := \mathbb{R}$, de dimensión $2^0 = 1$.

[14] Unos elementos $c_1,\dots,c_k$ son **generadores** de un álgebra A si todos los productos finitos $c_{i_1}\dots c_{i_l}$ generan A como espacio vectorial.

Ejemplo 5.36. Es claro que $C\ell_{1,0} \simeq \mathbb{R}\oplus\mathbb{R}$ como espacio vectorial; la primera copia de $\mathbb{R}$ denota los escalares, la segunda copia son los múltiplos de $\boldsymbol{e}_1$. Si $a,b,c,d \in \mathbb{R}$, entonces $(a+b\boldsymbol{e}_1)(c+d\boldsymbol{e}_1) = (ac+bd)+(ad+bc)\boldsymbol{e}_1$. Se puede considerar $\mathbb{R}\oplus\mathbb{R}$ como subálgebra de $M_2(\mathbb{R})$, al identificar

$$1 \leftrightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{e}_1 \leftrightarrow \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad a+b\boldsymbol{e}_1 \leftrightarrow \begin{bmatrix} a & b \\ b & a \end{bmatrix}.$$

Ejemplo 5.37. Por otro lado, es $C\ell_{0,1} \simeq \mathbb{C}$. En efecto, la regla de multiplicación

$$(a+b\boldsymbol{\varepsilon}_1)(c+d\boldsymbol{\varepsilon}_1) = (ac-bd)+(ad+bc)\boldsymbol{\varepsilon}_1, \quad \text{para} \quad a,b,c,d \in \mathbb{R},$$

revela el isomorfismo, con $\boldsymbol{\varepsilon}_1 \leftrightarrow i \in \mathbb{C}$. (Recuérdese que $\dim_{\mathbb{R}} \mathbb{C} = 2$.) Además, es posible considerar $\mathbb{C}$ como subálgebra de $M_2(\mathbb{R})$, al identificar

$$1 \leftrightarrow \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \qquad \boldsymbol{\varepsilon}_1 \leftrightarrow \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad a+b\boldsymbol{\varepsilon}_1 \leftrightarrow \begin{bmatrix} a & b \\ -b & a \end{bmatrix}.$$

Definición 5.38. Las **matrices de Pauli**[15] son las siguientes tres matrices en $M_2(\mathbb{C})$:

$$\sigma_1 := \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \sigma_2 := \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}, \qquad \sigma_3 := \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}. \tag{5.16}$$

Obsérvese que $\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = I_2$ y que estas matrices *anticonmutan*; en efecto,

$$\sigma_1\sigma_2 = i\sigma_3 = -\sigma_2\sigma_1, \qquad \sigma_3\sigma_1 = i\sigma_2 = -\sigma_1\sigma_3, \qquad \sigma_2\sigma_3 = i\sigma_1 = -\sigma_3\sigma_2.$$

Ejemplo 5.39. Resulta que $C\ell_{2,0} \simeq M_2(\mathbb{R})$. En primer lugar, es $C\ell_{2,0} = \operatorname{lin}\langle 1,\boldsymbol{e}_1,\boldsymbol{e}_2,\boldsymbol{e}_1\boldsymbol{e}_2\rangle$. Obsérvese que

$$(\boldsymbol{e}_1\boldsymbol{e}_2)^2 = \boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_1\boldsymbol{e}_2 = -\boldsymbol{e}_1\boldsymbol{e}_1\boldsymbol{e}_2\boldsymbol{e}_2 = -(+1)(+1) = -1.$$

Por tanto, como σ_3 y σ_1 anticonmutan, se puede identificar

$$\boldsymbol{e}_1 \leftrightarrow \sigma_3 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad \boldsymbol{e}_2 \leftrightarrow \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \boldsymbol{e}_1\boldsymbol{e}_2 \leftrightarrow J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$

Junto con $1 \leftrightarrow I_2$, estas matrices generan $M_2(\mathbb{R})$ como espacio vectorial real; es evidente que el producto de elementos de $C\ell_{2,0}$ corresponde con el producto de las matrices 2×2.

Ejemplo 5.40. También sucede que $C\ell_{1,1} \simeq M_2(\mathbb{R})$, con otra identificación. Ahora $C\ell_{2,0} = \operatorname{lin}\langle 1,\boldsymbol{e}_1,\boldsymbol{\varepsilon}_1,\boldsymbol{e}_1\boldsymbol{\varepsilon}_1\rangle$, con $(\boldsymbol{e}_1\boldsymbol{\varepsilon}_1)^2 = -\boldsymbol{e}_1\boldsymbol{e}_1\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1 = -(+1)(-1) = +1$. Como σ_1 y J_2 anticonmutan, se identifica

$$\boldsymbol{e}_1 \leftrightarrow \sigma_1 = \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \boldsymbol{\varepsilon}_1 \leftrightarrow J_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}, \qquad \boldsymbol{e}_1\boldsymbol{\varepsilon}_1 \leftrightarrow -\sigma_3 = \begin{bmatrix} -1 & 0 \\ 0 & 1 \end{bmatrix}.$$

[15]Estos matrices fueron usadas por el físico austriaco *Wolfgang Pauli* en 1927 para modelar el fenómeno de *espín* de un electrón. Su *principio de exclusión*, que afirma que dos partículas subatómicas de espín $\frac{1}{2}$ no pueden coexistir en un mismo estado material, es la base de la explicación moderna de la estructura de los átomos.

Definición 5.41. Los **cuaterniones** son elementos del álgebra real $\mathbb{H} := \operatorname{lin}\langle 1, i, j, k\rangle$, con $\dim_{\mathbb{R}} \mathbb{H} = 4$, cuyos generadores i, j, k obedecen las *relaciones de Hamilton*:[16]

$$i^2 = j^2 = k^2 = ijk = -1.$$

En consecuencia, estos generadores anticonmutan: $ij = k = -ji$, $ki = j = -ik$, $jk = i = -kj$.

Ejemplo 5.42. Resulta que $\mathrm{C}\ell_{0,2} \simeq \mathbb{H}$. En efecto, es cuestión de identificar $\boldsymbol{\varepsilon}_1 \leftrightarrow i$, $\boldsymbol{\varepsilon}_2 \leftrightarrow j$, $\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2 \leftrightarrow k$. Nótese que $(\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2)^2 = -\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_1\boldsymbol{\varepsilon}_2\boldsymbol{\varepsilon}_2 = -(-1)(-1) = -1$.

Ejemplo 5.43. Se obtiene $\mathrm{C}\ell_{3,0} \simeq M_2(\mathbb{C})$ al identificar los generadores $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3$ con las matrices de Pauli $\sigma_1, \sigma_2, \sigma_3$ respectivamente. Obsérvese que $\dim_{\mathbb{R}} M_2(\mathbb{C}) = 2 \dim_{\mathbb{C}} M_2(\mathbb{C}) = 8$. Una base para $M_2(\mathbb{C})$ sobre $\mathbb{R}$ es $\{I_2, \sigma_1, \sigma_2, \sigma_3, i\sigma_3, i\sigma_1, i\sigma_2, iI_2\}$.

Ejemplo 5.44. Sucede que $\mathrm{C}\ell_{3,0} \simeq M_2(\mathbb{H})$, las matrices 2×2 con entradas en $\mathbb{H}$. Si se reemplaza la segunda matriz de Pauli por

$$\tau_2 := \begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix},$$

se obtiene un elemento de $M_2(\mathbb{H})$ con $\tau_2^2 = I_2$, que anticonmuta con σ_1, σ_2 y σ_3. En efecto:

$$\sigma_2\tau_2 = \begin{bmatrix} 0 & -i \\ i & 0 \end{bmatrix}\begin{bmatrix} 0 & -j \\ j & 0 \end{bmatrix} = \begin{bmatrix} -k & 0 \\ 0 & -k \end{bmatrix} = -\tau_2\sigma_2.$$

Nótese también que $\dim_{\mathbb{R}} M_2(\mathbb{H}) = 4 \dim_{\mathbb{R}} \mathbb{H} = 16$. El Corolario 5.34 muestra el isomorfismo requerido, al identificar $\boldsymbol{e}_1, \boldsymbol{e}_2, \boldsymbol{e}_3, \boldsymbol{e}_4$ con $\sigma_1, \sigma_2, \sigma_3, \tau_2$ respectivamente.

Lema 5.45. *Hay isomorfismos $\mathrm{C}\ell_{p+1,q+1} \simeq M_2(\mathrm{C}\ell_{p,q})$, para todo $p, q \in \mathbb{N}$.*

Demostración. Por el Corolario 5.34, basta identificar la base $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_{p+1}, \boldsymbol{\varepsilon}_1, \dots, \boldsymbol{\varepsilon}_{q+1}\}$ de $\mathbb{R}^{p+q+2}$ con elementos de $M_2(\mathrm{C}\ell_{p,q})$ con cuadrados respectivos ± 1 que anticonmutan entre sí. Fíjese que $\dim \mathrm{C}\ell_{p+1,q+1} = 2^{p+q+2} = 4 \cdot 2^{p+q} = \dim M_2(\mathrm{C}\ell_{p,q})$. Las identificaciones requeridas son

$$\boldsymbol{e}_i \longmapsto \begin{bmatrix} \boldsymbol{e}_i & 0 \\ 0 & -\boldsymbol{e}_i \end{bmatrix} \quad \text{para} \quad i = 1, \dots, p, \qquad \boldsymbol{\varepsilon}_r \longmapsto \begin{bmatrix} \boldsymbol{\varepsilon}_r & 0 \\ 0 & -\boldsymbol{\varepsilon}_r \end{bmatrix}, \quad \text{para} \quad r = 1, \dots, q,$$

y también

$$\boldsymbol{e}_{p+1} \longmapsto \begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix}, \qquad \boldsymbol{\varepsilon}_{q+1} \longmapsto \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}.$$ □

[16] *William Rowan Hamilton* introdujo un cálculo de vectores en $\mathbb{R}^3$ en donde cada vector de la base estándar $\{i, j, k\}$ de $\mathbb{R}^3$ se combina con los escalares en $\mathbb{R}$ para formar una copia de los números complejos, al demandar que $i^2 = j^2 = k^2 = -1$. La dificultad esencial fue descubrir la manera de multiplicar vectores no paralelos; esto lo resolvió en 1847 mediante la receta $ijk = -1$. Esta solución se le ocurrió mientras viajaba por taxi en el puente Brougham sobre el río Tolka en Dublin; detuvo los caballos y saltó del taxi para cavar con un cuchillo en la madera del puente la fórmula $ijk = -1$. Desafortunadamente, la inscripción no sobrevivió el paso del tiempo.

Este resultado de “periodicidad-$(1,1)$” muestra que es suficiente clasificar las álgebras de Clifford $C\ell_{p,0}$ y $C\ell_{0,q}$ para $p,q \in \mathbb{N}$. De hecho, hay otro resultado importante, la *periodicidad módulo* 8, que afirma que[17]

$$C\ell_{p+8,q} \simeq C\ell_{p,q+8} \simeq M_{16}(C\ell_{p,q}).$$

De este modo, basta clasificar $C\ell_{p,0}$ y $C\ell_{0,q}$ para $p,q = 0,1,\dots,7$. Por ejemplo, vale $C\ell_{8,0} \simeq C\ell_{0,8} \simeq M_{16}(C\ell_{0,0}) = M_{16}(\mathbb{R})$. La lista de estos casos, completando los ejemplos anteriores, es[18]

$C\ell_{p,0}$ para $p = 0,1,\dots,7$:

$\mathbb{R},\quad \mathbb{R}\oplus\mathbb{R},\quad M_2(\mathbb{R}),\quad M_2(\mathbb{C}),\quad M_2(\mathbb{H}),\quad M_2(\mathbb{H})\oplus M_2(\mathbb{H}),\quad M_4(\mathbb{H}),\quad M_8(\mathbb{C});$

$C\ell_{0,q}$ para $q = 0,1,\dots,7$:

$\mathbb{R},\quad \mathbb{C},\quad \mathbb{H},\quad \mathbb{H}\oplus\mathbb{H},\quad M_2(\mathbb{H}),\quad M_4(\mathbb{C}),\quad M_8(\mathbb{R}),\quad M_8(\mathbb{R})\oplus M_8(\mathbb{R}).$

De hecho, estas listas ejemplifican un teorema de Wedderburn, que dice que cualquier álgebra real asociativa semisimple[19] es una suma directa de álgebras de matrices sobre una de las tres álgebras $\mathbb{R}$, $\mathbb{C}$, $\mathbb{H}$.

5.4 Ejercicios sobre álgebras exteriores y de Clifford

Ejercicio 5.1. (a) Demostrar que cualquier elemento de $\Lambda^2 V^*$ es de la forma

$$g = h_1 \wedge h_2 + h_3 \wedge h_4 + \cdots + h_{2r-1} \wedge h_{2r},$$

donde $h_1,\dots,h_{2r}$ son elementos linealmente independientes de V^*.
⟦ Indicación: Si $\{f_1,\dots,f_n\}$ es una base de V^*, entonces $g = \frac{1}{2}\sum_{i,j=1}^n a_{ij} f_i \wedge f_j$ par una matriz antisimétrica $A = [a_{ij}] \in M_n(\mathbb{F})$. Expresar $A = R^t(J_{2r} \oplus O_{n-2r})R$ y tomar $h_k := \sum_{j=1}^n r_{kj} f_j$. ⟧

(b) Verificar que la “potencia exterior” $\underline{g^{\wedge r}} := g \wedge g \wedge \cdots \wedge g$ (r veces) en $\Lambda^{2r}(V^*)$ cumple

$$g^{\wedge r} = (r!)\, h_1 \wedge h_2 \wedge \cdots \wedge h_{2r}.$$

[17]En el caso complejo (se puede desarrollar la teoría de las álgebras de Clifford sobre $\mathbb{C}$ en vez de $\mathbb{R}$) hay un resultado más sencillo: si $\mathbb{C}\ell_n$ es el álgebra de Clifford complejo generado por el espacio hilbertiano $\mathbb{C}^n$, entonces $\mathbb{C}\ell_{n+2} \simeq M_2(\mathbb{C}\ell_n)$. Esto fue descubierto por *Raoul Bott* en 1958, como consecuencia de su clasificación de los grupos de homotopía de los grupos unitarios. La esencia algebraica de esta construcción fue extraída, junto con la periodicidad módulo 8 para el caso real, en: Michael Atiyah, Raoul Bott y Arnold Shapiro, “Clifford modules”, Topology **3** (1964), 3–38.

[18]Aquí no se demuestran los casos superiores, ni el teorema de periodicidad módulo 8. Para estos y otros detalles sobre las álgebras de Clifford, se remite al Capítulo 5 del libro: José M. Gracia-Bondía, Joseph C. Várilly y Héctor Figueroa, *Elements of Noncommutative Geometry*, Birkhäuser, Boston, 2001.

[19]Un álgebra se llama *semisimple* si no posee ideales nilpotentes. Para el teorema de Wedderburn, véase el último capítulo de: Isadore Herstein, *Topics in Algebra*, Blaisdell, New York, 1964.

Ejercicio 5.2. Sea V un espacio vectorial de dimensión $n = 2m$ y sea $g \in \Lambda^2 V^*$ una forma bilineal alternada *no degenerada*. Con la notación del Ejercicio 5.1, demostrar que

$$h_1 \wedge h_2 \wedge \cdots \wedge h_{2m} = (\det R)\, f_1 \wedge f_2 \wedge \cdots \wedge f_{2m}.$$

A partir de la definición $\underline{\operatorname{Pf} A := \det R}$, usar el resultado del Ejercicio anterior para verificar la fórmula (4.10) para $\operatorname{Pf} A$.

Ejercicio 5.3. Sea V un espacio vectorial sobre $\mathbb{R}$ con una base ordenada $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_m\}$. Se dice que otra base $\mathcal{B}' = \{\boldsymbol{x}'_1, \dots, \boldsymbol{x}'_m\}$ **tiene la misma orientación**[20] que $\mathcal{B}$ si $\underline{\det [I]^{\mathcal{B}}_{\mathcal{B}'} > 0}$.

(a) Verificar que $\mathcal{B}$ y $\mathcal{B}'$ tienen la misma orientación si y sólo si $\underline{\mathrm{vol}}_{\mathcal{B}'} = c\, \underline{\mathrm{vol}}_{\mathcal{B}}$ en $\Lambda^n V$, con $c > 0$.

(b) Si $b = \frac{1}{2} \sum_{i,j=1}^n b_{ij} \boldsymbol{x}_i \wedge \boldsymbol{x}_j = \frac{1}{2} \sum_{r,s=1}^n b'_{rs} \boldsymbol{x}'_r \wedge \boldsymbol{x}'_s \in \Lambda^2 V$, mostrar que $\operatorname{Pf} B$ y $\operatorname{Pf} B'$ tienen el mismo signo si y sólo si $\mathcal{B}$ y $\mathcal{B}'$ tienen la misma orientación.

Ejercicio 5.4. Dada una forma bilineal simétrica no degenerada d sobre un espacio vectorial real V, sea $Q\colon \Lambda^\bullet V \to \mathrm{C}\ell(V,d)$ el inverso de la aplicación lineal biyectiva $\sigma\colon \mathrm{C}\ell(V,d) \to \Lambda^\bullet V$ dada por $\sigma(a) := a(1)$.

(a) Si $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z} \in V$, verificar que

$$Q(\boldsymbol{x} \wedge \boldsymbol{y} \wedge \boldsymbol{z}) = \boldsymbol{x}\boldsymbol{y}\boldsymbol{z} - d(\boldsymbol{y},\boldsymbol{z})\,\boldsymbol{x} + d(\boldsymbol{x},\boldsymbol{z})\,\boldsymbol{y} - d(\boldsymbol{x},\boldsymbol{y})\,\boldsymbol{z}.$$

(b) Si $\mathcal{B} = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\}$ es una base ortonormal de (V,d), comprobar que

$$Q(\underline{\mathrm{vol}}_{\mathcal{B}}) = \frac{1}{n!} \sum_{\sigma \in S_n} (-1)^\sigma \boldsymbol{x}_{\sigma(1)} \boldsymbol{x}_{\sigma(2)} \dots \boldsymbol{x}_{\sigma(n)}.$$

Ejercicio 5.5. Si $\mathrm{C}\ell(V,d)$ es un álgebra de Clifford real, es

$$\mathrm{C}\ell(V,d) = \mathrm{C}\ell^+(V,d) \oplus \mathrm{C}\ell^-(V,d),$$

donde la **subálgebra par** $\mathrm{C}\ell^+(V,d)$ y el **subespacio impar** $\mathrm{C}\ell^-(V,d)$ son generados por productos de un número par [respectivamente, por un número impar] de vectores en V.

(a) Mostrar que $\dim \mathrm{C}\ell^+(V,d) = \frac{1}{2} \dim \mathrm{C}\ell(V,d)$.

[[Indicación: Si $\boldsymbol{x} \in V$ es un vector no nulo, mostrar que $a \mapsto \boldsymbol{x}a$ es una biyección lineal entre $\mathrm{C}\ell^+(V,d)$ y $\mathrm{C}\ell^-(V,d)$.]]

(b) Se sabe que hay isomorfismos de álgebras reales tales que

$$\mathrm{C}\ell_{1,0} \simeq \mathbb{R} \oplus \mathbb{R}, \qquad \mathrm{C}\ell_{2,0} \simeq M_2(\mathbb{R}), \qquad \mathrm{C}\ell_{3,0} \simeq M_2(\mathbb{C}), \qquad \mathrm{C}\ell_{4,0} \simeq M_2(\mathbb{H}),$$

donde $\mathbb{H} = \{\, c_0 + c_1 i + c_2 j + c_3 k : c_i \in \mathbb{R} \,\}$ es el álgebra de *cuaterniones*. Verificar que

$$\mathrm{C}\ell^+_{1,0} \simeq \mathbb{R}, \qquad \mathrm{C}\ell^+_{2,0} \simeq \mathbb{C}, \qquad \mathrm{C}\ell^+_{3,0} \simeq \mathbb{H}, \qquad \mathrm{C}\ell^+_{4,0} \simeq \mathbb{H} \oplus \mathbb{H}.$$

[[Indicación: Si $\alpha, \beta \in \mathbb{C}$, se identifica $\alpha + \beta j \in \mathbb{H}$ con la matriz $\begin{bmatrix} \alpha & \beta \\ -\bar\beta & \bar\alpha \end{bmatrix} \in M_2(\mathbb{C})$.]]

[20] Una **orientación** sobre V es una de las dos clases de equivalencia de bases ordenadas determinadas por el signo de $\det [I]^{\mathcal{B}}_{\mathcal{B}'}$.

Bibliografía

Los siguientes libros amplifican y profundizan los tópicos vistos en este curso.

1. Mischa Cotlar y Cora Ratto de Sadosky, *Introducción al Algebra: Nociones de Algebra Lineal*, Editorial Universitaria de Buenos Aires, 1963.

2. Feliks R. Gantmacher, *The Theory of Matrices*, tomo 1, Chelsea, New York, 1959.

3. Lidia I. Goloviná, *Algebra Lineal y Algunas de sus Aplicaciones*, Mir, Moscú, 1974.

4. Paul R. Halmos, *Espacios Vectoriales de Dimensión Finita*, Compañía Editorial Continental, México, DF, 1971.

5. Kenneth Hoffman y Ray Kunze, *Algebra Lineal*, Prentice-Hall Internacional, Madrid, 1972.

6. Serge Lang, *Algebra Lineal*, Fondo Educativo Interamericano, México, DF, 1976.

7. Saunders MacLane y Garrett Birkhoff, *Algebra*, Macmillan, New York, 1967.

8. Anatoly I. Maltsev, *Fundamentos de Algebra Lineal*, Mir, Moscú, 1972.

9. Ben Noble y James W. Daniel, *Algebra Lineal Aplicada*, Prentice-Hall Hispanoamericana, México, DF, 1989.

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

11. Orlando E. Villamayor, *Algebra Lineal*, OEA, Washington, DC, 1981.

12. Valentin V. Voevodin, *Algebra Lineal*, Mir, Moscú, 1982.