Notas para el curso de Álgebra Lineal II

Centro de Matemática
Facultad de Ciencias
Universidad de la República

Andrés Abella

22 de febrero de 2023

Introducción

Estas son notas para el curso de Álgebra Lineal II de la Licenciatura en Matemática de la Facultad de Ciencias. Tratan sobre operadores (es decir, transformaciones lineales de un espacio vectorial en sí mismo), productos internos y formas bilineales simétricas, con énfasis en los espacios de dimensión finita. Como prerrequisito se asume que el lector tiene conocimientos básicos de espacios vectoriales, transformaciones lineales y matrices asociadas a transformaciones lineales. Si bien la mayoría de los resultados están escritos para $\mathbb{k}$-espacios en los cuales $\mathbb{k}$ es un cuerpo arbitrario, dado el caracter introductorio de las notas, todos los ejemplos son para $\mathbb{k} = \mathbb{R}$ (los números reales) o $\mathbb{k} = \mathbb{C}$ (los números complejos). A continuación damos una breve descripción de cada capítulo.

En el Capítulo 1 se estudian los operadores diagonalizables y sus propiedades básicas.

En el Capítulo 2 se estudian los espacios vectoriales con producto interno. En particular se prueba la existencia de bases ortonormales, de complementos ortogonales y el teorema de Riesz.

En el Capítulo 3 se estudian los operadores en espacios con producto interno. Las demostraciones están escritas para espacios complejos, mientras que las correspondientes a los espacios reales se obtienen como un caso particular de las anteriores. Si el lector está interesado solo en los espacios reales, entonces puede olvidarse de las referencias a los complejos, y las definiciones y demostraciones funcionan esencialmente igual. El único punto donde tendría problemas es en la prueba de que un operador real autoadjunto es diagonalizable sobre una base ortonormal, dado que la prueba en el caso complejo utiliza el *teorema fundamental del álgebra* que no es válido en los reales. Para solucionarlo se dan dos pruebas de ese teorema, una de las cuales no requiere números complejos.

En el Capítulo 4 se estudian las formas bilineales simétricas reales y sus formas cuadráticas asociadas.

En el Capítulo 5 se ve el teorema de Cayley-Hamilton y se introducen los subespacios propios generalizados.

En el Capítulo 6 se estudia la forma de Jordan. En la primera sección se estudian las propiedades de las bases de Jordan y se prueba su existencia. En la segunda se ven técnicas de cálculo que permiten obtener la forma de Jordan en dimensiones bajas.

En el Capítulo 7 se introduce el polinomio minimal. El mismo es útil para obtener información sobre un operador, a apartir de las relaciones polinomiales que el operador verifique.

El Capítulo 8 es un apéndice en el cual se ven algunos temas que son necesarios para seguir estas notas. No se necesita saber todo lo que aparece aquí para seguir todos los capítulos, pero la suma directa de subespacios se utiliza con frecuencia, así que conviene darle una mirada antes de empezar con el resto. Las matrices en bloques aparecen al estudiar matrices asociadas a operadores, en particular en diagonalización y principalmente en la forma de Jordan. Las matrices elementales aparecen en la diagonalización de formas bilineales simétricas.

Parte de este material está inspirado en el libro *Linear Algebra*, de S. M. Friedberg, A. J. Insel y L. E. Spence, de la editorial Prentice Hall.

Contenidos

Capítulo 1

Diagonalización

En este capítulo $\mathbb{k}$ es un cuerpo y V es un $\mathbb{k}$-espacio vectorial no nulo de dimensión finita n. Si $\mathcal{B}$ es una base de V, supondremos siempre que $\mathcal{B}$ es una *base ordenada* es decir una base en la cual hemos elegido un orden para sus elementos.

A las transformaciones lineales de V en V les llamaremos también *operadores*. Al espacio de los operadores en V lo denotaremos $\mathcal{L}(V)$ y a la transformación lineal identidad Id_V o simplemente Id. Al espacio de las matrices cuadradas de orden n con coeficientes en $\mathbb{k}$ lo denotaremos $M_n(\mathbb{k})$ y a la matriz identidad I_n o I.

Si $T \in \mathcal{L}(V)$ y $\mathcal{B}$ y $\mathcal{C}$ son bases de V, escribiremos ${}_{\mathcal{C}}[T]_{\mathcal{B}}$ a la matriz asociada a T de la base $\mathcal{B}$ en la base $\mathcal{C}$ y $[T]_{\mathcal{B}}$ para ${}_{\mathcal{B}}[T]_{\mathcal{B}}$. Si $A \in M_n(\mathbb{k})$, escribiremos L_A a la transformación lineal de $\mathbb{k}^n$ en $\mathbb{k}^n$ definida por $L_A(v) = Av$, para todo $v \in \mathbb{k}^n$ (el vector v en Av está escrito en forma vertical). Observar que si $\mathcal{C} = \{e_1, \dots, e_n\}$ es la base canónica de $\mathbb{k}^n$ y $A \in M_n(\mathbb{k})$, entonces es $[L_A]_{\mathcal{C}} = A$. Si $v_1, \dots, v_n \in \mathbb{k}^n$, el símbolo $[v_1|\cdots|v_n]$ denotará a la matriz cuadrada de orden n cuyas columnas son los vectores $v_1, \dots, v_n$ escritos en forma vertical. Por ejemplo, si $v_1 = (1, 2)$ y $v_2 = (0, 3)$, entonces $[v_1|v_2] = \left(\begin{smallmatrix} 1 & 0 \\ 2 & 3 \end{smallmatrix}\right)$.

1.1. Operadores diagonalizables

Recordar que una matriz cuadrada $A = (a_{ij})$ se dice *diagonal* si $a_{ij} = 0$, para todo $i \neq j$. Las matrices diagonales son simples para operar, por ejemplo valen las fórmulas siguientes.

$$\begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix} \times \begin{pmatrix} b_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & b_n \end{pmatrix} = \begin{pmatrix} a_1 b_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n b_n \end{pmatrix}; \quad \begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix}^k = \begin{pmatrix} a_1^k & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n^k \end{pmatrix}, \ \forall k \in \mathbb{N};$$

$$\begin{vmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{vmatrix} = a_1 \cdots a_n; \quad \text{si } a_1 \cdots a_n \neq 0 \ \Rightarrow \ \begin{pmatrix} a_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n \end{pmatrix}^{-1} = \begin{pmatrix} a_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & a_n^{-1} \end{pmatrix}.$$

Definición 1.1.1. Un operador $T \in \mathcal{L}(V)$ es *diagonalizable* si existe una base $\mathcal{B}$ de V en la cual $[T]_{\mathcal{B}}$ es diagonal.

Ejemplo 1.1.2. Sea r una recta del plano $\mathbb{R}^2$ que pasa por el origen y T la simetría axial respecto a r. Sean $0 \neq u \in r$ y $0 \neq v \in \mathbb{R}^2$ tales que u y v son ortogonales. El conjunto $\mathcal{B} = \{u, v\}$ es una base de $\mathbb{R}^2$ y la matriz asociada a T en la base $\mathcal{B}$ es $\left(\begin{smallmatrix} 1 & 0 \\ 0 & -1 \end{smallmatrix}\right)$. Luego la simetría axial de eje r es diagonalizable.

Definición 1.1.3. Dado $T \in \mathcal{L}(V)$, un escalar $\lambda \in \mathbb{k}$ se dice un *valor propio* de T si existe $0 \neq v \in V$ tal que $T(v) = \lambda v$, el vector v se dice que es un *vector propio* de T asociado al valor propio λ.

Ejemplo 1.1.4. En el caso en que $T : \mathbb{R}^2 \to \mathbb{R}^2$ es la simetría axial del ejemplo 1.1.2, vimos que existían vectores no nulos u y v tales que $T(u) = u$ y $T(v) = -v$; luego u y v son vectores propios de T con valores propios 1 y -1, respectivamente.

Proposición 1.1.5. *Un operador $T \in \mathcal{L}(V)$ es diagonalizable si y solo si existe una base $\mathcal{B}$ de V formada por vectores propios de T. En ese caso, las entradas diagonales de $[T]_{\mathcal{B}}$ son valores propios de T.*

Dem. El operador $T \in \mathcal{L}(V)$ es diagonalizable si y solo si existe una base $\mathcal{B}$ tal que $[T]_{\mathcal{B}}$ es de la forma

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix}. \tag{1.1}$$

Si $\mathcal{B} = \{v_1, \ldots, v_n\}$, entonces vale (1.1) si y solo si $T(v_i) = \lambda_i v_i$, para todo $i = 1, \ldots, n$, es decir, si y solo si los elementos de $\mathcal{B}$ son vectores propios de T con valores propios $\lambda_1, \ldots, \lambda_n$. $\square$

Ejemplo 1.1.6. Sea $T \in \mathcal{L}(\mathbb{R}_n[x])$, $n \geq 1$, definida por $T(p(x)) = p'(x)$. Un polinomio $p(x)$ es un vector propio de T si y solo si $p(x)$ es no nulo y existe $\lambda \in \mathbb{R}$ tal que $p'(x) = \lambda p(x)$. Esta última igualdad solo puede ocurrir cuando el grado de $p(x)$ es cero, es decir, cuando $p(x)$ es un polinomio constante; pero en ese caso es $p'(x) = 0$ y por lo tanto tiene que ser $\lambda = 0$. En conclusión el único valor propio de T es $\lambda = 0$ y los vectores propios correspondientes son los polinomios constantes no nulos. Pero el conjunto formado por dos polinomios constantes es LD, luego no puede existir una base de $\mathbb{R}_n[x]$ formada por vectores propios de T y por lo tanto T no es diagonalizable.

1.2. Polinomio característico

Por la proposición anterior, para saber si un operador T es diagonalizable hay que ver si podemos encontrar una base del espacio formada por vectores propios de T. Para eso debemos saber cómo encontrar los vectores propios. A continuación veremos primero cómo hallar los valores propios y luego usando estos hallaremos los vectores propios.

Definición 1.2.1. Decimos que dos matrices $A, B \in M_n(\Bbbk)$ son *semejantes* y escribimos $A \simeq B$, si existe una matriz invertible $Q \in M_n(\Bbbk)$ tal que $A = QBQ^{-1}$.

Proposición 1.2.2. *La relación de semejanza es una relación de equivalencia en $M_n(\Bbbk)$, es decir que valen*

$$A \simeq A,\ \forall A \in M_n(\Bbbk); \quad \textit{si } A \simeq B \ \Rightarrow\ B \simeq A; \quad \textit{si } A \simeq B \textit{ y } B \simeq C \ \Rightarrow\ A \simeq C.$$

Dem. Ejercicio. $\square$

Proposición 1.2.3. *Si $A, B \in M_n(\Bbbk)$ son semejantes, entonces* $\det(A) = \det(B)$.

Dem. Sea $Q \in M_n(\Bbbk)$ invertible tal que $A = QBQ^{-1}$. Entonces

$$\det(A) = \det\left(QBQ^{-1}\right) = \det(Q)\det(B)\det\left(Q^{-1}\right) = \det(Q)\det(B)\det(Q)^{-1} = \det(B). \quad \square$$

El siguiente resultado muestra que las matrices asociadas a una misma trasformación lineal son semejantes.

Proposición 1.2.4. *Sean $T \in \mathcal{L}(V)$ y $\mathcal{B}$, $\mathcal{C}$ bases de V. Entonces $[T]_{\mathcal{C}} = P\,[T]_{\mathcal{B}}\,P^{-1}$, siendo $P = {}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}}$.*

Dem. Si $P = {}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}}$, entonces

$$P\,[T]_{\mathcal{C}}\,P^{-1} = {}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}}\,[T]_{\mathcal{B}}\,({}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}})^{-1} = {}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}}\,[T]_{\mathcal{B}}\,{}_{\mathcal{B}}[\mathrm{Id}]_{\mathcal{C}} = [\mathrm{Id} \circ T \circ \mathrm{Id}]_{\mathcal{C}} = [T]_{\mathcal{C}}. \quad \square$$

De las dos proposiciones anteriores se deduce el siguiente resultado.

Corolario 1.2.5. *Si* $T \in \mathcal{L}(V)$ *y* $\mathcal{B}, \mathcal{C}$ *son dos bases de* V*, entonces* $\det([T]_{\mathcal{B}}) = \det([T]_{\mathcal{C}})$. $\square$

Visto el corolario anterior, tiene sentido la siguiente definición.

Definición 1.2.6. Si $T \in \mathcal{L}(V)$, definimos su *determinante* por $\det T := \det([T]_{\mathcal{B}})$ siendo $\mathcal{B}$ una base cualquiera de V.

Proposición 1.2.7. *Si* $T \in \mathcal{L}(V)$*, entonces vale lo siguiente.*

1. *El operador* T *es invertible si y solo si* $\det T \neq 0$.
2. *Para todo* λ *en* $\Bbbk$ *y toda base* $\mathcal{B}$ *de* V*, vale* $\det(T - \lambda \operatorname{Id}) = \det(A - \lambda I)$*, siendo* $A = [T]_{\mathcal{B}}$*.*

Dem. Sea $\mathcal{B}$ una base de V. La primera afirmación se deduce de lo siguiente

$$T \text{ es invertible} \quad \Leftrightarrow \quad [T]_{\mathcal{B}} \text{ es invertible} \quad \Leftrightarrow \quad \det([T]_{\mathcal{B}}) \neq 0 \quad \Leftrightarrow \quad \det T \neq 0.$$

Para la segunda, si $\lambda \in \Bbbk$, es

$$\det(T - \lambda \operatorname{Id}) = \det([T - \lambda \operatorname{Id}]_{\mathcal{B}}) = \det([T]_{\mathcal{B}} - \lambda [\operatorname{Id}]_{\mathcal{B}}) = \det(A - \lambda I). \quad \square$$

Definición 1.2.8. Si $T \in \mathcal{L}(V)$, definimos su *polinomio característico* por

$$\chi_T(t) := \det(T - t \operatorname{Id}).$$

Para hallar $\chi_T(t)$, elegimos una base cualquiera $\mathcal{B}$ de V y encontramos la matriz asociada

$$[T]_{\mathcal{B}} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix}, \text{ entonces } \chi_T(t) = \begin{vmatrix} a_{11} - t & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} - t \end{vmatrix}.$$

Es un ejercicio el probar que vale

$$\chi_T(t) = (-1)^n t^n + (-1)^{n-1} \operatorname{tr}(A)\, t^{n-1} + \cdots + \det(A),$$

siendo[1] $A = [T]_{\mathcal{B}}$. Luego $\chi_T(t)$ es un polinomio de grado n en la variable t, siendo $n = \dim V$.

Ejemplo 1.2.9. Sea $V = \mathbb{R}_2[x]$ y $T \in \mathcal{L}(V)$ definida por $T\big(p(x)\big) = p'(x) + p(x)$. Si consideramos la base $\mathcal{B} = \{1, x, x^2\}$, es $T(1) = 1$, $T(x) = 1 + x$ y $T\big(x^2\big) = 2x + x^2$. Luego

$$[T]_{\mathcal{B}} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}.$$

Entonces

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

El siguiente resultado explica nuestro interés en el polinomio característico.

Proposición 1.2.10. *Si* $T \in \mathcal{L}(V)$ *y* $\lambda \in \Bbbk$*, entonces las siguientes afirmaciones son equivalentes.*

[1] En la fórmula de $\chi_T(t)$, $\operatorname{tr}(A)$ es la *traza* de la matriz A.

1. *El escalar λ es un valor propio de T.*

2. $\operatorname{Ker}(T - \lambda \operatorname{Id}) \neq \{0\}$.

3. *El operador $T - \lambda \operatorname{Id}$ no es invertible.*

4. *El escalar λ es raíz de $\chi_T(t)$.*

Dem.

$$\begin{aligned}\lambda \text{ es valor propio de } T &\Leftrightarrow \exists\, v \neq 0: \ T(v) = \lambda\, v \quad \Leftrightarrow \quad \exists\, v \neq 0: \ v \in \operatorname{Ker}(T - \lambda \mathrm{Id}) \\ &\Leftrightarrow \operatorname{Ker}(T - \lambda \mathrm{Id}) \neq \{0\} \quad \Leftrightarrow \quad T - \lambda \mathrm{Id} \text{ no es inyectiva} \\ &\Leftrightarrow T - \lambda \mathrm{Id} \text{ no es biyectiva} \quad \Leftrightarrow \quad T - \lambda \mathrm{Id} \text{ no es invertible} \\ &\Leftrightarrow \det(T - \lambda \mathrm{Id}) = 0 \quad \Leftrightarrow \quad \chi_T(\lambda) = 0. \quad \square\end{aligned}$$

Observación 1.2.11. De la prueba anterior se deduce que si λ es un valor propio de $T \in \mathcal{L}(V)$, entonces $v \in V$ es un vector propio de T correspondiente a λ si y solo si $v \neq 0$ y $v \in \operatorname{Ker}(T - \lambda \operatorname{Id})$.

Corolario 1.2.12. *Si $T \in \mathcal{L}(V)$ y $\dim V = n$, entonces T tiene a lo más n valores propios.*

Dem. Los valores propios de T son las raíces de $\chi_T(t)$, y como este polinomio tiene grado n, entonces tiene a lo más n raíces. $\square$

Recordar que si $\mathcal{B} = \{v_1, \ldots, v_n\}$ es una base de V, entonces el mapa *coordenadas* $\operatorname{coord}_{\mathcal{B}} : V \to \Bbbk^n$ definido por

$$\operatorname{coord}_{\mathcal{B}}(v) = (x_1, \ldots, x_n) \quad \text{si} \quad v = x_1 v_1 + \cdots + x_n v_n,$$

es un isomorfismo lineal.

Proposición 1.2.13. *Sea $T \in \mathcal{L}(V)$, $\mathcal{B}$ una base de V, $A = [T]_{\mathcal{B}}$ y λ un valor propio de T. Entonces $v \in V$ es un vector propio de T correspondiente a λ si y solo si $\operatorname{coord}_{\mathcal{B}}(v) \in \Bbbk^n$ es un vector propio de L_A correspondiente a λ.*

Dem. Utilizando que $\operatorname{coord}_{\mathcal{B}} : V \to \Bbbk^n$ es un isomorfismo, tenemos que dado $v \in V$, es $v \neq 0$ si y solo si $\operatorname{coord}_{\mathcal{B}}(v) \neq 0$ y

$$\begin{aligned}T(v) = \lambda\, v &\Leftrightarrow \operatorname{coord}_{\mathcal{B}}(T(v)) = \operatorname{coord}_{\mathcal{B}}(\lambda\, v) \Leftrightarrow [T]_{\mathcal{B}} \operatorname{coord}_{\mathcal{B}}(v) = \lambda \operatorname{coord}_{\mathcal{B}}(v) \\ &\Leftrightarrow A \operatorname{coord}_{\mathcal{B}}(v) = \lambda \operatorname{coord}_{\mathcal{B}}(v). \quad \square\end{aligned}$$

Ejemplo 1.2.14. Sea $T \in \mathcal{L}(\mathbb{R}_2[x])$ definida por $T(p(x)) = p(x) + x\, p'(x) + p'(x)$. Considerando la base $\mathcal{C} = \{1, x, x^2\}$ de $\mathbb{R}_2[x]$, es

$$A = [T]_{\mathcal{C}} = \begin{pmatrix} 1 & 1 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 3 \end{pmatrix} \quad \Rightarrow \quad \chi_T(t) = \chi_A(t) = -(t-1)(t-2)(t-3),$$

luego los valores propios de T son 1, 2 y 3. Operando con la matriz A obtenemos[2]:

$$\operatorname{Ker}(L_A - \mathrm{Id}) = [(1,0,0)], \quad \operatorname{Ker}(L_A - 2\mathrm{Id}) = [(1,1,0)], \quad \operatorname{Ker}(L_A - 3\mathrm{Id}) = [(1,2,1)].$$

Luego aplicando la proposición anterior deducimos

$$\operatorname{Ker}(T - \mathrm{Id}) = [1], \quad \operatorname{Ker}(T - 2\mathrm{Id}) = [1 + x], \quad \operatorname{Ker}(T - 3\mathrm{Id}) = \left[1 + 2x + x^2\right].$$

[2]Estamos usando $[v_1, \ldots, v_n]$ para denotar el subespacio generado por $v_1, \ldots, v_n$.

El conjunto $\mathcal{B} = \{1,\, 1+x,\, 1+2x+x^2\}$ es LI y por lo tanto $\mathcal{B}$ es una base de $\mathbb{R}_2[x]$ formada por vectores propios de T; luego T es diagonalizable y la matriz asociada a T en la base $\mathcal{B}$ es

$$[T]_{\mathcal{B}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}.$$

Ejemplo 1.2.15. *Rotación.* La rotación de ángulo θ $(0 \leq \theta < 2\pi)$ es la transformación lineal $R_\theta : \mathbb{R}^2 \to \mathbb{R}^2$ definida por $R_\theta(x,y) = (x\cos\theta - y \operatorname{sen}\theta,\ x\operatorname{sen}\theta + y\cos\theta)$. Es

$$R_\theta\begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{pmatrix}\begin{pmatrix} x \\ y \end{pmatrix}, \quad \chi_{R_\theta}(t) = t^2 - (2\cos\theta)t + 1.$$

El discriminante de la ecuación $t^2 - (2\cos\theta)t + 1 = 0$ es

$$\Delta = 4\cos^2\theta - 4 = 4\big(\cos^2\theta - 1\big) \leq 0.$$

Observar que $\Delta = 0$ si y solo si $\theta = 0, \pi$. Si $\theta = 0$, es $R_\theta = \mathrm{Id}$ y todos los vectores no nulos de $\mathbb{R}^2$ son vectores propios de R_θ con valor propio 1. Si $\theta = \pi$, es $R_\theta = -\mathrm{Id}$ (simetría central de centro en el origen) y todos los vectores no nulos de $\mathbb{R}^2$ son vectores propios de R_θ con valor propio -1. Si $\theta \neq 0, \pi$, es $\Delta < 0$ y R_θ no tiene valores propios, por lo cual no es diagonalizable.

Teorema 1.2.16. *Sea $T \in \mathcal{L}(V)$, $\lambda_1, \ldots, \lambda_k$ valores propios distintos de T y $v_1, \ldots, v_k$ vectores propios correspondientes. Entonces $\{v_1, \ldots, v_k\}$ es LI.*

Dem. Lo probaremos por inducción en k.

Si $k = 1$, entonces $\{v_1\}$ es LI porque como v_1 es un vector propio, es $v_1 \neq 0$.

Supongamos que el resultado es cierto para $k - 1 \geq 1$ y sean $\lambda_1, \ldots, \lambda_k$ valores propios distintos de T y $v_1, \ldots, v_k$ vectores propios correspondientes. Sean $a_1, \ldots, a_k \in \Bbbk$ tales que

$$a_1 v_1 + \cdots + a_{k-1}v_{k-1} + a_k v_k = 0. \tag{1.2}$$

Aplicando T en ambos lados de la ecuación (1.2) y teniendo en cuenta que $v_1, \ldots, v_k$ son vectores propios obtenemos

$$a_1\lambda_1 v_1 + \cdots + a_{k-1}\lambda_{k-1}v_{k-1} + a_k\lambda_k v_k = 0. \tag{1.3}$$

Multiplicando la ecuación (1.2) por $-\lambda_k$ obtenemos

$$-a_1\lambda_k v_1 - \cdots - a_{k-1}\lambda_k v_{k-1} - a_k\lambda_k v_k = 0. \tag{1.4}$$

Sumando la ecuación (1.3) y la (1.4) obtenemos

$$a_1(\lambda_1 - \lambda_k)v_1 + \cdots + a_{k-1}(\lambda_{k-1} - \lambda_k)v_{k-1} = 0.$$

Por la hipótesis de inducción el conjunto $\{v_1, \ldots, v_{k-1}\}$ es LI, luego

$$a_1(\lambda_1 - \lambda_k) = \cdots = a_{k-1}(\lambda_{k-1} - \lambda_k) = 0.$$

Como $\lambda_i \neq \lambda_k$ para todo $i = 1, \ldots, k-1$, es

$$a_1 = \cdots = a_{k-1} = 0.$$

Substituyendo $a_1, \ldots, a_{k-1}$ por 0 en la ecuación (1.2) obtenemos $a_k v_k = 0$ y como $v_k \neq 0$, es $a_k = 0$; esto completa la prueba de que $\{v_1, \ldots, v_k\}$ es LI. $\square$

Corolario 1.2.17. *Si la dimensión de V es n y $T \in \mathcal{L}(V)$ tiene n valores propios distintos, entonces T es diagonalizable.*

Dem. Si $\lambda_1, \ldots, \lambda_n$ son los valores propios de T y $v_1, \ldots, v_n$ son vectores propios correspondientes, entonces la proposición anterior implica que $\mathcal{B} = \{v_1, \ldots, v_n\}$ es LI y por lo tanto $\mathcal{B}$ es una base de V formada por vectores propios de T. Luego la proposición 1.1.5 implica que T es diagonalizable. $\square$

Observación 1.2.18. El corolario anterior da una condición suficiente pero no necesaria. Si consideramos $\mathrm{Id} \in \mathcal{L}(V)$, siendo V un espacio arbitrario, entonces vale $[\mathrm{Id}]_{\mathcal{B}} = I$, para toda base $\mathcal{B}$ de V. Luego Id es diagonalizable, pero tiene un único valor propio que es $\lambda = 1$.

Definición 1.2.19. Decimos que un polinomio no constante $p(x) \in \Bbbk[x]$ *se escinde* en $\Bbbk[X]$ si existen escalares $a, a_1, \ldots, a_n$ tales que $p(x) = a(x-a_1)\cdots(x-a_n)$ (pueden haber elementos repetidos en $a_1, \ldots, a_n$).

Para abreviar diremos "$p(x)$ se escinde en $\Bbbk$" en vez de "$p(x)$ se escinde en $\Bbbk[x]$". También a veces diremos simplemente "$p(x)$ se escinde", cuando no sea necesario especificar el cuerpo $\Bbbk$.

Observación 1.2.20. Si el polinomio $p(x)$ se escinde en $\Bbbk$, entonces agrupando los factores repetidos obtenemos $p(x) = a(x-b_1)^{n_1}\cdots(x-b_r)^{n_r}$, con $n_i \geq 1$ para todo i, siendo $b_1, \ldots, b_r$ las distintas raíces de $p(x)$.

Ejemplo 1.2.21.

1. $x^2 - 1 = (x+1)(x-1)$ se escinde en $\mathbb{Q}$.
2. $x^2 - 2$ no se escinde en $\mathbb{Q}$, pero se escinde en $\mathbb{R}$: $x^2 - 2 = \big(x+\sqrt{2}\big)\big(x-\sqrt{2}\big)$.
3. $x^2 + 1$ y $\left(x^2+1\right)(x-2)$ no se escinden en $\mathbb{R}$, pero se escinden en $\mathbb{C}$: $x^2+1 = (x+i)(x-i)$.

Observaciones 1.2.22.

1. En $\mathbb{C}[x]$ todo polinomio no constante se escinde. Este es el llamado *teorema fundamental del álgebra*. Nosotros usaremos este resultado, aunque no veremos su prueba dado que requiere más conocimientos.
2. En estas notas todos los polinomios van a tener coeficientes reales o complejos. Si $p(x) \in \mathbb{R}[x]$, entonces $p(x)$ siempre se escinde en $\mathbb{C}$, y se escinde en $\mathbb{R}$ si y solo si no tiene raíces complejas que no sean reales. Por ejemplo, las raíces de x^2+1 son $\pm i$ que no están en $\mathbb{R}$, luego x^2+1 no se escinde en $\mathbb{R}$.

Proposición 1.2.23. *Sea $T \in \mathcal{L}(V)$. Si T es diagonalizable, entonces $\chi_T(t)$ se escinde.*

Dem. Sea $\mathcal{B}$ una base de V en la cual $[T]_{\mathcal{B}}$ es diagonal.

$$[T]_{\mathcal{B}} = \begin{pmatrix} a_1 & & \\ & \ddots & \\ & & a_n \end{pmatrix} \quad \Rightarrow \quad \chi_T(t) = (a_1 - t)\cdots(a_n - t) = (-1)^n (t-a_1)\cdots(t-a_n). \quad \square$$

Corolario 1.2.24. *Sea $T \in \mathcal{L}(V)$. Si $\chi_T(t)$ no se escinde en $\Bbbk$, entonces T no es diagonalizable.* $\square$

Observación 1.2.25. Que el polinomio característico se escinda en $\Bbbk$ es una condición necesaria pero no suficiente para que un operador sea diagonalizable. Por ejemplo, si consideramos $T \in \mathcal{L}(\mathbb{R}_n[x])$ definida por $T\big(p(x)\big) = p'(x)$, entonces es $\chi_T(t) = (-1)^n t^n$. Luego $\chi_T(t)$ se escinde, pero ya vimos en el ejemplo 1.1.6 que T no es diagonalizable. El problema en este caso es la existencia de raíces múltiples en $\chi_T(t)$, que es lo que estudiaremos en la próxima sección.

1.3. Multiplicidad geométrica y algebraica

Definición 1.3.1. Sea $T \in \mathcal{L}(V)$ y λ un valor propio de T.

1. El *subespacio propio* asociado al valor propio λ es $E_\lambda := \operatorname{Ker}(T - \lambda \mathrm{Id})$.
2. La *multiplicidad geométrica* de λ es $\mathrm{MG}(\lambda) := \dim E_\lambda$.
3. La *multiplicidad algebraica* de λ es $\mathrm{MA}(\lambda) := \text{máx}\left\{h : \ (t-\lambda)^h \text{ divide a } \chi_T(t)\right\}$.

Observaciones 1.3.2. Sea $T \in \mathcal{L}(V)$.

1. El escalar 0 es un valor propio de T si y solo si $\operatorname{Ker}(T) \neq \{0\}$; en ese caso es $E_0 = \operatorname{Ker}(T)$.
2. El subespacio propio asociado a un valor propio λ consiste en el vector nulo y los vectores propios correspondientes a λ.
3. Si λ es un valor propio de T, entonces[3]

$$\mathrm{MG}(\lambda) = \dim V - \operatorname{rango}(T - \lambda \mathrm{Id}).$$

   Eso se deduce directamente de la fórmula $\dim \operatorname{Ker}(T - \lambda \mathrm{Id}) + \dim \operatorname{Im}(T - \lambda \mathrm{Id}) = \dim V$.

4. Si λ es un valor propio de T, es $\mathrm{MA}(\lambda) = m$ si y solo si $\chi_T(t) = (t-\lambda)^m p(t)$ con $p(\lambda) \neq 0$.
5. Si $\chi_T(t)$ se escinde y se escribe de la forma $\chi_T(t) = (-1)^n (t-\lambda_1)^{n_1} \cdots (t-\lambda_h)^{n_h}$ con $\lambda_i \neq \lambda_j$ si $i \neq j$, entonces $\mathrm{MA}(\lambda_i) = n_i$, para todo $i = 1, \ldots, h$.

Ejemplo 1.3.3. Sea $T \in \mathcal{L}(\mathbb{R}^4)$ definida por $T(x,y,z,t) = (3x,\ 3y,\ 5z,\ z+5t)$. Es $\chi_T(t) = (t-3)^2(t-5)^2$, luego los valores propios de T son 3 y 5. En este caso es

$$\mathrm{MG}(3) = 4 - \operatorname{rango}(T - 3\mathrm{Id}) = 4 - 2 = 2; \quad \mathrm{MG}(5) = 4 - \operatorname{rango}(T - 5\mathrm{Id}) = 4 - 3 = 1.$$

Luego $\mathrm{MG}(3) = \mathrm{MA}(3) = 2$, $\mathrm{MG}(5) = 1$ y $\mathrm{MA}(5) = 2$.

Teorema 1.3.4. *Sea $T \in \mathcal{L}(V)$ y λ un valor propio de T, entonces*

$$1 \leq \mathrm{MG}(\lambda) \leq \mathrm{MA}(\lambda).$$

Dem. Como λ es un valor propio de T, es $E_\lambda \neq \{0\}$ y luego $\mathrm{MG}(\lambda) \geq 1$.

Sea $\mathcal{B}_1$ una base de E_λ y consideremos $\mathcal{B}_2$ un subconjunto LI de V, disjunto con $\mathcal{B}_1$, tal que $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$ es una base de V. Observar que $T(v) = \lambda v$ para todo $v \in \mathcal{B}_1$, luego la matriz asociada a T en la base $\mathcal{B}$ es una matriz en bloques de la forma

$$[T]_\mathcal{B} = \begin{pmatrix} \lambda I_m & B \\ 0 & C \end{pmatrix}$$

siendo $m = \#\mathcal{B}_1 = \dim E_\lambda = \mathrm{MG}(\lambda)$. Notar que $[T]_\mathcal{B}$ es una matriz en bloques, luego aplicando la proposición 8.2.1 obtenemos

$$\chi_T(t) = \begin{vmatrix} (\lambda - t) I_m & B \\ 0 & C - tI \end{vmatrix} = |(\lambda - t) I_m| \cdot |C - tI| = (\lambda - t)^m \chi_C(t).$$

Como puede ser $\chi_C(\lambda) = 0$, deducimos $\mathrm{MA}(\lambda) \geq m = \mathrm{MG}(\lambda)$. □

[3] Recordar que el *rango* de una transformación lineal T es la dimensión de la imagen de T. El rango de T coincide con el rango de la matriz asociada a T en cualquier par de bases.

Corolario 1.3.5. *Sea $T \in \mathcal{L}(V)$ y λ un valor propio de T, entonces* $\mathrm{MA}(\lambda) = 1$ *implica* $\mathrm{MG}(\lambda) = 1$. □

En lo que sigue usaremos la suma directa de subespacios (§8.1).

Proposición 1.3.6. *Sea $T \in \mathcal{L}(V)$ y $\lambda_1, \ldots, \lambda_k$ valores propios distintos de T. Entonces $E_{\lambda_1}, \ldots, E_{\lambda_k}$ son subespacios independientes.*

Dem. Sean $v_1 \in E_{\lambda_1}, \ldots, v_k \in E_{\lambda_k}$ tales que $v_1 + \cdots + v_k = 0$. Si existiese algún $i \in \{1, \ldots, k\}$ tal que $v_i \neq 0$, eventualmente reordenando los subíndices, existiría algún l, $1 \leq l \leq k$, tal que $v_i \neq 0$, para todo $i = 1, \ldots, l$ y $v_{l+1} = \cdots = v_k = 0$. Luego es

$$v_1 + \cdots + v_l = 0, \tag{1.5}$$

con $v_i \neq 0$, para todo $i = 1, \ldots, l$. Para cada $i \in \{1, \ldots, l\}$ es $0 \neq v_i \in E_{\lambda_i}$, luego v_i es un vector propio de T correspondiente al valor propio λ_i y por lo tanto el teorema 1.2.16 implica que $\{v_1, \ldots, v_l\}$ es LI. Esto contradice (1.5). Luego es $v_1 = \cdots = v_k = 0$ y por lo tanto $E_{\lambda_1}, \ldots, E_{\lambda_k}$ son subespacios independientes. □

Terminamos esta sección con el siguiente resultado, que establece condiciones necesarias y suficientes para que un operador sea diagonalizable.

Teorema 1.3.7. *Sean $T \in \mathcal{L}(V)$ y $\lambda_1, \ldots, \lambda_h$ sus valores propios. Las siguientes afirmaciones son equivalentes.*

1. *El operador T es diagonalizable.*
2. *El polinomio $\chi_T(t)$ se escinde y* $\mathrm{MG}(\lambda_i) = \mathrm{MA}(\lambda_i)$ *para todo $i = 1, \ldots, h$.*
3. *Vale $V = \sum_{i=1}^{h} E_{\lambda_i}$.*
4. *Vale $V = \bigoplus_{i=1}^{h} E_{\lambda_i}$.*

Dem. Observemos primero que la proposición 1.3.6 nos dice que $E_{\lambda_1}, \ldots, E_{\lambda_k}$ son subespacios independientes y por lo tanto su suma es directa. Luego las afirmaciones 3 y 4 son equivalentes.
($1 \Rightarrow 2$). Como T es diagonalizable, sabemos que $\chi_T(t)$ se escinde y que V admite una base $\mathcal{B}$ formada por vectores propios de T. Agrupando los vectores propios correspondientes a cada valor propio, podemos suponer que esta base es de la forma $\mathcal{B} = \left\{v_1^1, \ldots, v_{n_1}^1, \ldots, v_1^h, \ldots, v_{n_h}^h\right\}$, siendo $\lambda_1, \ldots, \lambda_h$ los valores propios de T, con $\lambda_i \neq \lambda_j$ si $i \neq j$. Luego

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & & & & & & \\ & \ddots & & & & & \\ & & \lambda_1 & & & & \\ & & & \ddots & & & \\ & & & & \lambda_h & & \\ & & & & & \ddots & \\ & & & & & & \lambda_h \end{pmatrix}.$$

Notar que cada valor propio λ_i aparece exactamente n_i veces en $[T]_{\mathcal{B}}$. Luego

$$\chi_T(t) = (\lambda_1 - t)^{n_1} \cdots (\lambda_h - t)^{n_h} = (-1)^n (t - \lambda_1)^{n_1} \cdots (t - \lambda_h)^{n_h}.$$

Observar que por la forma de la matriz $[T]_{\mathcal{B}}$ es $\left\{v_1^i, \ldots, v_{n_i}^i\right\} \subset E_{\lambda_i}$, luego $\mathrm{MG}(\lambda_i) \geq n_i = \mathrm{MA}(\lambda_i)$. Pero siempre vale $\mathrm{MG}(\lambda_i) \leq \mathrm{MA}(\lambda_i)$, luego $\mathrm{MG}(\lambda_i) = \mathrm{MA}(\lambda_i)$, para todo $i = 1, \ldots, h$.

(2 ⇒ 4). Sea $\chi_T(t) = (-1)^n (t-\lambda_1)^{n_1}\cdots(t-\lambda_h)^{n_h}$ con $\lambda_i \neq \lambda_j$ si $i \neq j$. Entonces

$$\dim\left(\bigoplus_{i=1}^{h} E_{\lambda_i}\right) = \sum_{i=1}^{h}\dim E_{\lambda_i} = \sum_{i=1}^{h}\mathrm{MG}(\lambda_i) = \sum_{i=1}^{h}\mathrm{MA}(\lambda_i) = \sum_{i=1}^{h} n_i = \mathrm{gr}\ \chi_T(t) = \dim V \ \Rightarrow\ \bigoplus_{i=1}^{h} E_{\lambda_i} = V.$$

(4 ⇒ 1). Si $\mathcal{B}_i$ es una base de E_{λ_i} para todo $i = 1,\dots,h$, entonces $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_h$ es una base de V formada por vectores propios de T. Luego la proposición 1.1.5 implica que T es diagonalizable. □

1.4. Matrices diagonalizables

A cada matriz $A \in M_n(\Bbbk)$ le podemos asociar la transformación lineal $L_A \in \mathcal{L}(\Bbbk^n)$. Recíprocamente, dada una transformación lineal $T \in \mathcal{L}(V)$, con V de dimensión n, a cada base $\mathcal{B}$ de V le podemos asociar la matriz $[T]_\mathcal{B} \in M_n(\Bbbk)$. Los dos resultados siguientes muestran que estas correspondencias se relacionan bien, lo cual nos permitirá trasladar lo que vimos de transformaciones lineales a matrices.

Lema 1.4.1. *Sea $\mathcal{B} = \{v_1,\dots,v_n\}$ una base de un espacio V y $A = (a_{ij}) \in M_n(\Bbbk)$ una matriz invertible. Definimos $w_j := \sum_{i=1}^n a_{ij}\, v_i$, para todo $j = 1\dots,n$. Entonces $\mathcal{C} = \{w_1,\dots,w_n\}$ es base de V y ${}_\mathcal{B}[\mathrm{Id}]_\mathcal{C} = A$.*

Dem. Sean $x_1\dots,x_n \in \Bbbk$ tales que $\sum_{j=1}^n x_j w_j = 0$. Entonces

$$0 = \sum_{j=1}^{n} x_j w_j = \sum_{j=1}^{n} x_j \sum_{i=1}^{n} a_{ij}\, v_i = \sum_{j=1}^{n}\sum_{i=1}^{n} x_j a_{ij}\, v_i = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} a_{ij} x_j\right) v_i.$$

Como $\mathcal{B} = \{v_1,\dots,v_n\}$ es LI, esto implica $\sum_{j=1}^n a_{ij}x_j = 0$, para todo $j = 1\dots,n$. Esto puede escribirse

$$\begin{cases} a_{11}x_1 + \cdots + a_{1n}x_n &= \ 0 \\ & \ \vdots \\ a_{n1}x_1 + \cdots + a_{nn}x_n &= \ 0 \end{cases} \quad\Leftrightarrow\quad \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn}\end{pmatrix}\begin{pmatrix} x_1 \\ \vdots \\ x_n\end{pmatrix} = \begin{pmatrix} 0 \\ \vdots \\ 0\end{pmatrix}.$$

Como A es invertible, esta última igualdad implica $x_1 = \cdots = x_n = 0$. Esto prueba que $\mathcal{C}$ es LI y por lo tanto es base de V. Además vale $\mathrm{coord}_\mathcal{B}(w_j) = (a_{1j},\dots,a_{nj})$, para todo $j = 1\dots,n$. Luego ${}_\mathcal{B}[\mathrm{Id}]_\mathcal{C} = A$. □

Proposición 1.4.2. *Sean $T \in \mathcal{L}(V)$ y $\mathcal{B}$ una base de V.*

1. *Si $\mathcal{B}'$ es otra base de V, entonces $[T]_{\mathcal{B}'} \simeq [T]_\mathcal{B}$. Explícitamente, $[T]_{\mathcal{B}'} = P\,[T]_\mathcal{B}\,P^{-1}$, siendo $P = {}_{\mathcal{B}'}[\mathrm{Id}]_\mathcal{B}$.*
2. *Si $A \in M_n(\Bbbk)$ es tal que $A \simeq [T]_\mathcal{B}$, entonces existe una base $\mathcal{B}''$ de V tal que $A = [T]_{\mathcal{B}''}$.*

Dem. La primera afirmación es la proposición 1.2.4. Veamos la segunda. Supongamos $A = Q\,[T]_\mathcal{B}\,Q^{-1}$, siendo $Q^{-1} = (r_{ij})$ y $\mathcal{B} = \{v_1,\dots,v_n\}$. Definimos $w_j := \sum_{i=1}^n r_{ij}\, v_i$, para todo $j = 1\dots,n$. Como la matriz Q^{-1} es invertible, entonces el conjunto $\mathcal{B}'' = \{w_1,\dots,w_n\}$ es una base de V y ${}_\mathcal{B}[\mathrm{Id}]_{\mathcal{B}''} = Q^{-1}$. Luego

$$A = Q\,[T]_\mathcal{B}\,Q^{-1} = {}_{\mathcal{B}''}[\mathrm{Id}]_\mathcal{B} \cdot [T]_\mathcal{B} \cdot {}_\mathcal{B}[\mathrm{Id}]_{\mathcal{B}''} = [T]_{\mathcal{B}''}. \qquad \square$$

Corolario 1.4.3. *Dos matrices son semejantes si y solo si representan a una misma transformación lineal.*

Dem. Sean A y B en $M_n(\Bbbk)$ tales que $A \simeq B$. Si consideramos $T = L_A \in \mathcal{L}(\Bbbk^n)$, es $A = [T]_\mathcal{C}$ siendo $\mathcal{C}$ la base canónica de $\Bbbk^n$. Luego $B \simeq [T]_\mathcal{C}$ y la parte 2 de la Proposición 1.4.2 implica que existe $\mathcal{B}$ base de $\Bbbk^n$ tal que $B = [T]_\mathcal{B}$. El recíproco es la parte 1 de la Proposición 1.4.2. □

Definición 1.4.4. Una matriz $A \in M_n(\Bbbk)$ es *diagonalizable* si es semejante a una matriz diagonal.

El siguiente resultado relaciona la diagonalización en matrices con la diagonalización en operadores.

Proposición 1.4.5. *Sea $T \in \mathcal{L}(V)$.*

1. *Si T es diagonalizable y $\mathcal{B}$ es una base cualquiera de V, entonces $[T]_{\mathcal{B}}$ es diagonalizable.*
2. *Si existe una base $\mathcal{B}$ de V tal que $[T]_{\mathcal{B}}$ es diagonalizable, entonces T es diagonalizable.*

Dem. Si T es diagonalizable, entonces existe una base $\mathcal{C}$ de V en la cual $[T]_{\mathcal{C}}$ es diagonal. Luego la parte 1 de la Proposición 1.4.2 implica que para toda base $\mathcal{B}$ se cumple que $[T]_{\mathcal{B}}$ es semejante a la matriz $[T]_{\mathcal{C}}$.

Si existe una base $\mathcal{B}$ de V tal que $[T]_{\mathcal{B}}$ es diagonalizable, entonces $[T]_{\mathcal{B}}$ es semejante a una matriz diagonal D. Luego la parte 2 de la Proposición 1.4.2 implica que existe una base $\tilde{\mathcal{B}}$ de V tal que $[T]_{\tilde{\mathcal{B}}} = D$. □

Como, dada una matriz $A \in M_n(\mathbb{k})$, la matriz asociada al operador $L_A \in \mathcal{L}(\mathbb{k}^n)$ en la base canónica es la matriz A, se obtiene el siguiente resultado.

Corolario 1.4.6. *Una matriz $A \in M_n(\mathbb{k})$ es diagonalizable si y solo su operador asociado $L_A \in \mathcal{L}(\mathbb{k}^n)$ es diagonalizable.* □

Lo que sigue es la versión matricial de la proposición 1.1.5.

Proposición 1.4.7. *Una matriz $A \in M_n(\mathbb{k})$ es diagonalizable si y solo si existe una base $\mathcal{B}$ de $\mathbb{k}^n$ formada por vectores propios de L_A. En este caso, si $\mathcal{B} = \{v_1, \dots, v_n\}$ y $\lambda_1, \dots, \lambda_n$ son los valores propios correspondientes, es*

$$A = QDQ^{-1}, \quad \textit{siendo} \quad D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{pmatrix} \quad \textit{y} \quad Q = [v_1|\cdots|v_n].$$

Dem. La primera afirmación se deduce de la proposición 1.1.5 aplicada al operador $L_A : \mathbb{k}^n \to \mathbb{k}^n$. Para la segunda, si $\mathcal{C}$ es la base canónica de $\mathbb{k}^n$, es $A = [L_A]_{\mathcal{C}} = {}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}} \cdot [L_A]_{\mathcal{B}} \cdot {}_{\mathcal{B}}[\mathrm{Id}]_{\mathcal{C}} = QDQ^{-1}$. □

El corolario anterior nos dice que para estudiar la diagonalizabilidad de $A \in M_n(\mathbb{k})$ podemos aplicar todo lo que vimos anteriormente al operador $L_A \in \mathcal{L}(\mathbb{k}^n)$, en particular el teorema 1.3.7. Para no estar pasando todo el tiempo traduciendo de operadores a matrices, daremos las siguientes definiciones. Dada $A \in M_n(\mathbb{k})$, sus valores propios, vectores propios, polinomio característico, etc., son los correspondientes al operador $L_A \in \mathcal{L}(\mathbb{k}^n)$. En particular destacamos los siguientes.

1. Un escalar $\lambda \in \mathbb{k}$ es un *valor propio* de A si existe $0 \neq v \in \mathbb{k}^n$ tal que $Av = \lambda v$; en ese caso v es un *vector propio* de A asociado a λ.
2. El *polinomio característico* de A es $\chi_A(t) := \det(A - tI)$. Es un polinomio de grado n en la variable t.
3. El *subespacio propio* correspondiente a un valor propio λ de A es

$$E_\lambda = \mathrm{Ker}(A - \lambda I) := \{v \in \mathbb{k}^n : \ (A - \lambda I)v = 0\} = \{v \in \mathbb{k}^n : \ Av = \lambda v\}.$$

   La *multiplicidad geométrica* de λ es $\mathrm{MG}(\lambda) = \dim \mathrm{Ker}(A - \lambda I) = n - \mathrm{rango}(A - \lambda I)$.

Ejemplos 1.4.8. 1. Sea $A = \left(\begin{smallmatrix} 1 & 3 \\ 4 & 2 \end{smallmatrix}\right) \in M_2(\mathbb{R})$. Es $\chi_A(t) = \left|\begin{smallmatrix} 1-t & 3 \\ 4 & 2-t \end{smallmatrix}\right| = (t+2)(t-5)$. Luego A tiene dos valores propios distintos y por lo tanto es diagonalizable. Los subespacios propios son $E_{-2} = \mathrm{Ker}(A - 2I) = [(1,-1)]$ y $E_5 = \mathrm{Ker}(A + 5I) = [(3,4)]$. Luego

$$\begin{pmatrix} 1 & 3 \\ 4 & 2 \end{pmatrix} = \begin{pmatrix} 1 & 3 \\ -1 & 4 \end{pmatrix} \begin{pmatrix} -2 & 0 \\ 0 & 5 \end{pmatrix} \begin{pmatrix} 1 & 3 \\ -1 & 4 \end{pmatrix}^{-1}, \quad \text{siendo} \ \begin{pmatrix} 1 & 3 \\ -1 & 4 \end{pmatrix}^{-1} = \begin{pmatrix} 4/7 & -3/7 \\ 1/7 & 1/7 \end{pmatrix}.$$

2. Si $A=\left(\begin{smallmatrix}0&1\\-1&0\end{smallmatrix}\right)\in M_2(\mathbb{R})$, es $\chi_A(t)=t^2+1$ que no se escinde en $\mathbb{R}$, luego A no es diagonalizable en $M_2(\mathbb{R})$. Observar que si consideramos $A\in M_2(\mathbb{C})$, es $\chi_A(t)=t^2+1=(t-i)(t+i)$ que se escinde en $\mathbb{C}$ con raíces simples. Luego A es diagonalizable en $M_2(\mathbb{C})$ y es semejante a $\left(\begin{smallmatrix}i&0\\0&-i\end{smallmatrix}\right)$. Hallando los subespacios propios obtenemos $E_i=[(1,i)]$ y $E_{-i}=[(1,-i)]$. Luego
$$\begin{pmatrix}0&1\\-1&0\end{pmatrix}=\begin{pmatrix}1&1\\i&-i\end{pmatrix}\begin{pmatrix}i&0\\0&-i\end{pmatrix}\begin{pmatrix}1&1\\i&-i\end{pmatrix}^{-1},\quad \text{siendo } \begin{pmatrix}1&1\\i&-i\end{pmatrix}^{-1}=\begin{pmatrix}1/2&-i/2\\1/2&i/2\end{pmatrix}.$$
3. Sea $A=\left(\begin{smallmatrix}\lambda&1\\0&\lambda\end{smallmatrix}\right)\in M_2(\mathbb{C})$, siendo $\lambda\in\mathbb{C}$ arbitrario. Es $\chi_A(t)=(t-\lambda)^2$ y por lo tanto λ es el único valor propio de A. Notar que es $\mathrm{MA}(\lambda)=2$ y $\mathrm{MG}(\lambda)=2-\operatorname{rango}(A-\lambda I)=1$. Luego A no es diagonalizable en $M_2(\mathbb{C})$, para ningún valor de λ.

Finalizamos esta sección con el siguiente resultado, que no es obvio a partir de las definiciones.

Proposición 1.4.9. *Sean $A,D\in M_n(\mathbb{R})$ tales que D es una matriz diagonal. Si $A\simeq D$ en $M_n(\mathbb{C})$, entonces $A\simeq D$ en $M_n(\mathbb{R})$.*

Dem. Consideremos las transformaciones lineales $T:\mathbb{C}^n\to\mathbb{C}^n$ y $S:\mathbb{R}^n\to\mathbb{R}^n$ definidas por $T(v)=Av$, para todo $v\in\mathbb{C}^n$ y $S(v)=Av$, para todo $v\in\mathbb{R}^n$. Si $\mathcal{B}$ es la base canónica de $\mathbb{C}^n$ y $\mathcal{C}$ es la base canónica de $\mathbb{R}^n$, entonces es $[T]_\mathcal{B}=[S]_\mathcal{C}=A$. Además sabemos $A\simeq D$ en $M_n(\mathbb{C})$, luego $\chi_T(t)=\chi_S(t)=\chi_A(t)=\chi_D(t)$. Como $D\in M_n(\mathbb{R})$ es una matriz diagonal, es $\chi_D(t)=(-1)^n(t-\lambda_1)^{n_1}\cdots(t-\lambda_h)^{n_h}$, siendo $\lambda_1,\dots,\lambda_h\in\mathbb{R}$ las distintas entradas diagonales de D. Luego
$$\chi_T(t)=\chi_S(t)=(-1)^n(t-\lambda_1)^{n_1}\cdots(t-\lambda_h)^{n_h}.$$
Como $[T]_\mathcal{B}=A\simeq D$ en $M_n(\mathbb{C})$, entonces $T:\mathbb{C}^n\to\mathbb{C}^n$ es diagonalizable. Esto implica que para cada valor propio λ_i, la multiplicidad algebraica coincide con la multiplicidad geométrica, es decir,
$$n_i=n-\operatorname{rango}(T-\lambda_i\mathrm{Id})=n-\operatorname{rango}(A-\lambda_iI),\quad \forall i=1,\dots,h.$$
Pero $A-\lambda_iI\in M_n(\mathbb{R})$, y el rango de una matriz real es el mismo[4] pensada en $M_n(\mathbb{R})$ o $M_n(\mathbb{C})$, entonces el teorema 1.3.7 aplicado a $S:\mathbb{R}^n\to\mathbb{R}^n$ implica que S es diagonalizable con valores propios $\lambda_1,\dots,\lambda_h$ y multiplicidades respectivas $n_1,\dots,n_h$. Al ser $[S]_\mathcal{C}=A$, concluimos $A\simeq D$ en $M_n(\mathbb{R})$. $\square$

[4] El rango de una matriz real es el mismo pensada en $M_n(\mathbb{R})$ o $M_n(\mathbb{C})$. Esto se prueba usando la definición del rango por determinantes. También, con un poco más de trabajo, se lo puede probar usando la definición del rango por columnas o por filas.

Capítulo 2

Espacios con producto interno

En este tema el cuerpo de base $\mathbb{k}$ será $\mathbb{R}$ o $\mathbb{C}$. Diremos que un $\mathbb{k}$-espacio vectorial es *real* si $\mathbb{k} = \mathbb{R}$ y que es *complejo* si $\mathbb{k} = \mathbb{C}$. Recordar que si $z = a + bi \in \mathbb{C}$, entonces su *conjugado* es $\overline{z} = a - bi \in \mathbb{C}$. Como es usual pensamos $\mathbb{R} \subset \mathbb{C}$; en particular, si $z \in \mathbb{C}$, entonces $z \in \mathbb{R}$ si y solo si $\overline{z} = z$.

2.1. Definiciones y propiedades básicas

La definición que sigue es la generalización del producto escalar de $\mathbb{R}^3$ a un espacio vectorial, real o complejo, arbitrario.

Definición 2.1.1. Sea V un espacio vectorial. Un *producto interno* en V es una función $\langle\ ,\ \rangle : V \times V \to \mathbb{k}$ que verifica:

1. $\langle u + v, w\rangle = \langle u, w\rangle + \langle v, w\rangle,\ \forall u, v, w \in V$.
2. $\langle au, v\rangle = a\langle u, v\rangle,\ \forall a \in \mathbb{k},\ u, v \in V$.
3. $\langle u, v\rangle = \overline{\langle v, u\rangle},\ \forall u, v \in V$.
4. $\langle v, v\rangle > 0,\ \forall v \neq 0$.

Notar que si $\mathbb{k} = \mathbb{R}$, entonces la condición (3) queda en $\langle u, v\rangle = \langle v, u\rangle$, para todo $u, v \in V$.

Un *espacio con producto interno* es un par $(V, \langle\ ,\ \rangle)$ en el cual V es un espacio vectorial y $\langle\ ,\ \rangle : V \times V \to \mathbb{k}$ es un producto interno en V.

Proposición 2.1.2. *Si $\langle\ ,\ \rangle$ es un producto interno en V, entonces valen las siguientes propiedades.*

1. $\langle u, v + w\rangle = \langle u, v\rangle + \langle u, w\rangle$, *para todo* $u, v, w \in V$.
2. $\langle u, av\rangle = \overline{a}\langle u, v\rangle$, *para todo* $u, v \in V$ *y* $a \in \mathbb{k}$.

Dem.

1. $\langle u, v + w\rangle = \overline{\langle v + w, u\rangle} = \overline{\langle v, u\rangle + \langle w, u\rangle} = \overline{\langle v, u\rangle} + \overline{\langle w, u\rangle} = \langle u, v\rangle + \langle u, w\rangle$.
2. $\langle u, av\rangle = \overline{\langle av, u\rangle} = \overline{a\langle v, u\rangle} = \overline{a}\,\overline{\langle v, u\rangle} = \overline{a}\langle u, v\rangle$. □

Observación 2.1.3. Todo producto interno es lineal en la primera variable. Respecto a la segunda variable, en el caso real también es lineal, pero en el caso complejo es conjugado-lineal.

Ejemplo 2.1.4. El producto escalar de $\mathbb{R}^3$ definido por

$$(x, y, z) \cdot (x', y', z') = xx' + yy' + zz'$$

es un producto interno. Este producto se generaliza a $\mathbb{R}^n$ $(n \geq 1)$, definiendo

$$\langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle := \sum_{i=1}^{n} x_i y_i = x_1 y_1 + x_2 y_2 + \dots + x_n y_n.$$

Este es un producto interno llamado el *producto escalar* de $\mathbb{R}^n$. Con una pequeña modificación, obtenemos un producto interno en $\mathbb{C}^n$ definiendo

$$\langle (x_1, \dots, x_n), (y_1, \dots, y_n) \rangle := \sum_{i=1}^{n} x_i \overline{y_i} = x_1 \overline{y_1} + x_2 \overline{y_2} + \dots + x_n \overline{y_n}. \tag{2.1}$$

Este es el producto interno usual de $\mathbb{C}^n$. Notar que si pensamos $\mathbb{R} \subset \mathbb{C}$, entonces el producto interno usual de $\mathbb{C}^n$ restringido a $\mathbb{R}^n$ nos da el producto escalar, así que podemos usar la formula (2.1) como definición del producto interno usual en $\mathbb{k}^n$, tanto para el caso $\mathbb{k} = \mathbb{R}$ o $\mathbb{k} = \mathbb{C}$. Cuando consideremos a $\mathbb{k}^n$ como espacio vectorial con producto interno, nos estaremos refiriendo siempre al producto interno (2.1), a menos que explicitemos lo contrario.

Ejemplo 2.1.5. Si $A = (a_{ij}), B = (b_{ij}) \in M_{m \times n}(\mathbb{k})$, entonces definimos $\langle A, B \rangle = \sum_{i,j=1}^{m,n} a_{ij} \overline{b_{ij}}$. Es fácil de probar que la fórmula anterior define un producto interno en $M_{m \times n}(\mathbb{k})$. Observar que si identificamos $M_{m \times n}(\mathbb{k})$ con $\mathbb{k}^{mn}$ mediante el isomorfismo $T : M_{m \times n}(\mathbb{k}) \to \mathbb{k}^{mn}$ definido por

$$T\big((a_{ij})\big) = (a_{11}, \dots, a_{1n}, \ \dots \ , a_{m1}, \dots, a_{mn}),$$

entonces este producto interno en $M_{m \times n}(\mathbb{k})$ coincide con el producto interno usual de $\mathbb{k}^{mn}$.

Ejemplo 2.1.6. Sea $C\big([0,1]\big) := \{f : [0,1] \to \mathbb{R} : \ f \text{ es continua}\}$, entonces

$$\langle f, g \rangle = \int_0^1 f(t)\, g(t)\, dt$$

define un producto interno en $C\big([0,1]\big)$. Probaremos solo la propiedad 4, que es la única que requiere cierta elaboración. Sea $f \in C\big([0,1]\big)$ tal que $f \neq 0$. Esto último implica que existe $x_0 \in [0,1]$ tal que $f(x_0) \neq 0$. Supongamos $x_0 \in (0,1)$. Como f^2 es continua y $f^2(x_0) > 0$, entonces existen $\epsilon > 0$ y $\delta > 0$ tales que $(x_0 - \delta, x_0 + \delta) \subset [0,1]$ y $f^2(x) > \epsilon$, para todo $x \in (x_0 - \delta, x_0 + \delta)$. Luego

$$\langle f, f \rangle = \int_0^1 f^2(t)\, dt = \int_0^{x_0 - \delta} f^2(t)\, dt + \int_{x_0 - \delta}^{x_0 + \delta} f^2(t)\, dt + \int_{x_0 + \delta}^{1} f^2(t)\, dt \geq \int_{x_0 - \delta}^{x_0 + \delta} f^2(t)\, dt > 2\delta\epsilon > 0.$$

Los casos $x_0 = 0$ y $x_0 = 1$ se prueban en forma similar y quedan como ejercicio.

Notar que si en vez de funciones continuas consideramos funciones integrables, entonces la propiedad 4 deja de valer y el anterior no es un producto interno.

De ahora en adelante $(V, \langle\ ,\ \rangle)$ es un espacio con producto interno.

Proposición 2.1.7. *Valen las siguientes propiedades.*

1. $\langle v, v \rangle = 0$ *si y solo si* $v = 0$.

2. Si $\langle v, u\rangle = 0$ para todo $v \in V$, entonces $u = 0$.

3. Si $\langle v, u\rangle = \langle v, w\rangle$ para todo $v \in V$, entonces $u = w$.

Dem.

1. Para que sea menos confuso, en esta parte escribiremos o al vector nulo. Por un lado, es $\langle o, o\rangle = \langle 0o, o\rangle = 0\langle o, o\rangle = 0$. Por el otro lado, si $v \neq o$, entonces $\langle v, v\rangle > 0$ y por lo tanto $\langle v, v\rangle \neq 0$. Luego el único caso en que vale $\langle v, v\rangle = 0$ es cuando $v = o$.

2. Si $\langle v, u\rangle = 0$ para todo $v \in V$, entonces tomando $v = u$ es $\langle u, u\rangle = 0$, luego $u = 0$.

3. Si $\langle v, u\rangle = \langle v, w\rangle$ para todo $v \in V$, entonces es $0 = \langle v, u\rangle - \langle v, w\rangle = \langle v, u - w\rangle$ para todo $v \in V$; luego es $u - w = 0$ y resulta $u = w$. □

Observación 2.1.8. De la proposición anterior, se deduce que la condición 4 de la definición de producto interno equivale a las dos siguientes.

$$\langle v, v\rangle \geq 0,\ \forall v \in V \quad \text{y} \quad \langle v, v\rangle = 0 \ \Rightarrow\ v = 0.$$

Definición 2.1.9. La *norma* de un vector $v \in V$ es el número real definido por $||v|| := \sqrt{\langle v, v\rangle}$. La norma define una función $||\ || : V \to \mathbb{R}$.

Ejemplo 2.1.10. Si $V = \mathbb{k}^n$ es

$$||(x_1, \dots, x_n)\,|| = \sqrt{|x_1|^2 + |x_2|^2 + \cdots + |x_n|^2} = \sqrt{\sum_{i=1}^{n} |x_i|^2}.$$

Notar que en la fórmula anterior, $|x|$ denota el módulo en el caso complejo y el valor absoluto en el caso real (que coinciden cuando pensamos $\mathbb{R} \subset \mathbb{C}$).

Observación 2.1.11. Si $z = a + bi \in \mathbb{C}$ con $a, b \in \mathbb{R}$, entonces $\operatorname{Re}(z) := a$ e $\operatorname{Im}(z) := b$ son su *parte real* e *imaginaria*, respectivamente. Para todo $z \in \mathbb{C}$ vale

$$|\operatorname{Re}(z)| \leq |z|, \quad |\operatorname{Im}(z)| \leq |z|, \quad z + \overline{z} = 2\operatorname{Re}(z), \quad z - \overline{z} = 2i\operatorname{Im}(z) \quad \text{y} \quad z\overline{z} = |z|^2.$$

Proposición 2.1.12. *Vale la siguiente igualdad*

$$||u + v||^2 = ||u||^2 + 2\operatorname{Re}(\langle u, v\rangle) + ||v||^2, \quad \forall u, v \in V. \tag{2.2}$$

Dem.

$$\begin{aligned} ||u + v||^2 &= \langle u + v, u + v\rangle = \langle u, u\rangle + \langle u, v\rangle + \langle v, u\rangle + \langle v, v\rangle = ||u||^2 + \langle u, v\rangle + \overline{\langle u, v\rangle} + ||v||^2 \\ &= ||u||^2 + 2\operatorname{Re}(\langle u, v\rangle) + ||v||^2. \quad \square \end{aligned}$$

Observar que si $\mathbb{k} = \mathbb{R}$, la relación (2.2) queda en

$$||u + v||^2 = ||u||^2 + 2\,\langle u, v\rangle + ||v||^2, \quad \forall u, v \in V. \tag{2.3}$$

Proposición 2.1.13. *Valen las siguientes propiedades.*

1. $||av|| = |a|\,||v||$, para todo $v \in V$ y $a \in \mathbb{k}$.

2. $||v|| \geq 0$ para todo $v \in V$, y $||v|| = 0$ si y solo si $v = 0$.

3. Desigualdad de Cauchy-Schwarz.

$$|\langle u,v\rangle| \leq ||u||\,||v||, \quad \forall u,v\in V.$$

Además vale la igualdad si y solo si $\{u,v\}$ es LD.

4. Desigualdad triangular.

$$||u+v|| \leq ||u|| + ||v||, \quad \forall u,v\in V.$$

Dem. La primera afirmación se deduce de lo siguiente

$$||av|| = \sqrt{\langle av, av\rangle} = \sqrt{a\overline{a}\langle v,v\rangle} = \sqrt{|a|^2\langle v,v\rangle} = |a|\,||v||.$$

La segunda afirmación es consecuencia directa de la definición de norma y de la proposición 2.1.7. Para probar la desigualdad de Cauchy-Schwarz, notar primero que si $v=0$, entonces $\{u,v\}=\{u,0\}$ es LD y

$$|\langle u,v\rangle| = |\langle u,0\rangle| = 0 = ||u||\,||0|| = ||u||\,||v||.$$

Supongamos ahora $v\neq 0$. Aplicando la fórmula (2.2) obtenemos

$$\begin{aligned}\left\|u-\frac{\langle u,v\rangle}{||v||^2}v\right\|^2 &= ||u||^2 - 2\operatorname{Re}\left(\left\langle u, \frac{\langle u,v\rangle}{||v||^2}v\right\rangle\right) + \left\|\frac{\langle u,v\rangle}{||v||^2}v\right\|^2 = ||u||^2 - 2\operatorname{Re}\left(\frac{\overline{\langle u,v\rangle}}{||v||^2}\langle u,v\rangle\right) + \frac{|\langle u,v\rangle|^2}{||v||^4}||v||^2 \\ &= ||u||^2 - 2\frac{|\langle u,v\rangle|^2}{||v||^2} + \frac{|\langle u,v\rangle|^2}{||v||^2} = ||u||^2 - \frac{|\langle u,v\rangle|^2}{||v||^2}.\end{aligned}$$

Luego

$$0 \leq \left\|u-\frac{\langle u,v\rangle}{||v||^2}v\right\|^2 = ||u||^2 - \frac{|\langle u,v\rangle|^2}{||v||^2}. \tag{2.4}$$

De (2.4) se obtiene inmediatamente la desigualdad de Cauchy-Schwarz.

Si $\{u,v\}$ es LD, como $v\neq 0$ entonces existe $a\in\Bbbk$ tal que $u=av$, luego

$$|\langle u,v\rangle| = |\langle av,v\rangle| = |a\langle v,v\rangle| = |a|\,||v||^2 = |a|\,||v||\,||v|| = ||av||\,||v|| = ||u||\,||v||.$$

Recíprocamente, si $|\langle u,v\rangle| = ||u||\,||v||$ y $v\neq 0$, entonces (2.4) implica $\left\|u-\frac{\langle u,v\rangle}{||v||^2}v\right\|^2 = 0$. Luego $u=\frac{\langle u,v\rangle}{||v||^2}v$ y por lo tanto $\{u,v\}$ es LD.

La desigualdad triangular se deduce de la desigualdad Cauchy-Schwarz:

$$\begin{aligned}||u+v||^2 &= ||u||^2 + 2\operatorname{Re}(\langle u,v\rangle) + ||v||^2 \leq ||u||^2 + 2|\langle u,v\rangle| + ||v||^2 \leq ||u||^2 + 2||u||\,||v|| + ||v||^2 \\ &= (||u||+||v||)^2 \quad \Rightarrow \quad ||u+v|| \leq ||u|| + ||v||. \quad \square\end{aligned}$$

Aplicación 2.1.1. *En el caso $V=\Bbbk^n$, la desigualdad de Cauchy-Schwarz y la desigualdad triangular nos dan las siguientes relaciones.*

$$\left|\sum_{i=1}^n x_i\overline{y_i}\right| \leq \sqrt{\sum_{i=1}^n |x_i|^2}\sqrt{\sum_{i=1}^n |y_i|^2}, \qquad \sqrt{\sum_{i=1}^n |x_i+y_i|^2} \leq \sqrt{\sum_{i=1}^n |x_i|^2} + \sqrt{\sum_{i=1}^n |y_i|^2}.$$

Si $\Bbbk=\mathbb{R}$, entonces las fórmulas quedan más simples

$$\left|\sum_{i=1}^n x_i y_i\right| \leq \sqrt{\sum_{i=1}^n x_i^2}\sqrt{\sum_{i=1}^n y_i^2}, \qquad \sqrt{\sum_{i=1}^n (x_i+y_i)^2} \leq \sqrt{\sum_{i=1}^n x_i^2} + \sqrt{\sum_{i=1}^n y_i^2}.$$

2.2. Ortogonalidad

En esta sección V es un espacio con producto interno.

Definición 2.2.1. Un *versor* es un vector de norma 1. Si dos vectores $u, v \in V$ verifican $\langle u, v\rangle = 0$, entonces decimos que son *ortogonales* y escribimos $u \perp v$.

Observaciones 2.2.2. Si v es un vector no nulo, entonces $\frac{v}{\|v\|} := \frac{1}{\|v\|}v$ es un versor colineal con v. El vector nulo es el único vector que es ortogonal a todos los vectores del espacio (proposición 2.1.7).

Definición 2.2.3. Un subconjunto S de V se dice *ortogonal* si para todo u, v en S, con $u \neq v$, es $u \perp v$; si además los elementos de S son versores, entonces decimos que S es un conjunto *ortonormal*. Observar que si $S = \{v_1, \dots, v_n\}$, entonces S es ortonormal si y solo si[1] $\langle v_i, v_j\rangle = \delta_{ij}$ para todo i, j. Una *base ortonormal* de V es un conjunto que es ortonormal y es base del espacio V.

Ejemplos 2.2.4.

1. La base canónica de $\Bbbk^n$ es una base ortonormal.
2. El conjunto $S = \{(1,1,0),\ (1,-1,1),\ (-1,1,2)\} \subset \mathbb{R}^3$ es ortogonal pero no es ortonormal. Calculando las normas de los elementos de S obtenemos $\|(1,1,0)\| = \sqrt{2}$, $\|(1,-1,1)\| = \sqrt{3}$, $\|(-1,1,2)\| = \sqrt{6}$. Luego $S' = \left\{\frac{1}{\sqrt{3}}(1,1,0),\ \frac{1}{\sqrt{2}}(1,-1,1),\ \frac{1}{\sqrt{6}}(-1,1,2)\right\} \subset \mathbb{R}^3$ es un conjunto ortonormal.

A continuación veremos algunas propiedades de los conjuntos ortonormales.

Proposición 2.2.5 (Teorema de Pitágoras)**.** *Si u y v son ortogonales, entonces $||u+v||^2 = ||u||^2 + ||v||^2$.*

Dem. Aplicar la fórmula (2.2). □

Corolario 2.2.6. *Si $\{v_1, \dots, v_n\}$ es un conjunto ortogonal, es $||\sum_{i=1}^n v_i||^2 = \sum_{i=1}^n ||v_i||^2$.*

Dem. Lo demostramos por inducción en n. Si $n = 2$ es el teorema de Pitágoras. Supongamos que vale para n y consideremos un conjunto ortogonal $\{v_1, \dots, v_n, v_{n+1}\}$. Como v_{n+1} es ortogonal a $v_1, \dots, v_n$ entonces resulta ortogonal a $\sum_{i=1}^n v_i$. Luego aplicando Pitágoras y la hipótesis de inducción obtenemos

$$\left|\left|\sum_{i=1}^{n+1} v_i\right|\right|^2 = \left|\left|\left(\sum_{i=1}^{n} v_i\right) + v_{n+1}\right|\right|^2 = \left|\left|\sum_{i=1}^{n} v_i\right|\right|^2 + ||v_{n+1}||^2 = \sum_{i=1}^{n} ||v_i||^2 + ||v_{n+1}||^2 = \sum_{i=1}^{n+1} ||v_i||^2. \quad \square$$

Proposición 2.2.7. *Sea $S = \{v_1, \dots, v_n\}$ un conjunto ortogonal de vectores no nulos y $u \in [v_1, \dots, v_n]$.*

$$\textit{Si } u = \sum_{i=1}^n a_i v_i, \ \textit{entonces } a_i = \frac{\langle u, v_i\rangle}{||v_i||^2}, \ \forall i = 1, \dots, n.$$

Dem.

$$u = \sum_{j=1}^n a_j v_j \quad \Rightarrow \quad \langle u, v_i\rangle = \left\langle \sum_{j=1}^n a_j v_j, v_i \right\rangle = \sum_{j=1}^n a_j \langle v_j, v_i\rangle = a_i ||v_i||^2 \quad \Rightarrow \quad a_i = \frac{\langle u, v_i\rangle}{||v_i||^2}. \quad \square$$

Corolario 2.2.8. *Si S es un subconjunto ortogonal (finito o infinito) de V formado por vectores no nulos, entonces S es LI.*

Dem. Si $v_1, \dots, v_n \in S$ y $a_1, \dots, a_n \in \Bbbk$ son tales que $\sum_{i=1}^n a_i v_i = 0$, entonces $a_i = \frac{\langle 0, v_i\rangle}{||v_i||^2} = 0$ para todo $i = 1, \dots, n$. □

[1] El símbolo δ_{ij} se llama la *delta de Kronecker* y está definido por $\delta_{ii} = 1$ y $\delta_{ij} = 0$, si $i \neq j$.

Observación 2.2.9. La proposición anterior implica que si V tiene dimensión finita y $\mathcal{B} = \{v_1, \ldots, v_n\}$ es una base ortonormal de V, entonces vale

$$u = \sum_{i=1}^{n} \langle u, v_i \rangle \, v_i, \quad \forall u \in V. \tag{2.5}$$

Definición 2.2.10. Sea $\mathcal{B}$ un conjunto (posiblemente infinito) ortonormal en V. Si $w \in V$, entonces los escalares $\langle w, v \rangle$, con $v \in \mathcal{B}$, se llaman los *coeficientes de Fourier* de w respecto a $\mathcal{B}$.

Teorema 2.2.11 (Método de Gram-Schmidt)**.** *Sea $\{v_1, \ldots, v_n\}$ un conjunto LI. Definimos $\{w_1, \ldots, w_n\}$ mediante*

$$\begin{aligned}
w_1 &= v_1 \\
w_2 &= v_2 - \frac{\langle v_2, w_1 \rangle}{||w_1||^2} w_1 \\
w_3 &= v_3 - \frac{\langle v_3, w_1 \rangle}{||w_1||^2} w_1 - \frac{\langle v_3, w_2 \rangle}{||w_2||^2} w_2 \\
&\vdots \\
w_n &= v_n - \frac{\langle v_n, w_1 \rangle}{||w_1||^2} w_1 - \cdots - \frac{\langle v_n, w_{n-1} \rangle}{||w_{n-1}||^2} w_{n-1}.
\end{aligned} \tag{2.6}$$

Entonces $\{w_1, \ldots, w_n\}$ es un conjunto ortogonal LI que verifica $[w_1, \ldots, w_n] = [v_1, \ldots, v_n]$.

Dem. Probaremos por inducción en n, que si $\{v_1, \ldots, v_n\}$ es un conjunto LI y definimos $w_1, \ldots, w_n$ como arriba, entonces $\{w_1, \ldots, w_n\}$ es un conjunto ortogonal LI que verifica $[w_1, \ldots, w_n] = [v_1, \ldots, v_n]$.

Para $n = 1$, es $w_1 = v_1$ y por lo tanto $\{w_1\} = \{v_1\}$ y $[w_1] = [v_1]$; luego la tesis se cumple trivialmente.

Sea ahora $\{v_1, \ldots, v_{n-1}, v_n\}$ un conjunto LI. El conjunto $\{v_1, \ldots, v_{n-1}\}$ es LI, luego la hipótesis de inducción nos dice que $\{w_1, \ldots, w_{n-1}\}$ es un conjunto ortogonal LI que verifica $[w_1, \ldots, w_{n-1}] = [v_1, \ldots, v_{n-1}]$.
Afirmación 1: el conjunto $\{w_1, \ldots, w_{n-1}, w_n\}$ es ortogonal. Ya sabemos que $\{w_1, \ldots, w_{n-1}\}$ es ortogonal, así que solo resta probar que w_n es ortogonal a $w_1, \ldots, w_{n-1}$. Sea $i \in \{1, \ldots, n-1\}$. Entonces usando (2.6) y que $\{w_1, \ldots, w_{n-1}\}$ es ortogonal, obtenemos

$$\begin{aligned}
\langle w_n, w_i \rangle &= \left\langle v_n - \frac{\langle v_n, w_1 \rangle}{||w_1||^2} w_1 - \cdots - \frac{\langle v_n, w_{n-1} \rangle}{||w_{n-1}||^2} w_{n-1},\, w_i \right\rangle \\
&= \langle v_n, w_i \rangle - \frac{\langle v_n, w_1 \rangle}{||w_1||^2} \langle w_1, w_i \rangle - \cdots - \frac{\langle v_n, w_i \rangle}{||w_i||^2} \langle w_i, w_i \rangle - \cdots - \frac{\langle v_n, w_{n-1} \rangle}{||w_{n-1}||^2} \langle w_{n-1}, w_i \rangle \\
&= \langle v_n, w_i \rangle - \frac{\langle v_n, w_i \rangle}{||w_i||^2} \langle w_i, w_i \rangle = \langle v_n, w_i \rangle - \langle v_n, w_i \rangle = 0.
\end{aligned}$$

Luego $\{w_1, \ldots, w_{n-1}, w_n\}$ es ortogonal.
Afirmación 2: los vectores $w_1, \ldots, w_{n-1}, w_n$ son no nulos. Como $\{w_1, \ldots, w_{n-1}\}$ es LI, entonces sabemos que $w_1, \ldots, w_{n-1}$ son no nulos. Si fuese $w_n = 0$, entonces usando (2.6) obtendríamos

$$v_n = \frac{\langle v_n, w_1 \rangle}{||w_1||^2} w_1 + \cdots + \frac{\langle v_n, w_{n-1} \rangle}{||w_{n-1}||^2} w_{n-1} \in [w_1, \ldots, w_{n-1}] = [v_1, \ldots, v_{n-1}].$$

Pero esto implica que $\{v_1, \ldots, v_n\}$ es un conjunto LD contradiciendo nuestra hipótesis. Luego es $w_n \neq 0$.
Juntando las dos afirmaciones anteriores, obtenemos que $\{w_1, \ldots, w_{n-1}, w_n\}$ es un conjunto ortogonal formado por vectores no nulos, luego es LI (corolario 2.2.8).

Para la última parte, observar primero que (2.6) implica $w_n \in [w_1, \ldots, w_{n-1}, v_n]$. Además, de la igualdad $[w_1, \ldots, w_{n-1}] = [v_1, \ldots, v_{n-1}]$, deducimos $w_1, \ldots, w_{n-1} \in [v_1, \ldots, v_{n-1}]$. Luego

$$w_n \in [w_1, \ldots, w_{n-1}, v_n] \subset [v_1, \ldots, v_{n-1}, v_n].$$

Usando esta relación y que $w_1, \ldots, w_{n-1} \in [v_1, \ldots, v_{n-1}]$, concluimos que vale $[w_1, \ldots, w_n] \subset [v_1, \ldots, v_n]$. Además sabemos que $\{v_1, \ldots, v_n\}$ y $\{w_1, \ldots, w_n\}$ son LI, luego $\dim[w_1, \ldots, w_n] = \dim[v_1, \ldots, v_n] = n$. Entonces vale $[w_1, \ldots, w_n] \subset [v_1, \ldots, v_n]$ y tienen la misma dimensión, luego $[w_1, \ldots, w_n] = [v_1, \ldots, v_n]$. $\square$

Observación 2.2.12. Si $S = \{v_1, \ldots, v_l, v_{l+1}, \ldots, v_n\}$ es un conjunto LI tal que $\{v_1, \ldots, v_l\}$ es un conjunto ortonormal, entonces aplicándole a S el método de Gram-Schmidt obtenemos $w_i = v_i$, para todo $i = 1, \ldots, l$.

Corolario 2.2.13. *Si la dimensión de V es finita, entonces V admite una base ortonormal.*

Dem. Sea $\mathcal{C} = \{v_1, \ldots, v_n\}$ una base cualquiera de V. Aplicando el método de Gram-Schmidt obtenemos una base ortogonal $\mathcal{C}' = \{u_1, \ldots, u_n\}$. Si definimos $w_i = u_i/||u_i||$, $i = 1 \ldots, n$, entonces $\mathcal{B} = \{w_1, \ldots, w_n\}$ es una base ortonormal de V. $\square$

Observación 2.2.14. Si W es un subespacio de V, entonces W es un espacio vectorial y la restrición del producto interno de V a W le da a W una estructura de espacio vectorial con producto interno.

Ejemplo 2.2.15. Consideremos el plano $W = \{(x, y, z) \in \mathbb{R}^3 : \ 2x+3y-z = 0\}$, veremos de hallar una base ortonormal de W. Empezamos por encontrar una base cualquiera de W, por ejemplo $\mathcal{B}_1 = \{(1, 0, 2),\ (0, 1, 3)\}$. A esta base le aplicamos el proceso de Gram-Schmidt.

$$\begin{aligned} w_1 &= (1, 0, 2), \\ w_2 &= (0, 1, 3) - \frac{\langle (0, 1, 3), (1, 0, 2)\rangle}{\|(1, 0, 2)\|^2}(1, 0, 2) = (0, 1, 3) - \frac{6}{5}(1, 0, 2) = \left(-\frac{6}{5}, 1, \frac{3}{5}\right). \end{aligned}$$

Luego $\mathcal{B}_2 = \left\{(1, 0, 2), \left(-\frac{6}{5}, 1, \frac{3}{5}\right)\right\}$ es una base ortogonal de W. Notar que es $\left(-\frac{6}{5}, 1, \frac{3}{5}\right) = \frac{1}{5}(-6, 5, 3)$. Luego $\mathcal{B}_3 = \{(1, 0, 2),\ (-6, 5, 3)\}$ también es una base ortogonal de W. Finalmente normalizando los vectores de $\mathcal{B}_3$ obtenemos que $\mathcal{B}_4 = \left\{\frac{1}{\sqrt{5}}(1, 0, 2),\ \frac{1}{\sqrt{70}}(-6, 5, 3)\right\}$ es una base ortonormal de W.

Observación 2.2.16. Supongamos que la dimensión de V es finita y que $\mathcal{B} = \{v_1, \ldots, v_n\}$ es una base ortonormal de V. Sean $v = \sum_{i=1}^n x_i v_i$ y $w = \sum_{i=1}^n y_i v_i$, entonces

$$\langle v, w\rangle = \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j v_j \right\rangle = \sum_{i,j=1}^n x_i \overline{y_j}\, \langle v_i, v_j\rangle = \sum_{i,j=1}^n x_i \overline{y_j} \delta_{ij} = \sum_{i=1}^n x_i \overline{y_i}.$$

Luego

$$\langle v, w\rangle = \sum_{i=1}^n x_i \overline{y_i}, \quad \|v\|^2 = \sum_{i=1}^n |x_i|^2.$$

Notar la analogía con el caso de $\mathbb{k}^n$ con el producto interno usual.

De la observación anterior y la fórmula (2.5) se obtiene lo siguiente.

Proposición 2.2.17 (Identidad de Parseval)**.** *Sea V un espacio de dimensión finita y $\mathcal{B} = \{v_1, \ldots, v_n\}$ una base ortonormal de V. Entonces*

$$\langle v, w\rangle = \sum_{i=1}^n \langle v, v_i\rangle\, \overline{\langle w, v_i\rangle}, \quad \forall v, w \in V.$$

En particular, $\|v\|^2 = \sum_{i=1}^n |\langle v, v_i\rangle|^2$. $\square$

El complemento ortogonal.

Definición 2.2.18. Si S es un subconjunto cualquiera de V, el conjunto

$$S^\perp := \{v \in V : \ v \perp w, \ \forall w \in S\}$$

se llama el *complemento ortogonal* de S.

Ejemplo 2.2.19. Valen $\{0\}^\perp = V$ y $V^\perp = \{0\}$.

Observaciones 2.2.20. Los siguientes resultados son fáciles de probar.

1. El conjunto $S^\perp$ siempre es un subespacio de V (aunque S no lo sea).

2. Si $S_1 \subset S_2$, entonces $S_2^\perp \subset S_1^\perp$.

Proposición 2.2.21. *Si S es un subconjunto cualquiera de V, entonces $S^\perp = [S]^\perp$.*

Dem. Observar que que $S \subset [S]$, implica $[S]^\perp \subset S^\perp$; así que solo necesitamos probar la otra inclusión. Sea $v \in S^\perp$. Si $w \in [S]$, entonces existen $w_1, \ldots, w_n \in S$ y $a_1, \ldots, a_n \in \Bbbk$ tales que $w = \sum_{i=1}^n a_i w_i$, luego $\langle w, v\rangle = \langle \sum_{i=1}^n a_i w_i, v\rangle = \sum_{i=1}^n a_i \langle w_i, v\rangle = \sum_{i=1}^n a_i 0 = 0$. Como esto vale para todo $w \in [S]$, concluimos $v \in [S]^\perp$. Luego $S^\perp \subset [S]^\perp$. $\square$

Observación 2.2.22. Este último resultado implica que si W es un subespacio de V y $\{w_1, \ldots, w_n\}$ es un generador de W, entonces $v \in W^\perp$ si y solo si $v \perp w_i$ para todo $i = 1, \ldots, n$.

Teorema 2.2.23. *Si W es un subespacio de dimensión finita de V, entonces $V = W \oplus W^\perp$.*

Dem. Sea $\mathcal{B} = \{w_1, \ldots, w_n\}$ una base ortonormal de W, si $v \in V$ es

$$v = \sum_{i=1}^n \langle v, w_i\rangle w_i + v - \sum_{i=1}^n \langle v, w_i\rangle w_i.$$

Observar que $\langle v - \sum_{i=1}^n \langle v, w_i\rangle w_i, w_j\rangle = \langle v, w_j\rangle - \sum_{i=1}^n \langle v, w_i\rangle \langle w_i, w_j\rangle = 0$ para todo $j = 1, \ldots, n$, luego por la observación anterior es $v - \sum_{i=1}^n \langle v, w_i\rangle w_i \in W^\perp$ y claramente $\sum_{i=1}^n \langle v, w_i\rangle w_i \in W$, de donde $V = W + W^\perp$. Por otro lado si $w \in W \cap W^\perp$ es $w \perp w$ y luego $w = 0$; esto prueba $W \cap W^\perp = \{0\}$. $\square$

Corolario 2.2.24. *Si la dimensión de V es finita y $W \subset V$ es un subespacio, entonces*

$$\dim W + \dim W^\perp = \dim V. \quad \square$$

Definición 2.2.25. Sea W un subespacio de dimensión finita de V. Sean $P_W \in \mathcal{L}(V)$ y $P_{W^\perp} \in \mathcal{L}(V)$ las proyecciones asociadas a la descomposición $V = W \oplus W^\perp$, siendo $\operatorname{Im} P_W = W$ y $\operatorname{Im} P_{W^\perp} = W^\perp$. La proyección P_W se llama la *proyección ortogonal* sobre el espacio W.
De la prueba del teorema anterior se deduce que si $\mathcal{B} = \{w_1, \ldots, w_n\}$ es una base ortonormal de W, entonces

$$P_W(v) = \sum_{i=1}^n \langle v, w_i\rangle w_i, \quad P_{W^\perp}(v) = v - \sum_{i=1}^n \langle v, w_i\rangle w_i, \quad \forall v \in V.$$

Ejemplo 2.2.26. Sea $W = \{(x, y, z, t) \in \mathbb{R}^4 : \ x - y + z = 0 \text{ y } 2y - z + t = 0\}$ y consideramos $\mathbb{R}^4$ con el producto escalar. Veamos de hallar $W^\perp$. Lo primero es encontrar una base cualquiera de W, por ejemplo $\mathcal{B} = \{(-1, 0, 1, 1), (1, 1, 0, -2)\}$. Sabemos que es $W^\perp = \mathcal{B}^\perp$. Luego

$(x, y, z, t) \in W^\perp \Leftrightarrow \langle (x, y, z, t), (1, 1, 0, -2)\rangle = \langle (x, y, z, t), (-1, 0, 1, 1)\rangle = 0 \Leftrightarrow x + y - 2t = 0$ y $-x + z + t = 0$, luego $W^\perp = \{(x,y,z,t) \in \mathbb{R}^4 : \ x+y-2t = 0 \text{ y } -x+z+t = 0\}$. Vamos a hallar la proyección P_W. Aplicando Gram-Schmidt a la base $\mathcal{B}$ y normalizando obtenemos $\mathcal{B}_1 = \left\{\frac{1}{\sqrt{3}}(-1,0,1,1), \frac{1}{\sqrt{3}}(0,1,1,-1)\right\}$ base ortonormal de W. Luego aplicando la fórmula de arriba obtenemos

$$\begin{aligned} P_W(x,y,z,t) &= \frac{1}{3}(-x+z+t)(-1,0,1,1) + \frac{1}{3}(y+z-t)(0,1,1,-1) \\ &= \frac{1}{3}(x-z-t,\, y+z-t,\, -x+y+2z,\, -x-y+2t), \quad \forall (x,y,z,t) \in \mathbb{R}^4. \end{aligned}$$

Corolario 2.2.27. *Sea W un subespacio de dimensión finita de V, $P_W \in \mathcal{L}(V)$ la proyección ortogonal sobre W y $v \in V$. Entonces:*

1. *Para todo $w \in W$ es*
$$||v-w|| \geq ||v - P_W(v)||. \tag{2.7}$$
2. *Si $w_0 \in W$ verifica*
$$||v-w|| \geq ||v-w_0||, \tag{2.8}$$
para todo $w \in W$, entonces $w_0 = P_W(v)$.

Dem. Observar que si $w \in W$, es $v - P_W(v) = P_{W^\perp}(v) \in W^\perp$ y $P_W(v) - w \in W$, luego aplicando el teorema de Pitágoras resulta

$$||v-w||^2 = ||v - P_W(v) + P_W(v) - w||^2 = ||v-P_W(v)||^2 + ||P_W(v)-w||^2 \geq ||v-P_W(v)||^2. \tag{2.9}$$

Esto implica la primera afirmación. Para la segunda afirmación, tomando $w = P_W(v)$ en (2.8) resulta $||v-P_W(v)|| \geq ||v-w_0||$. Por otro lado tomando $w = w_0$ en (2.7) resulta $||v-w_0|| \geq ||v-P_W(v)||$. Luego $||v-w_0|| = ||v-P_W(v)||$. Teniendo en cuenta esta última relación, tomando $w = w_0$ en (2.9) deducimos $||P_W(v) - w_0|| = 0$ y luego $P_W(v) = w_0$. □

Definición 2.2.28.
1. S $v, w \in V$, entonces definimos la *distancia* entre v y w por $d(v,w) := \|v-w\|$.
2. Sea $S \subset V$ un conjunto no vacío y $v \in V$. Definimos la *distancia* de v a S por
$$d(v,S) := \inf\{d(v,w) : \ w \in S\}.$$

Observación 2.2.29. El corolario 2.2.27 se puede interpretar en términos de distancias. La relación (2.7) es

$$d(v,w) \geq d(v, P_W(v)), \quad \forall w \in W.$$

Esto implica que $d(v,P_W(v))$ es una cota inferior para $\{d(v,w) : \ w \in W\}$, pero como $P_W(v) \in W$, resulta

$$d(v,P_W(v)) = \min\{d(v,w) : \ w \in W\} = d(v,W).$$

La parte 2 del corolario muestra que $P_W(v)$ es el único elemento de W que realiza la distancia de v a W.

Corolario 2.2.30 (Desigualdad de Bessel). *Sea $\{v_1,\ldots,v_n\}$ un subconjunto ortonormal de V, entonces*

$$||v||^2 \geq \sum_{i=1}^n |\langle v, v_i\rangle|^2, \quad \forall v \in V.$$

Dem. Sea $W = [v_1,\ldots,v_n]$. El teorema 2.2.23 implica $V = W \oplus W^\perp$. Si $v \in V$, entonces aplicando Pitágoras obtenemos

$$||v||^2 = ||P_W(v) + v - P_W(v)||^2 = ||P_W(v)||^2 + ||v-P_W(v)||^2 \geq ||P_W(v)||^2 = \sum_{i=1}^n |\langle v, v_i\rangle|^2. \quad \square$$

El teorema de Riesz. Recordar que el espacio dual de V es $V^* = \mathcal{L}(V, \Bbbk)$. Si $w \in V$, entonces el mapa $\alpha : V \to \Bbbk$ definido por $\alpha(v) = \langle v, w \rangle$, para todo $v \in V$, es una transformación lineal, luego $\alpha \in V^*$.

Ejemplo 2.2.31. Si $\alpha \in (\mathbb{R}^n)^*$, entonces existen $a_1, \ldots, a_n \in \mathbb{R}$ tales que $\alpha(x_1, \ldots, x_n) = a_1 x_1 + \cdots + a_n x_n$, para todo $(x_1, \ldots, x_n) \in \mathbb{R}^n$. Luego considerando el producto interno usual (el escalar) en $\mathbb{R}^n$, vemos que podemos escribir α de la forma

$$\alpha(x_1, \ldots, x_n) = \langle (x_1, \ldots, x_n), (a_1, \ldots, a_n) \rangle, \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n.$$

Si ahora en vez de $\mathbb{R}^n$ consideramos $\mathbb{C}^n$, entonces manteniendo las notaciones anteriores obtenemos

$$\alpha(x_1, \ldots, x_n) = a_1 x_1 + \cdots + a_n x_n = \langle (x_1, \ldots, x_n), (\overline{a_1}, \ldots, \overline{a_n}) \rangle, \quad \forall (x_1, \ldots, x_n) \in \mathbb{C}^n.$$

Luego todo $\alpha \in (\Bbbk^n)^*$ (siendo $\Bbbk = \mathbb{R}$ o $\mathbb{C}$) es de la forma $\alpha(v) = \langle v, w \rangle$, para cierto $w \in \Bbbk^n$.

El siguiente resultado muestra que si estamos en un espacio con producto interno de dimensión finita, entonces siempre sucede lo anterior.

Teorema 2.2.32 (Riesz)**.** *Si V tiene dimensión finita y $\alpha \in V^*$, entonces existe un único $w \in V$ tal que*

$$\alpha(v) = \langle v, w \rangle, \quad \forall v \in V. \tag{2.10}$$

Dem. Notar que la unicidad de w se deduce de la parte 3 de la proposición 2.1.7, así que solo tenemos que probar su existencia. Si es $\alpha = 0$, entonces es claro que $w = 0$ verifica (2.10). Supongamos ahora que es $\alpha \neq 0$. Como es $\alpha : V \to \Bbbk$, entonces $\dim \operatorname{Ker}(\alpha) = n - 1$, siendo $n = \dim V$. Esto implica $\dim \operatorname{Ker}(\alpha)^\perp = 1$. Elegimos $0 \neq w_0 \in \operatorname{Ker}(\alpha)^\perp$, luego $\operatorname{Ker}(\alpha)^\perp = [w_0]$. Buscamos encontrar $w \in \operatorname{Ker}(\alpha)^\perp$ tal que $\alpha(w_0) = \langle w_0, w \rangle$. Sea $w = a w_0$ $a \in \Bbbk$. Entonces

$$\alpha(w_0) = \langle w_0, w \rangle \quad \Leftrightarrow \quad \alpha(w_0) = \langle w_0, a w_0 \rangle \quad \Leftrightarrow \quad \alpha(w_0) = \overline{a} \|w_0\|^2 \quad \Leftrightarrow \quad a = \frac{\overline{\alpha(w_0)}}{\|w_0\|^2}.$$

Luego si consideramos $w = \frac{\overline{\alpha(w_0)}}{\|w_0\|^2} w_0$, entonces vale $\alpha(w_0) = \langle w_0, w \rangle$. Probaremos que w verifica (2.10). Sea $v \in V$. Como es $V = \operatorname{Ker}(\alpha) \oplus \operatorname{Ker}(\alpha)^\perp$, con $\operatorname{Ker}(\alpha)^\perp = [w_0]$, entonces existe $v_1 \in \operatorname{Ker}(\alpha)$ y $b \in \Bbbk$ tales que $v = v_1 + b w_0$. Luego

$$\alpha(v) = \alpha(v_1 + b w_0) = \alpha(v_1) + b\alpha(w_0) = b\alpha(w_0); \quad \langle v, w \rangle = \langle v_1 + b w_0, w \rangle = \langle v_1, w \rangle + b \langle w_0, w \rangle = b \langle w_0, w \rangle.$$

En la última igualdad usamos que es $v_1 \in \operatorname{Ker}(\alpha)$ y $w \in \operatorname{Ker}(\alpha)^\perp$. Luego $\alpha(v) = \langle v, w \rangle$, para todo $v \in V$. $\square$

Ejemplo 2.2.33. Sea $V = \mathbb{R}_1[x] = \{a + bx : \ a, b \in \mathbb{R}\}$. Es un ejercicio el probar que $\langle p, q \rangle = \int_0^1 pq$ define un producto interno en V. Consideremos $\alpha \in V^*$ definido por $\alpha(a + bx) = a + b$, para todo $a + bx \in V$. Buscamos encontrar $w \in V$ que verifique (2.10). Para eso, primero observar que vale $\operatorname{Ker}(\alpha) = [1 - x]$. Operando obtenemos

$$\langle a + bx,\, 1 - x \rangle = \int_0^1 (a + bx)(1 - x)\, dx = \frac{3a + b}{6} \quad \Rightarrow \quad \operatorname{Ker}(\alpha)^\perp = \{1 - x\}^\perp = [1 - 3x].$$

Finalmente usando la fórmula $w = \frac{\alpha(w_0)}{\|w_0\|^2} w_0$, obtenemos $w = \frac{1-3}{1}(1 - 3x)$. Luego $w = -2(1 - 3x)$.

Observación 2.2.34. Si $\mathcal{B} = \{v_1, \ldots, v_n\}$ es una base ortonormal de V, entonces vale $\langle v_i, v_j \rangle = \delta_{ij}$, para todo i, j. Luego si definimos $\alpha_i \in V^*$ mediante $\alpha_i(v) = \langle v, v_i \rangle$, para todo $v \in V$, entonces es $\alpha_i(v_j) = \delta_{ij}$, para todo i, j. Luego la base dual de $\mathcal{B}$ es $\{\alpha_1, \ldots, \alpha_n\}$.

Capítulo 3

Operadores en espacios con producto interno

En este capítulo, V será siempre un espacio vectorial (real o complejo) con producto interno y de dimensión finita. Nuestro objetivo es estudiar las propiedades de los operadores en V. Como sabemos, los operadores diagonalizables tienen buenas propiedades y también las tienen las bases ortonormales. Por eso primero nos concentraremos en los operadores que son diagonalizables en bases ortonormales. Luego estudiaremos los operadores que preservan el producto interno. El siguiente resultado será usado con frecuencia.

Proposición 3.0.1. *Sean $T \in \mathcal{L}(V)$, $\mathcal{B} = \{v_1, \dots, v_n\}$ una base ortonormal de V y $[T]_\mathcal{B} = (a_{ij})$. Entonces $a_{ij} = \langle T(v_j), v_i \rangle$, para todo $i, j = 1, \dots, n$.*

Dem. Si $[T]_\mathcal{B} = (a_{ij})$ es $T(v_i) = \sum_{j=1}^n a_{ji} v_j$, para todo $i = 1, \dots, n$. Como $\mathcal{B}$ es ortonormal, entonces aplicando la fórmula (2.5) sabemos que también vale $T(v_i) = \sum_{j=1}^n \langle T(v_i), v_j \rangle v_j$. Luego $a_{ji} = \langle T(v_i), v_j \rangle$. $\square$

3.1. Operadores autoadjuntos

Empezamos definiendo un concepto que veremos que está directamente relacionado con la diagonalización de operadores en bases ortonormales.

Definición 3.1.1. Un operador $T \in \mathcal{L}(V)$ es *autoadjunto*[1] si verifica

$$\langle T(u), v \rangle = \langle u, T(v) \rangle, \quad \forall u, v \in V.$$

Proposición 3.1.2. *Sea $T \in \mathcal{L}(V)$ y $\mathcal{B} = \{v_1, \dots, v_n\}$ una base de V. Entonces T es autoadjunto si y solo si verifica*

$$\langle T(v_i), v_j \rangle = \langle v_i, T(v_j) \rangle, \quad \forall i, j = 1, \dots, n. \tag{3.1}$$

Dem. Es claro que vale el directo, luego lo único que hay que probar es el recíproco. Si $u, v \in V$ son vectores arbitrarios, y escribimos $u = \sum_{i=1}^n x_i v_i$ y $v = \sum_{j=1}^n y_j v_j$, entonces

$$\langle T(u), v \rangle = \left\langle \sum_{i=1}^n x_i T(v_i), \sum_{j=1}^n y_j v_j \right\rangle = \sum_{i=j=1}^n x_i \overline{y_j} \langle T(v_i), v_j \rangle,$$
$$\langle u, T(v) \rangle = \left\langle \sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j T(v_j) \right\rangle = \sum_{i=j=1}^n x_i \overline{y_j} \langle v_i, T(v_j) \rangle.$$

[1] A los operadores autoadjuntos también se les llama operadores *simétricos* en el caso real y *hermitianos* en el caso complejo, pero nosotros solo usaremos autoadjunto para ambos. El porqué llamarles “autoadjuntos” va a quedar claro en la próxima sección.

Luego (3.1) implica que T es autoadjunto. $\square$

Definición 3.1.3. Una matriz $A = (a_{ij}) \in M_n(\mathbb{C})$ se dice que es *hermitiana* si verifica $a_{ij} = \overline{a_{ji}}$, para todo i, j. Notar que las entradas diagonales de una matriz hermitiana siempre son números reales.

Observación 3.1.4. Si $A = (a_{ij}) \in M_n(\mathbb{R}) \subset M_n(\mathbb{C})$, entonces la condición de ser hermitiana queda en $a_{ij} = a_{ji}$, para todo i, j. Luego una matriz real es hermitiana si y solo si es simétrica.
Notar que las matrices hermitianas coinciden con las simétricas solo en el caso real. Por ejemplo, si consideramos $A = \left(\begin{smallmatrix} 3 & 1+2i \\ 1-2i & 4 \end{smallmatrix}\right)$ y $B = \left(\begin{smallmatrix} 3 & 1+2i \\ 1+2i & 4 \end{smallmatrix}\right)$, entonces A es hermitiana pero no es simétrica, mientras que B es simétrica pero no es hermitiana.

La proposición siguiente permite determinar los operadores autoadjuntos de $\Bbbk^n$.

Proposición 3.1.5. *Sea $T \in \mathcal{L}(V)$, $\mathcal{B}$ base ortonormal de V y $A = [T]_{\mathcal{B}}$. Entonces T es autoadjunto si y solo si A es simétrica en el caso real o hermitiana en el caso complejo.*

Dem. Sea $\mathcal{B} = \{v_1, \dots, v_n\}$. Si $A = (a_{ij})$, entonces la proposición 3.0.1 implica que es $a_{ji} = \langle T(v_i), v_j \rangle$. Luego vale $a_{ij} = \overline{a_{ji}}$, para todo i, j, si y solo si vale $\langle T(v_j), v_i \rangle = \overline{\langle T(v_i), v_j \rangle}$, para todo i, j, lo cual equivale a la fórmula (3.1). $\square$

Corolario 3.1.6. *Sean $A \in M_n(\Bbbk)$ y el operador asociado $L_A \in \mathcal{L}(\Bbbk^n)$. Si en $\Bbbk^n$ consideramos el producto interno usual, entonces $L_A : \Bbbk^n \to \Bbbk^n$ es autoadjunto, es decir vale*

$$\langle Ax, y \rangle = \langle x, Ay \rangle, \quad \forall x, y \in \Bbbk^n,$$

si y solo si A es simétrica en el caso real o hermitiana en el caso complejo.

Dem. La base canónica $\mathcal{B}$ de $\Bbbk^n$ es ortonormal y $A = [L_A]_{\mathcal{B}}$. $\square$

Proposición 3.1.7. *Sea $T \in \mathcal{L}(V)$. Si $\Bbbk = \mathbb{R}$ y T es diagonalizable en una base ortonormal o $\Bbbk = \mathbb{C}$ y T es diagonalizable en una base ortonormal con valores propios reales, entonces T es autoadjunto.*

Dem. En el caso real, si existe una base ortonormal $\mathcal{B}$ de V tal que $[T]_{\mathcal{B}}$ es una matriz diagonal, entonces $[T]_{\mathcal{B}}$ es simétrica (real) y por lo tanto T es autoadjunto.
En el caso complejo pasa lo mismo, porque como $[T]_{\mathcal{B}}$ es una matriz diagonal y sus entradas diagonales son los valores propios de T (que son reales), entonces $[T]_{\mathcal{B}}$ es una matriz simétrica real y por lo tanto es una matriz hermitiana compleja, lo cual implica que T es autoadjunto. $\square$

Observación 3.1.8. En el caso real, el ser autoadjunto es una condición necesaria para que el operador sea diagonalizable en una base ortonormal. En el caso complejo no lo es, como lo muestra el siguiente ejemplo.

Ejemplo 3.1.9. Sea $A = \left(\begin{smallmatrix} i & 0 \\ 0 & 1 \end{smallmatrix}\right) \in M_2(\mathbb{C})$. Considerando $\mathbb{C}^2$ con el producto interno usual, el operador $L_A : \mathbb{C}^2 \to \mathbb{C}^2$ es diagonalizable en una base ortonormal (la canónica), pero como A no es hermitiana, deducimos que L_A no es autoadjunto.

Observación 3.1.10. El siguiente resultado muestra que los operadores autoadjuntos solo pueden tener valores propios reales.

Proposición 3.1.11. *Si $T \in \mathcal{L}(V)$ es autoadjunto y $\lambda \in \mathbb{C}$ es un valor propio de T, entonces $\lambda \in \mathbb{R}$.*

Dem. Sea $v \in V$ un vector propio correspondiente a λ. Entonces

$$\lambda \|v\|^2 = \langle \lambda v, v \rangle = \langle T(v), v \rangle = \langle v, T(v) \rangle = \langle v, \lambda v \rangle = \overline{\lambda} \|v\|^2.$$

Como es $\|v\| \neq 0$, deducimos $\overline{\lambda} = \lambda$; luego $\lambda \in \mathbb{R}$. $\square$

A continuación probaremos el recíproco de la proposición 3.1.7. Para eso necesitamos el siguiente teorema. Veremos dos pruebas del mismo. La primera usa diferenciación en varias variables; en particular requiere lo siguiente.

Lema 3.1.12. *Sean $\alpha, \beta : \mathbb{R} \to \mathbb{R}^n$ funciones diferenciables y $A \in M_n(\mathbb{R})$. Definimos $\gamma : \mathbb{R} \to \mathbb{R}^n$ mediante $\gamma(t) = A\alpha(t)$, para todo $t \in \mathbb{R}$. Sea $\langle\ ,\ \rangle$ el producto escalar de $\mathbb{R}^n$. Entonces*

$$\gamma'(t) = A\alpha'(t), \quad \langle \alpha(t), \beta(t)\rangle' = \langle \alpha'(t), \beta(t)\rangle + \langle \alpha(t), \beta'(t)\rangle, \quad \forall t \in \mathbb{R}.$$

Dem. La prueba consiste simplemente en hacer el cálculo correspondiente. Sean $\alpha(t) = \big(\alpha_1(t), \ldots, \alpha_n(t)\big)$ y $\beta(t) = \big(\beta_1(t), \ldots, \beta_n(t)\big)$. Que α y β sean diferenciables quiere decir que las funciones $\alpha_i, \beta_j : \mathbb{R} \to \mathbb{R}$ son derivables, para todo i, j. Si escribimos $A = (a_{ij})$ y $\gamma(t) = \big(\gamma_1(t), \ldots, \gamma_n(t)\big)$, entonces $\gamma(t) = A\alpha(t)$ equivale a $\gamma_i(t) = a_{i1}\alpha_1(t) + \cdots + a_{in}\alpha_n(t)$, para todo i. Luego $\gamma_i'(t) = a_{i1}\alpha_1'(t) + \cdots + a_{in}\alpha_n'(t)$, para todo i. Esto equivale a $\gamma'(t) = A\alpha'(t)$.
Para la segunda afirmación, es $\langle \alpha(t), \beta(t)\rangle = \big\langle \big(\alpha_1(t), \ldots, \alpha_n(t)\big), \big(\beta_1(t), \ldots, \beta_n(t)\big)\big\rangle = \sum_{i=1}^n \alpha_i(t)\beta_i(t)$. Luego

$$\langle \alpha(t), \beta(t)\rangle' = \sum_{i=1}^n \big(\alpha_i(t)\beta_i(t)\big)' = \sum_{i=1}^n \alpha_i'(t)\beta_i(t) + \sum_{i=1}^n \alpha_i(t)\beta_i'(t) = \langle \alpha'(t), \beta(t)\rangle + \langle \alpha(t), \beta'(t)\rangle. \quad \square$$

Teorema 3.1.13. *Si $A \in M_n(\mathbb{R})$ es simétrica, entonces A tiene algún valor propio real.*

Dem. Veremos dos pruebas.

Prueba 1 (sin usar números complejos). Sea $\langle\ ,\ \rangle$ el producto interno usual (escalar) de $\mathbb{R}^n$. Definimos $f : \mathbb{R}^n \to \mathbb{R}$ mediante $f(x) = \langle x, Ax\rangle$, para todo $x \in \mathbb{R}^n$. Observar que si $A = (a_{ij})$, entonces

$$f(x_1, \ldots, x_n) = \sum_{i=j=1}^n a_{ij} x_i x_j, \quad \forall (x_1, \ldots, x_n) \in \mathbb{R}^n,$$

luego f es una función polinomial y por lo tanto es continua.
Sea $S = \{x \in \mathbb{R}^n : \ \|x\| = 1\}$. Como S es compacto y f es continua, entonces el teorema de Weierstrass nos dice que f tiene un mínimo en S, es decir, existe v_0 en S tal que $f(v_0) \leq f(x)$, para todo x en S. Notar que al ser $\|v_0\| = 1$ se tiene $v_0 \neq 0$.
Sea w un vector arbitrario de $\mathbb{R}^n$ tal que $w \perp v_0$ y $\|w\| = 1$. Definimos $\alpha : \mathbb{R} \to \mathbb{R}^n$ mediante

$$\alpha(t) = \cos(t)\, v_0 + \operatorname{sen}(t)\, w, \qquad \forall t \in \mathbb{R}.$$

Observar que al ser $\|v_0\| = \|w\| = 1$ y $v_0 \perp w$, el teorema de Pitágoras implica $\|\alpha(t)\|^2 = \cos^2 t + \operatorname{sen}^2 t = 1$, luego $\alpha(t) \in S$ y por lo tanto $f(v_0) \leq f\big(\alpha(t)\big)$, para todo t en $\mathbb{R}$.
Sea $\beta = f \circ \alpha : \mathbb{R} \to \mathbb{R}$. Por lo anterior es $\beta(0) \leq \beta(t)$, para todo t en $\mathbb{R}$; esto nos dice que β tiene un mínimo absoluto en 0, y como β es derivable, deducimos $\beta'(0) = 0$. Para calcular $\beta'(0)$, empezamos observando que vale $\alpha'(t) = -\operatorname{sen}(t)\, v_0 + \cos(t)\, w$ y por lo tanto $\alpha(0) = v_0$ y $\alpha'(0) = w$. Por otro lado es $\beta(t) = \langle \alpha(t), A\alpha(t)\rangle$, luego $\beta'(t) = \langle \alpha'(t), A\alpha(t)\rangle + \langle \alpha(t), A\alpha'(t)\rangle$. Sustituyendo t por 0 y usando que A es simétrica obtenemos

$$0 = \beta'(0) = \langle w, Av_0\rangle + \langle v_0, Aw\rangle = \langle w, Av_0\rangle + \langle Av_0, w\rangle = 2\langle w, Av_0\rangle.$$

Luego Av_0 es ortogonal con w, para todo w en $\mathbb{R}^n$ tal que $w \perp v_0$ y $\|w\| = 1$. Pero es fácil de probar que esto implica que Av_0 es ortogonal con w, para todo w en $\mathbb{R}^n$ tal que $w \perp v_0$. Esto implica $Av_0 \in \{v_0\}^{\perp\perp} = [v_0]$ y por lo tanto existe λ en $\mathbb{R}$ tal que $Av_0 = \lambda v_0$.

Prueba 2 (usando números complejos). Consideremos $L_A : \mathbb{C}^n \to \mathbb{C}^n$. Como A es simétrica real, entonces A pensada en $M_n(\mathbb{C})$ es hermitiana y por lo tanto $L_A : \mathbb{C}^n \to \mathbb{C}^n$ es autoadjunto. Como estamos en $\mathbb{C}$ sabemos que L_A tiene algún valor propio λ y al ser L_A autoadjunto deducimos que es $\lambda \in \mathbb{R}$. Como los valores propios de L_A coinciden con los de A, concluimos que A tiene algún valor propio en $\mathbb{R}$. $\square$

Corolario 3.1.14. *Si $T \in \mathcal{L}(V)$ es autoadjunto, entonces T tiene algún valor propio.*

Dem. Si $\Bbbk = \mathbb{C}$ el resultado es inmediato, dado que el polinomio característico de T siempre tiene alguna raíz, la cual es un valor propio de T (esto no requiere que T sea autoadjunto). Supongamos ahora que es $\Bbbk = \mathbb{R}$. Sea $\mathcal{B}$ una base ortonormal de V. Como $T \in \mathcal{L}(V)$ es autoadjunto, entonces $[T]_{\mathcal{B}}$ es una matriz simétrica real; luego el teorema anterior implica que $[T]_{\mathcal{B}}$ tiene algún valor propio (real) y por lo tanto T tiene algún valor propio. □

Observación 3.1.15. Si $T \in \mathcal{L}(V)$, un subespacio $W \subset V$ se dice T-*invariante* si verifica $T(w) \in W$, para todo $w \in W$. En ese caso $T|_W : W \to W$ definida por $(T|_W)(w) = T(w)$, para todo $w \in W$ es un operador en W. Notar que si $T \in \mathcal{L}(V)$ es autoadjunto y $W \subset V$ es un subespacio T-invariante, entonces $T|_W \in \mathcal{L}(W)$ también es autoadjunto.

Proposición 3.1.16. *Si $T \in \mathcal{L}(V)$ es autoadjunto, entonces T es diagonalizable en una base ortonormal.*

Dem. La prueba es por inducción en $n = \dim V$. Para $n = 1$ la prueba es trivial, porque en ese caso todo vector no nulo es un vector propio[2] de T. Luego tomando $v \in V$ con $\|v\| = 1$, obtenemos que $\{v\}$ es una base ortonormal de V formada por vectores propios de T.

Sea ahora $n > 1$ y supongamos razonando inductivamente que si tenemos un operador autoadjunto en un espacio de dimensión $n-1$, entonces existe una base ortonormal del espacio formada por vectores propios del operador.
Como T es autoadjunto, entonces el corolario anterior implica que existe λ valor propio de T. Sea v un vector propio de T correspondiente a λ y consideremos $W := [v]^\perp = \{w \in V \mid \langle w, v\rangle = 0\}$. Sabemos que W es un subespacio de V de dimensión $n-1$. Sea $w \in W$,

$$\langle T(w), v\rangle = \langle w, T(v)\rangle = \langle w, \lambda v\rangle = \lambda\langle w, v\rangle = \lambda 0 = 0 \quad \Rightarrow \quad T(w) \in W.$$

Luego W es T-invariante y como T es autoadjunto se deduce que $T|_W \in \mathcal{L}(W)$ es autoadjunto. Aplicando la hipótesis inductiva a $T|_W$, deducimos que existe $\{v_1, \ldots, v_{n-1}\}$ base ortonormal de W formada por vectores propios de T. Sea $v_n = \frac{v}{||v||}$, como $v_1, \ldots, v_{n-1} \in W = [v]^\perp$ y $v_n \in [v]$ es $v_i \perp v_n$, $\forall i = 1, \ldots, n-1$. Luego $\mathcal{B} = \{v_1, \ldots, v_{n-1}, v_n\}$ es una base ortonormal de V formada por vectores propios de T. □

Corolario 3.1.17. *Si $A \in M_n(\mathbb{R})$ es simétrica, entonces A es diagonalizable en $M_n(\mathbb{R})$.*

Dem. Como $A \in M_n(\mathbb{R})$ es simétrica, entonces $L_A : \mathbb{R}^n \to \mathbb{R}^n$ es autoadjunto (con el producto escalar), luego el operador L_A es diagonalizable y sabemos que eso equivale a que la matriz A sea diagonalizable. □

Juntando las proposiciones 3.1.7, 3.1.11 y 3.1.16, obtenemos el siguiente resultado.

Teorema 3.1.18. *Sea $T \in \mathcal{L}(V)$.*

1. *Si $\Bbbk = \mathbb{R}$, entonces T es autoadjunto si y solo si T es diagonalizable en una base ortonormal.*

2. *Si $\Bbbk = \mathbb{C}$, entonces T es autoadjunto si y solo si T es diagonalizable en una base ortonormal y sus valores propios son reales.* □

Observación 3.1.19. En el caso real, el teorema anterior nos dice que ser autoadjunto es una condición necesaria y suficiente para que un operador sea diagonalizable en una base ortonormal. En el caso complejo es una condición suficiente, pero vimos en el ejemplo 3.1.9 que no es necesaria. Para obtener una condición necesaria y suficiente en el caso complejo se necesita la normalidad, que veremos más adelante.

[2] Si $\dim V = 1$, entonces toda transformación lineal de V en V es de la forma $T(v) = \lambda v$ para algún $\lambda \in \Bbbk$.

Terminamos esta sección dando una caracterización de las proyecciones ortogonales sobre subespacios. Recordar que una *proyección* en un espacio V es un operador $P \in \mathcal{L}(V)$ que verifica $P^2 = P$.

Proposición 3.1.20. *Sea $P \in \mathcal{L}(V)$. Entonces P es la proyección ortogonal (sobre algún subespacio) si y solo si P es una proyección y es un operador autoadjunto.*

Dem. Sea $W \subset V$ un subespacio y P_W la proyección ortogonal sobre W. Probaremos que P_W es autoadjunto. Sean $v_1, v_2 \in V$ arbitrarios. Escribimos $v_1 = w_1 + w_1'$ y $v_2 = w_2 + w_2'$, donde $w_1, w_2 \in W$, $w_1', w_2' \in W^\perp$, $i = 1, 2$. Entonces

$$\begin{aligned} \langle P_W(v_1), v_2\rangle &= \langle w_1, w_2 + w_2'\rangle = \langle w_1, w_2\rangle + \langle w_1, w_2'\rangle = \langle w_1, w_2\rangle, \\ \langle v_1, P_W(v_2)\rangle &= \langle w_1 + w_1', w_2\rangle = \langle w_1, w_2\rangle + \langle w_1', w_2\rangle = \langle w_1, w_2\rangle. \end{aligned}$$

Luego $\langle P_W(v_1), v_2\rangle = \langle v_1, P_W(v_2)\rangle$, para todo $v_1, v_2 \in V$.

Sea ahora P una proyección que además es un operador autoadjunto. La relación $P^2 = P$ implica $V = \operatorname{Im} P \oplus \operatorname{Ker} P$ y P es la proyección sobre $\operatorname{Im} P$ en la dirección de $\operatorname{Ker} P$. Para ver que P es la proyección ortogonal sobre $\operatorname{Im} P$, lo único que nos falta probar es $\operatorname{Ker} P = (\operatorname{Im} P)^\perp$. Sean $u \in \operatorname{Im} P$ (luego $P(u) = u$) y $v \in \operatorname{Ker} P$, es

$$\langle u, v\rangle = \langle P(u), v\rangle = \langle u, P(v)\rangle = \langle u, 0\rangle = 0 \quad \Rightarrow \quad \operatorname{Ker} P \subset (\operatorname{Im} P)^\perp.$$

De $V = \operatorname{Im} P \oplus \operatorname{Ker} P$ y $V = \operatorname{Im} P \oplus (\operatorname{Im} P)^\perp$ se deduce que $\dim \operatorname{Ker} P = \dim(\operatorname{Im} P)^\perp$, luego la relación $\operatorname{Ker} P \subset (\operatorname{Im} P)^\perp$ implica $\operatorname{Ker} P = (\operatorname{Im} P)^\perp$. Luego $P = P_W$, siendo $W = \operatorname{Im} P$. $\square$

Definición 3.1.21. Un operador $P \in \mathcal{L}(V)$ es una *proyección ortogonal* si es una proyección y además es un operador autoadjunto. La proposición anterior prueba que toda proyección ortogonal es la proyección ortogonal sobre algún subespacio (este subespacio es la imagen de la proyección).

3.2. El adjunto de un operador

En la sección anterior definimos que un operador T es autoadjunto si verifica $\langle T(v), w\rangle = \langle v, T(w)\rangle$, para todo $v, w \in V$. En general, vale el siguiente resultado.

Teorema 3.2.1. *Si $T \in \mathcal{L}(V)$, entonces existe un único $T^* \in \mathcal{L}(V)$ tal que*

$$\langle T(v), w\rangle = \langle v, T^*(w)\rangle, \quad \forall v, w \in V. \tag{3.2}$$

Dem. Sea $\mathcal{B} = \{v_1, \dots, v_n\}$ una base ortonormal de V.
Supongamos que existe $T^* \in \mathcal{L}(V)$ que verifica (3.2) y sea $w \in V$ arbitrario. Entonces usando la fórmula (2.5) obtenemos

$$T^*(w) = \sum_{i=1}^n \langle T^*(w), v_i\rangle v_i = \sum_{i=1}^n \overline{\langle v_i, T^*(w)\rangle} v_i = \sum_{i=1}^n \overline{\langle T(v_i), w\rangle} v_i = \sum_{i=1}^n \langle w, T(v_i)\rangle v_i.$$

Luego si existe $T^* \in \mathcal{L}(V)$ que verifica (3.2), entonces $T^*(w)$ está determinado por la fórmula anterior. Esto prueba la unicidad. Para ver la existencia, definamos $T^* : V \to V$ mediante

$$T^*(w) = \sum_{i=1}^n \langle w, T(v_i)\rangle v_i, \quad \forall w \in V. \tag{3.3}$$

La prueba de que T^* así definida es lineal, es simple y queda como ejercicio. Sean $v, w \in V$ arbitrarios. Aplicando las fórmulas (2.5) y (3.3) obtenemos

$$\langle v, T^*(w)\rangle = \left\langle \sum_{i=1}^n \langle v, v_i\rangle v_i, \sum_{j=1}^n \langle w, T(v_j)\rangle v_j \right\rangle = \sum_{i=j=1}^n \langle v, v_i\rangle \overline{\langle w, T(v_j)\rangle} \langle v_i, v_j\rangle = \sum_{i=1}^n \langle v, v_i\rangle \langle T(v_i), w\rangle;$$

$$\langle T(v), w\rangle = \left\langle T\left(\sum_{i=1}^n \langle v, v_i\rangle v_i\right), w \right\rangle = \left\langle \sum_{i=1}^n \langle v, v_i\rangle T(v_i), w \right\rangle = \sum_{i=1}^n \langle v, v_i\rangle \langle T(v_i), w\rangle.$$

Luego $\langle v, T^*(w)\rangle = \langle T(v), w\rangle$, para todo $v, w \in V$. $\square$

Definición 3.2.2. El operador T^* se llama el *adjunto* del operador T.

Observación 3.2.3. La propiedad que caracteriza al adjunto vale también intercambiando los lugares de T y T^*. Es decir, si $T \in \mathcal{L}(V)$ entonces vale $\langle v, T(w)\rangle = \langle T^*(v), w\rangle$, para todo $v, w \in V$. Esto se obtiene tomando conjugados en la fórmula (3.2).

Observación 3.2.4. Un operador T es autoadjunto si y solo si $T^* = T$ (de ahí viene el término "autoadjunto").

A continuación veremos algunas propiedades de los adjuntos.

Proposición 3.2.5. *Sean $T, S \in \mathcal{L}(V)$ y $a \in \mathbb{k}$. Entonces*

$$(aT + S)^* = \bar{a}T^* + S^*, \qquad (T \circ S)^* = S^* \circ T^*, \qquad (T^*)^* = T, \qquad \mathrm{Id}^* = \mathrm{Id}.$$

Dem. Consideremos la primera igualdad.

$$\begin{aligned} \langle (aT+S)(v), w\rangle &= \langle aT(v) + S(v), w\rangle = a\langle T(v), w\rangle + \langle S(v), w\rangle = a\langle v, T^*(w)\rangle + \langle v, S^*(w)\rangle \\ &= \langle v, \overline{a}T^*(w) + S^*(w)\rangle = \langle v, (\overline{a}T^* + S^*)(w)\rangle, \quad \forall v, w \in V, \end{aligned}$$

luego la unicidad del adjunto implica $(aT + S)^* = \bar{a}T^* + S^*$. La prueba de la segunda afirmación es la siguiente

$$\langle (T \circ S)(v), w\rangle = \big\langle T\big(S(v)\big), w\big\rangle = \langle S(v), T^*(w)\rangle = \big\langle v, S^*\big(T^*(w)\big)\big\rangle = \langle v, (S^* \circ T^*)(w)\rangle, \quad \forall v, w \in V,$$

Luego $(T \circ S)^* = S^* \circ T^*$. Para la tercera, usando la observación 3.2.3 obtenemos

$$\langle T^*(v), w\rangle = \langle v, T(w)\rangle, \quad \forall v, w \in V,$$

luego el adjunto de T^* es T, es decir, $(T^*)^* = T$. La cuarta igualdad es inmediata. $\square$

Definición 3.2.6. Si $A = (a_{ij}) \in M_n(\mathbb{C})$, definimos $\overline{A} := (\overline{a_{ij}}) \in M_n(\mathbb{C})$ y $A^* := \overline{A}^t \in M_n(\mathbb{C})$, es decir, $A^* = (b_{ij})$, siendo $b_{ij} = \overline{a_{ji}}$. La matriz A^* se llama la *adjunta* de A. Notar que si $A \in M_n(\mathbb{C})$, entonces vale $A^* = A^t$ si y solo si $A \in M_n(\mathbb{R})$.

Proposición 3.2.7. *Si $T \in \mathcal{L}(V)$ y $\mathcal{B}$ es una base ortonormal de V entonces $[T^*]_\mathcal{B} = ([T]_\mathcal{B})^*$.*

Dem. Sea $[T]_\mathcal{B} = (a_{ij})$ y $[T^*]_\mathcal{B} = (c_{ij})$. Entonces $a_{ij} = \langle T(v_j), v_i\rangle = \langle v_j, T^*(v_i)\rangle = \overline{\langle T^*(v_i), v_j\rangle} = \overline{c_{ji}}$. $\square$

Observación 3.2.8. La proposición anterior nos da una forma de hallar el adjunto. Dado $T \in \mathcal{L}(V)$, elegimos $\mathcal{B}$ una base ortonormal de V y calculamos $[T^*]_\mathcal{B} = ([T]_\mathcal{B})^*$. Luego a partir de $[T^*]_\mathcal{B}$ obtenemos T^*.

Corolario 3.2.9. *Si $A \in M_n(\mathbb{k})$, entonces $(L_A)^* = L_{A^*}$ en $\mathcal{L}(\mathbb{k}^n)$, es decir vale*

$$\langle Ax, y\rangle = \langle x, A^*y\rangle, \quad \forall x, y \in \mathbb{k}^n,$$

considerando en $\mathbb{k}^n$ el producto interno usual.

Dem. La base canónica $\mathcal{B}$ es ortonormal respecto al producto interno usual y $[L_A]_\mathcal{B} = A$, luego $[(L_A)^*]_\mathcal{B} = A^*$, lo cual implica $(L_A)^* = L_{A^*}$. $\square$

Ejemplo 3.2.10. Consideremos $\mathbb{C}^2$ con el producto interno usual y $T : \mathbb{C}^2 \to \mathbb{C}^2$ definido por $T(x, y) = (2ix + 3y, x - y)$, $(x, y) \in \mathbb{C}^2$. Si $\mathcal{B}$ es la base canónica de $\mathbb{C}^2$, entonces

$$[T]_\mathcal{B} = \begin{pmatrix} 2i & 3 \\ 1 & -1 \end{pmatrix} \quad \Rightarrow \quad [T^*]_\mathcal{B} = ([T]_\mathcal{B})^* = \overline{\begin{pmatrix} 2i & 3 \\ 1 & -1 \end{pmatrix}^t} = \begin{pmatrix} -2i & 1 \\ 3 & -1 \end{pmatrix},$$

luego $T^*(x, y) = (-2ix + y, 3x - y)$.

De las proposiciones 3.2.5 y 3.2.7 se deduce directamente el siguiente resultado.

Corolario 3.2.11. *Sean $A, B \in M_n(\Bbbk)$ y $a \in \Bbbk$. Entonces*

$$(aA + B)^* = \bar{a}A^* + B^*, \qquad (AB)^* = B^*A^*, \qquad (A^*)^* = A, \qquad I^* = I. \quad \square$$

Proposición 3.2.12. *1. Si $T \in \mathcal{L}(V)$, entonces* $\det(T^*) = \overline{\det(T)}$.

2. Si $A \in M_n(\Bbbk)$, entonces $\det(A^*) = \overline{\det(A)}$.

Dem. Por la proposición 3.2.7 alcanza con probar la segunda afirmación.

$$\det(A^*) = \det\left(\overline{A}^t\right) = \det\left(\overline{A}\right) = \overline{\det(A)}.$$

En lo anterior usamos $\det\left(\overline{A}\right) = \overline{\det(A)}$, lo cual es fácil de probar. $\square$

El siguiente resultado relaciona los valores propios de un operador con los de su adjunto.

Proposición 3.2.13. *Sea $T \in \mathcal{L}(V)$. Si λ es un valor propio de T, entonces $\overline{\lambda}$ es un valor propio de T^*.*

Dem.

$$\chi_{T^*}\left(\overline{\lambda}\right) = \det\left(T^* - \overline{\lambda}\mathrm{Id}\right) = \det[(T - \lambda\mathrm{Id})^*] = \overline{\det(T - \lambda\mathrm{Id})} = \overline{\chi_T(\lambda)}.$$

Entonces si λ es valor propio de T, es $\chi_{T^*}(\overline{\lambda}) = \overline{\chi_T(\lambda)} = \overline{0} = 0$. Luego $\overline{\lambda}$ es un valor propio de T^*. $\square$

Observación 3.2.14. La proposición anterior dice que los valores propios de T^* son los conjugados de los valores propios de T, pero no dice nada sobre los vectores propios correspondientes, los cuales no suelen estar relacionados.

3.3. Operadores normales

Definición 3.3.1. Un operador $T \in \mathcal{L}(V)$ es *normal* si verifica $T \circ T^* = T^* \circ T$. Análogamente, decimos que una matriz $A \in M_n(\Bbbk)$ es *normal* si verifica $AA^* = A^*A$

La prueba del siguiente resultado es inmediata a partir de la proposición 3.2.7.

Proposición 3.3.2. *Si $T \in \mathcal{L}(V)$ y $\mathcal{B}$ es una base ortonormal de V, entonces T es normal si y solo si $[T]_\mathcal{B}$ es normal.* $\square$

Observación 3.3.3. Es claro que si $T \in \mathcal{L}(V)$ es autoadjunto, entonces T es normal. El siguiente ejemplo muestra que el recíproco es falso.

Ejemplo 3.3.4. Sea $0 < \theta < \pi$ y $A = \left(\begin{smallmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{smallmatrix}\right) \in M_2(\mathbb{R})$. Es $AA^t = A^tA = \mathrm{Id}$, luego A es normal. Por otro lado es claro que A no es simétrica, luego la rotación $L_A : \mathbb{R}^2 \to \mathbb{R}^2$ es normal pero no es autoadjunta.

El siguiente resultado muestra que la normalidad es una condición necesaria para que un operador sea diagonalizable en una base ortonormal.

Proposición 3.3.5. *Si $T \in \mathcal{L}(V)$ es diagonalizable en una base ortonormal de V, entonces T es normal.*

Dem. Sea $\mathcal{B}$ base ortonormal de V tal que $[T]_{\mathcal{B}} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$, en que esta última es una forma abreviada de escribir la matriz diagonal cuya diagonal principal es $(\lambda_1, \ldots, \lambda_n)$. La proposición 3.2.7 implica $[T^*]_{\mathcal{B}} = \operatorname{diag}\big(\overline{\lambda_1}, \ldots, \overline{\lambda_n}\big)$. Multiplicando estas matrices obtenemos

$$[T]_{\mathcal{B}}[T^*]_{\mathcal{B}} = [T^*]_{\mathcal{B}}[T]_{\mathcal{B}} = \operatorname{diag}\big(|\lambda_1|^2, \ldots, |\lambda_n|^2\big)\,.$$

En particular esto implica $T \circ T^* = T^* \circ T$. □

Observación 3.3.6. En el caso real la proposición anterior no nos dice nada nuevo, ya que en ese caso sabemos que el operador es autoadjunto (teorema 3.1.18), y eso es más fuerte que ser normal. Lo interesante de la misma es que nos muestra que en el caso complejo, la normalidad es una condición necesaria para que un operador sea diagonalizable en una base ortonormal. A continuación veremos que vale también el recíproco. Para probarlo necesitamos algunas propiedades de los operadores normales, que son interesantes en sí mismas y valen tanto para espacios reales como complejos.

Proposición 3.3.7. *Sea $T \in \mathcal{L}(V)$ un operador normal. Entonces*

1. *$||T(v)|| = ||T^*(v)||$, para todo $v \in V$.*
2. *Si v es un vector propio de T correspondiente al valor propio λ, entonces v es un vector propio de T^* correspondiente al valor propio $\overline{\lambda}$.*
3. *Si $\lambda_1 \neq \lambda_2$ son valores propios de T con vectores propios correspondientes v_1 y v_2, entonces $v_1 \perp v_2$.*

Dem.

1. $||T(v)||^2 = \langle T(v), T(v)\rangle = \langle v, T^*(T(v))\rangle = \langle v, T(T^*(v))\rangle = \langle T^*(v), T^*(v)\rangle = ||T^*(v)||^2$.
2. Veamos primero que $T - \lambda\mathrm{Id}$ es normal, para todo $\lambda \in \Bbbk$.

$$\begin{aligned}(T - \lambda\mathrm{Id}) \circ (T - \lambda\mathrm{Id})^* &= (T - \lambda\mathrm{Id}) \circ (T^* - \overline{\lambda}\mathrm{Id}) = T \circ T^* - \lambda T^* - \overline{\lambda}T + |\lambda|^2\mathrm{Id} \\ &= T^* \circ T - \lambda T^* - \overline{\lambda}T + |\lambda|^2\mathrm{Id} = (T - \lambda\mathrm{Id})^* \circ (T - \lambda\mathrm{Id}).\end{aligned}$$

   Luego aplicando la parte 1) a $T - \lambda\mathrm{Id}$, obtenemos que si $T(v) = \lambda v$, es

$$0 = ||T(v) - \lambda v|| = ||(T - \lambda\mathrm{Id})(v)|| = ||(T - \lambda\mathrm{Id})^*(v)|| = \left\|T^*(v) - \overline{\lambda}v\right\|.$$

   Luego $T^*(v) = \overline{\lambda}v$.
3. La afirmación se deduce del cálculo siguiente

$$\lambda_1\langle v_1, v_2\rangle = \langle \lambda_1 v_1, v_2\rangle = \langle T(v_1), v_2\rangle = \langle v_1, T^*(v_2)\rangle = \langle v_1, \overline{\lambda_2}v_2\rangle = \lambda_2\langle v_1, v_2\rangle.$$

   Como $\lambda_1 \neq \lambda_2$, resulta $\langle v_1, v_2\rangle = 0$. □

Lema 3.3.8. *Sea $T \in \mathcal{L}(V)$ y $W \subset V$ un subespacio que es T-invariante y T^*-invariante. Si consideramos la restricción $T|_W \in \mathcal{L}(W)$, entonces $(T|_W)^* = T^*|_W$. Si además T es normal, entonces también lo es $T|_W$.*

Dem. Si $w_1, w_2 \in W$, es

$$\langle T|_W(w_1), w_2\rangle = \langle T(w_1), w_2\rangle = \langle w_1, T^*(w_2)\rangle = \langle w_1, T^*|_W(w_2)\rangle,\ \forall w_1, w_2 \in W.$$

Luego $(T|_W)^* = T^*|_W$. Por lo tanto si T es normal, es

$$T|_W \circ (T|_W)^* = T|_W \circ T^*|_W = (T \circ T^*)|_W = (T^* \circ T)|_W = T^*|_W \circ T|_W = (T|_W)^* \circ T|_W. \quad \square$$

El siguiente resultado muestra que en el caso complejo vale el recíproco de la proposición 3.3.5. La prueba es muy similar a la de la proposición 3.1.16.

Proposición 3.3.9. *Si $\Bbbk = \mathbb{C}$ y $T \in \mathcal{L}(V)$ es normal, entonces T es diagonalizable en una base ortonormal*

Dem. La prueba es por inducción en $n = \dim V$.

Si $n = 1$, entonces tomando $v \in V$ con $\|v\| = 1$, obtenemos que $\{v\}$ es una base ortonormal de V formada por vectores propios de T.

Sea ahora $n > 1$ y supongamos razonando inductivamente que si tenemos un operador normal en un espacio de dimensión $n - 1$, entonces existe una base ortonormal del espacio formada por vectores propios del operador.

Como $\Bbbk = \mathbb{C}$ existe λ valor propio de T. Sea v un vector propio de T correspondiente a λ y consideremos $W := [v]^\perp = \{w \in V \mid \langle w, v\rangle = 0\}$. Sabemos que W es un subespacio de V y que $\dim W = n - 1$. Sea $w \in W$,

$$\begin{aligned} \langle T(w), v\rangle &= \langle w, T^*(v)\rangle = \left\langle w, \overline{\lambda} v\right\rangle = \lambda\langle w, v\rangle = \lambda 0 = 0 \quad \Rightarrow \quad T(w) \in W, \\ \langle T^*(w), v\rangle &= \langle w, T(v)\rangle = \langle w, \lambda v\rangle = \overline{\lambda}\langle w, v\rangle = \overline{\lambda} 0 = 0 \quad \Rightarrow \quad T^*(w) \in W. \end{aligned}$$

Luego W es T y T^*-invariante, como T es normal, el lema anterior implica que $T|_W \in \mathcal{L}(W)$ es normal. Tenemos que $T|_W \in \mathcal{L}(W)$ normal y que $\dim W = n - 1$ entonces por la hipótesis inductiva existe $\{v_1, \ldots, v_{n-1}\}$ base ortonormal de W formada por vectores propios de $T|_W$ (y por lo tanto vectores propios de T). Sea $v_n = \frac{v}{||v||}$, como $v_1, \ldots, v_{n-1} \in W = [v]^\perp$ y $v_n \in [v]$ es $v_i \perp v_n$, para todo $i = 1, \ldots, n - 1$. Luego $\mathcal{B} = \{v_1, \ldots, v_{n-1}, v_n\}$ es una base ortonormal de V formada por vectores propios de T. $\square$

En resumen, podemos juntar el teorema 3.1.18 y las proposiciones 3.3.5 y 3.3.9 en el siguiente resultado.

Teorema 3.3.10 (Teorema espectral)**.** *Un operador $T \in \mathcal{L}(V)$ es diagonalizable en una base ortonormal si y solo si T es normal en el caso complejo o autoadjunto en el caso real.* $\square$

Lo que sigue profundiza en la relación entre normal y autoadjunto.

Proposición 3.3.11. *Si $T \in \mathcal{L}(V)$ y $\Bbbk = \mathbb{C}$, entonces T es autoadjunto si y solo si T es normal y todos sus valores propios son reales.*

Dem. El directo es inmediato. Para el recíproco, como T es normal, entonces T es diagonalizable en una base ortonormal. Además los valores propios de T son reales, luego el teorema 3.1.18 implica que T es autoadjunto. $\square$

Proposición 3.3.12. *Si $T \in \mathcal{L}(V)$, $\Bbbk = \mathbb{R}$ y T es normal, entonces las siguientes afirmaciones son equivalentes.*

1. *T es autoadjunto.*
2. *T es diagonalizable.*

3. El polinomio característico de T se escinde en $\mathbb{R}$.

Dem. ($1 \Rightarrow 2 \Rightarrow 3$). Esto es inmediato

($3 \Rightarrow 2$). Sea $\mathcal{B}$ una base ortonormal de V y $A = [T]_\mathcal{B}$. Como T es normal, entonces $A \in M_n(\mathbb{R}) \subset M_n(\mathbb{C})$ es normal. Luego $L_A \in \mathcal{L}(\mathbb{C}^n)$ es un operador normal. Entonces $L_A \in \mathcal{L}(\mathbb{C}^n)$ es diagonalizable y por lo tanto A es semejante en $M_n(\mathbb{C})$ a una matriz diagonal D, siendo las entradas diagonales de D los valores propios de A. Como los valores propios de A son las raíces del polinomio característico de T y estas son reales, concluimos que los valores propios de A son reales; luego $D \in M_n(\mathbb{R})$. Entonces aplicando la proposición 1.4.9 concluimos que A es semejante en $M_n(\mathbb{R})$ a la matriz diagonal D. Luego T es diagonalizable.

($2 \Rightarrow 1$). Como T es diagonalizable, es $V = E_{\lambda_1} \oplus \cdots \oplus E_{\lambda_r}$, siendo $\lambda_1, \ldots, \lambda_r$ los valores propios de T. Al ser T normal, sabemos que estos subespacios propios son ortogonales entre sí (proposición 3.3.7). Luego si para cada $i = 1, \ldots, r$ elegimos $\mathcal{B}_i$ una base ortonormal de E_{λ_i}, entonces $\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_r$ es una base ortonormal de V formada por vectores propios de T. Luego T es autoadjunto (por el teorema anterior). $\square$

3.4. Isometrías

Definición 3.4.1. Sean V y W dos espacios con producto interno. Una *isometría* de V en W es una transformación lineal $T : V \to W$, que verifica

$$\langle T(u), T(v) \rangle = \langle u, v \rangle, \quad \forall u, v \in V.$$

Si $V = W$, entonces a las isometrías también se les llama operadores *ortogonales* en el caso real o *unitarios* en el caso complejo. Nosotros estudiaremos solo este tipo de isometrías.

En lo que sigue V es un espacio con producto interno de dimensión finita.

Proposición 3.4.2. *Si $T \in \mathcal{L}(V)$, entonces T es una isometría si y solo si T preserva la norma:*

$$\|T(v)\| = \|v\|, \quad \forall v \in V.$$

Dem. Si T es una isometría y $v \in V$, entonces $\|T(v)\| = \sqrt{\langle T(v), T(v) \rangle} = \sqrt{\langle v, v \rangle} = \|v\|$.
Recíprocamente, si T preserva la norma entonces de la linealidad de T y de las fórmulas de polarización

$$\langle u, v \rangle = \frac{1}{4}\big(\|u+v\|^2 - \|u-v\|^2\big), \ \Bbbk = \mathbb{R}; \quad \langle u, v \rangle = \frac{1}{4}\sum_{k=0}^{k=3} i^k \left\|u + i^k v\right\|^2, \quad \Bbbk = \mathbb{C}; \quad \forall u, v \in V,$$

se deduce que T preserva el producto interno. Por ejemplo, en el caso real es

$$\begin{aligned}\langle T(u), T(v) \rangle &= \frac{1}{4}\big(\|T(u)+T(v)\|^2 - \|T(u)-T(v)\|^2\big) = \frac{1}{4}\big(\|T(u+v)\|^2 - \|T(u-v)\|^2\big) \\ &= \frac{1}{4}\big(\|u+v\|^2 - \|u-v\|^2\big) = \langle u, v \rangle.\end{aligned}$$

El caso complejo es análogo. $\square$

Ejemplo 3.4.3. Sea $\theta \in \mathbb{R}$ y consideremos el operador $T = L_A : \mathbb{R}^2 \to \mathbb{R}^2$ con $A = \left(\begin{smallmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{smallmatrix}\right)$. Entonces

$$\begin{aligned}\|T(x,y)\|^2 &= \|\big(x\cos(\theta) - y\operatorname{sen}(\theta), x\operatorname{sen}(\theta) + y\cos(\theta)\big)\|^2 \\ &= \big(x\cos(\theta) - y\operatorname{sen}(\theta)\big)^2 + \big(x\operatorname{sen}(\theta) + y\cos(\theta)\big)^2 \\ &= x^2\big(\cos^2(\theta) + \operatorname{sen}^2(\theta)\big) + y^2\big(\cos^2(\theta) + \operatorname{sen}^2(\theta)\big) = x^2 + y^2 = \|(x,y)\|^2.\end{aligned}$$

Luego $\|T(x,y)\| = \|(x,y)\|$, para todo $(x,y) \in \mathbb{R}^2$, y por lo tanto T es una isometría. Notar que T es la rotación en el plano $\mathbb{R}^2$ de centro en el origen y ángulo θ.

Proposición 3.4.4. *Sea $T \in \mathcal{L}(V)$. Las siguientes afirmaciones son equivalentes:*

1. *T es una isometría.*

2. *Para toda base ortonormal $\mathcal{B}$ de V, el conjunto $T(\mathcal{B})$ es una base ortonormal de V.*

3. *Existe una base ortonormal $\mathcal{B}$ de V tal que $T(\mathcal{B})$ es una base ortonormal de V.*

Dem. $(1 \Rightarrow 2)$. Sea $\mathcal{B} = \{v_1, \ldots, v_n\}$ una base ortonormal de V, es $\langle T(v_i), T(v_j)\rangle = \langle v_i, v_j\rangle = \delta_{ij}$, luego $T(\mathcal{B}) = \{T(v_1), \ldots, T(v_n)\}$ es un conjunto ortonormal (y por lo tanto es LI) con la misma cantidad de elementos que $\mathcal{B}$, luego $T(\mathcal{B})$ es una base ortonormal de V.

$(2 \Rightarrow 3)$. Esto es obvio, dado que V siempre admite una base ortonormal.

$(3 \Rightarrow 1)$. Supongamos que existe una base ortonormal $\mathcal{B} = \{w_1, \ldots, w_n\}$ tal que $T(\mathcal{B}) = \{T(w_1), \ldots, T(w_n)\}$ es también una base ortonormal. Sea $v \in V$. Como $\mathcal{B}$ es una base de V, entonces existen únicos $a_1, \ldots, a_n \in \Bbbk$ tales que $v = \sum_{i=1}^n a_i w_i$, luego $T(v) = \sum_{i=1}^n a_i T(w_i)$. Como $\mathcal{B}$ es una base ortonormal es $\|v\|^2 = \sum_{i=1}^n |a_i|^2$ y como $T(\mathcal{B})$ es una base ortonormal es $\|T(v)\|^2 = \sum_{i=1}^n |a_i|^2$. Luego $\|T(v)\| = \|v\|$, para todo $v \in V$ y se aplica la proposición anterior. $\square$

Proposición 3.4.5. *Sea $T \in \mathcal{L}(V)$. Las siguientes afirmaciones son equivalentes*

$$T \text{ es una isometría}; \qquad T^* \circ T = \mathrm{Id}; \qquad T \circ T^* = \mathrm{Id}; \qquad T \text{ es un isomorfismo y } T^{-1} = T^*.$$

Dem. La equivalencia entre las tres últimas condiciones es una propiedad bien conocida de los operadores en espacios de dimensión finita. Probaremos la equivalencia entre la primera y la segunda. Observar que es

$$\langle T(u), T(v)\rangle = \langle u, T^*(T(v))\rangle = \langle u, (T^* \circ T)(v)\rangle, \quad \forall u, v \in V.$$

Luego T es una isometría si y solo si

$$\langle u, (T^* \circ T)(v)\rangle = \langle u, v\rangle, \quad \forall u, v \in V \quad \Leftrightarrow \quad (T^* \circ T)(v) = v, \quad \forall v \in V \quad \Leftrightarrow \quad T^* \circ T = \mathrm{Id}. \quad \square$$

Observación 3.4.6. Es claro que la función identidad Id_V es una isometría y es fácil de probar que la composición de isometrías es una isometría, y que la inversa de una isometría es también una isometría. Luego las isometrías de V en V forman un grupo, con la operación de composición.

Observaciones 3.4.7. Sea T una isometría.

1. Vimos que vale $T \circ T^* = T^* \circ T = \mathrm{Id}$; luego T es normal.

2. Si $\lambda \in \Bbbk$ es un valor propio de T, entonces $|\lambda| = 1$ ($\lambda = \pm 1$, si $\Bbbk = \mathbb{R}$):

$$\text{sea } 0 \neq v \in V : T(v) = \lambda v \quad \Rightarrow \quad \|v\| = \|T(v)\| = |\lambda|\,\|v\| \quad \Rightarrow \quad |\lambda| = 1.$$

3. Vale $|\det(T)| = 1$ ($\det T = \pm 1$, si $\Bbbk = \mathbb{R}$):

$$1 = \det(\mathrm{Id}) = \det(T \circ T^*) = \det(T)\det(T^*) = \det(T)\,\overline{\det(T)} = |\det(T)|^2 \quad \Rightarrow \quad |\det(T)| = 1.$$

4. Si $\Bbbk = \mathbb{C}$, como T es normal, entonces T es diagonalizable en una base ortonormal. Si $\Bbbk = \mathbb{R}$ esto no es necesariamente cierto. Por ejemplo, la rotación de centro en el origen y ángulo θ (con $\theta \neq k\pi$, $k \in \mathbb{Z}$) es una isometría y no es diagonalizable.

Proposición 3.4.8. *Sea $T \in \mathcal{L}(V)$.*

1. *Si $\Bbbk = \mathbb{C}$, entonces T es una isometría si y solo si T es normal y vale $|\lambda| = 1$ para todo valor propio λ de T.*

2. *Si $\Bbbk = \mathbb{R}$, entonces T es una isometría y es diagonalizable si y solo si T es autoadjunta y vale $\lambda = \pm 1$ para todo valor propio λ de T.*

Dem. Consideremos primero el caso $\Bbbk = \mathbb{C}$. El directo es inmediato, así que probaremos solo el recíproco. Como T es normal, entonces existe una base ortonormal $\mathcal{B} = \{v_1, \ldots, v_n\}$ de V tal que $T(v_i) = \lambda_i v_i$, para todo i, siendo $\lambda_1, \ldots, \lambda_n$ los valores propios correspondientes. Por hipótesis sabemos que es $|\lambda_i| = 1$, para todo $i = 1, \ldots, n$. Luego

$$\begin{aligned} \langle T(v_i), T(v_j) \rangle &= \langle \lambda_i v_i, \lambda_j v_j \rangle = \lambda_i \overline{\lambda_j} \langle v_i, v_j \rangle = \lambda_i \overline{\lambda_j} 0 = 0, \quad \forall i \neq j, \\ \|T(v_i)\| &= \|\lambda_i v_i\| = |\lambda_i| \, \|v_i\| = 1, \quad \forall i. \end{aligned}$$

Entonces $T(\mathcal{B})$ es una base ortonormal y por lo tanto T es una isometría (proposición 3.4.4).

Consideremos ahora el caso $\Bbbk = \mathbb{R}$. Para el directo, como T es una isometría, entonces sabemos que T es normal y $\lambda = \pm 1$ para todo valor propio λ de T. Además, T es normal y diagonalizable, lo cual implica que T es autoadjunto (proposición 3.3.12). El recíproco es análogo al caso complejo. Como T es autoadjunto, entonces existe una base ortonormal $\mathcal{B}$ formada por vectores propios de T; además sabemos que si λ es un valor propio de T, entonces $\lambda = \pm 1$, lo cual equivale a $|\lambda| = 1$. Luego de la misma forma que en el caso complejo se deduce que $T(\mathcal{B})$ es una base ortonormal, lo cual implica que T es una isometría. □

Observación 3.4.9. En el caso complejo, la proposición anterior muestra que un operador es una isometría si y solo si es diagonalizable en una base ortonormal y sus valores propios tienen módulo uno. En el caso real, obtuvimos una caracterización similar para el caso de operadores diagonalizables. En el final de este capítulo se verá la clasificación de las isometrías en $\mathbb{R}^2$ y $\mathbb{R}^3$ (diagonalizables o no).

Matrices ortogonales y unitarias.

Definición 3.4.10. Sea $\Bbbk$ un cuerpo arbitrario, una matriz $A \in M_n(\Bbbk)$ se dice *ortogonal* si $A^t A = A A^t = I$. Si $\Bbbk = \mathbb{C}$, una matriz $A \in M_n(\mathbb{C})$ se dice *unitaria* si $A^* A = A A^* = I$.

Observaciones 3.4.11.

1. Como una matriz cuadrada es invertible si y solo si es invertible por la derecha o por la izquierda, se deduce que, dada $A \in M_n(\Bbbk)$, las condiciones

$$A \text{ es ortogonal}; \quad A^t A = I; \quad A A^t = I; \quad A \text{ es invertible y } A^{-1} = A^t$$

son equivalentes (esto vale en todo cuerpo $\Bbbk$). Análogamente, si $A \in M_n(\mathbb{C})$, entonces las condiciones

$$A \text{ es unitaria}; \quad A^* A = I; \quad A A^* = I; \quad A \text{ es invertible y } A^{-1} = A^*$$

son equivalentes.

2. La ortogonalidad se puede definir para matrices con coeficientes en un cuerpo cualquiera, mientras que el ser unitaria se define solo para matrices complejas. Si a una matriz real A la consideramos como matriz compleja, entonces es unitaria si y solo si es ortogonal, ya que en ese caso es $A^* = A^t$. Esto es falso para matrices complejas que no son reales. Esta situación es análoga a la que se da entre las matrices hermitianas y la matrices complejas simétricas, que solo coinciden cuando son reales.

Proposición 3.4.12. *Sea $T \in \mathcal{L}(V)$, $\mathcal{B}$ una base ortonormal de V y $A = [T]_{\mathcal{B}}$. Entonces T es una isometría si y solo si A es ortogonal en el caso real o unitaria en el caso complejo.*

Dem. Sea $A = [T]_{\mathcal{B}}$. Como $\mathcal{B}$ es una base ortonormal, es $[T^*]_{\mathcal{B}} = A^*$. Luego

$$T \circ T^* = \mathrm{Id} \;\Leftrightarrow\; [T \circ T^*]_{\mathcal{B}} = [\mathrm{Id}]_{\mathcal{B}} \;\Leftrightarrow\; [T]_{\mathcal{B}}\,[T^*]_{\mathcal{B}} = I \;\Leftrightarrow\; AA^* = I. \quad \square$$

Usando la proposición 3.4.5 y las observaciones anteriores, se obtiene directamente el siguiente resultado.

Corolario 3.4.13. *Sea $A \in M_n(\mathbb{k})$ y consideramos el operador $L_A \in \mathcal{L}(\mathbb{k}^n)$. Entonces L_A es una isometría si y solo si A es ortogonal en el caso real o unitaria en el caso complejo.* $\square$

Proposición 3.4.14. *Sea $A \in M_n(\mathbb{k})$. Las siguientes afirmaciones son equivalentes:*

1. *La matriz A es unitaria si $\mathbb{k} = \mathbb{C}$ o es ortogonal $\mathbb{k} = \mathbb{R}$.*
2. *Las filas de A forman una base ortonormal de $\mathbb{k}^n$.*
3. *Las columnas de A forman una base ortonormal de $\mathbb{k}^n$.*

Dem. Observar primero que si $A = (a_{ij}) \in M_n(\mathbb{k})$, entonces vale

$$AA^* = I \;\Leftrightarrow\; \sum_{k=1}^{n} a_{ik}\overline{a_{jk}} = \delta_{ij}; \quad A^*A = I \;\Leftrightarrow\; \sum_{k=1}^{n} \overline{a_{ki}}a_{kj} = \delta_{ij} \Leftrightarrow \sum_{k=1}^{n} a_{ki}\overline{a_{kj}} = \delta_{ij}; \quad \forall i, j.$$

Luego $AA^* = I$ si y solo si las filas de A foman una base ortonormal de $\mathbb{k}^n$, y $A^*A = I$ si y solo si las columnas de A forman una base ortonormal de $\mathbb{k}^n$. $\square$

El siguiente resultado relaciona bases ortonormales con matrices ortogonales o unitarias.

Proposición 3.4.15. *Sean $\mathcal{B}$ y $\mathcal{C}$ bases de $\mathbb{k}^n$ y $A = {}_{\mathcal{B}}[\mathrm{Id}]_{\mathcal{C}}$. Entonces:*

1. *Si $\mathcal{B}$ y $\mathcal{C}$ son bases ortonormales, entonces A es unitaria si $\mathbb{k} = \mathbb{C}$, u ortogonal si $\mathbb{k} = \mathbb{R}$.*
2. *Si A es unitaria si $\mathbb{k} = \mathbb{C}$ u ortogonal si $\mathbb{k} = \mathbb{R}$ y $\mathcal{B}$ o $\mathcal{C}$ es una base ortonormal, entonces la otra base también es ortonormal.*

Dem. Veremos solo el caso complejo, el caso real es análogo. Probaremos primero que vale lo siguiente.

1. Si $\mathcal{B}$ es una base ortonormal, entonces $\mathcal{C}$ es una base ortonormal si y solo si $A^*A = I$.
2. Si $\mathcal{C}$ es una base ortonormal, entonces $\mathcal{B}$ es una base ortonormal si y solo si $AA^* = I$.

Sean $\mathcal{B} = \{v_1, \dots, v_n\}$, $\mathcal{C} = \{w_1, \dots, w_n\}$ y $A = {}_{\mathcal{B}}[\mathrm{Id}]_{\mathcal{C}} = (a_{ij})$. Luego $w_i = \sum_{k=1}^{n} a_{ki}v_k$, para todo $i = 1, \dots, n$. Supongamos que $\mathcal{B}$ es una base ortonormal. Si tomamos dos elementos de $\mathcal{C}$, $w_i = \sum_{k=1}^{n} a_{ki}v_k$ y $w_j = \sum_{k=1}^{n} a_{kj}v_k$, entonces como $\mathcal{B}$ es ortonormal (recordar la observación 2.2.16) vale $\langle w_i, w_j\rangle = \sum_{k=1}^{n} a_{ki}\overline{a_{kj}}$. Luego $\mathcal{C}$ es una base ortonormal si y solo si $\sum_{k=1}^{n} a_{ki}\overline{a_{kj}} = \delta_{ij}$, para todo i, j, lo cual equivale a $A^*A = I$. Por otro lado, si $\mathcal{C}$ es una base ortonormal, como es $A^{-1} = {}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}}$, deducimos aplicando lo anterior que $\mathcal{B}$ es una base ortonormal si y solo si $(A^{-1})^*A^{-1} = I$. Pero tomando inversos, obtenemos que esta última igualdad equivale a $AA^* = I$. Así $\mathcal{B}$ es una base ortonormal si y solo si $AA^* = I$.

Como las condiciones $AA^* = I$ o $A^*A = I$ equivalen con que A sea unitaria, entonces tenemos lo siguiente.

1. Si $\mathcal{B}$ es una base ortonormal, entonces $\mathcal{C}$ es una base ortonormal si y solo si A es unitaria.
2. Si $\mathcal{C}$ es una base ortonormal, entonces $\mathcal{B}$ es una base ortonormal si y solo si A es unitaria.

Es claro que esto equivale al enunciado de la proposición. $\square$

Definición 3.4.16. 1. Dos matrices $A, B \in M_n(\mathbb{C})$ son *unitariamente equivalentes* si existe una matriz unitaria $Q \in M_n(\mathbb{C})$ tal que $A = QBQ^*$.

2. Dos matrices $A, B \in M_n(\Bbbk)$ ($\Bbbk$ cuerpo arbitrario) son *ortogonalmente equivalentes* si existe una matriz ortogonal $Q \in M_n(\Bbbk)$ tal que $A = QBQ^t$.

Proposición 3.4.17. *La relación de ser unitariamente equivalentes es de equivalencia en $M_n(\mathbb{C})$ y la de ser ortogonalmente equivalentes es de equivalencia en $M_n(\Bbbk)$ ($\Bbbk$ cuerpo arbitrario).*

Dem. Lo probaremos solo para el caso ortogonal, dejando el unitario (que es análogo) como ejercicio.
Notar que el ser ortogonalmente equivalentes es un caso particular de ser semejantes, ya que en ese caso vale $A = QBQ^t = QBQ^{-1}$. Recordando cómo es que se prueba que la semejanza es una relación de equivalencia, vemos lo único que nos falta probar es que la matriz identidad I es ortogonal (lo cual es obvio), que si Q es ortogonal, entonces Q^{-1} también lo es y que si Q_1, Q_2 son ortogonales, entonces Q_1Q_2 también lo es.
Es un ejercicio el probar que si $A \in M_n(\Bbbk)$ es invertible, entonces A^t es invertible y vale $(A^t)^{-1} = (A^{-1})^t$.
Luego si Q es ortogonal, vale

$$QQ^t = I \Rightarrow (QQ^t)^{-1} = I \Rightarrow (Q^t)^{-1}Q^{-1} = I \Rightarrow (Q^{-1})^tQ^{-1} = I.$$

Esto muestra que Q^{-1} es una matriz ortogonal. Además, si Q_1, Q_2 son ortogonales, vale

$$(Q_1Q_2)^tQ_1Q_2 = Q_2^tQ_1^tQ_1Q_2 = Q_2^tIQ_2 = Q_2^tQ_2 = I,$$

luego Q_1Q_2 es ortogonal. $\square$

Teorema 3.4.18. *1. Si $\Bbbk = \mathbb{C}$, entonces una matriz es normal si y solo si es unitariamente equivalente a una matriz diagonal.*

2. Si $\Bbbk = \mathbb{R}$, entonces una matriz es simétrica si y solo si es ortogonalmente equivalente a una matriz diagonal.

Dem.

1. ($\Rightarrow$): Sea $A \in M_n(\mathbb{C})$ y consideremos $L_A : \mathbb{C}^n \to \mathbb{C}^n$. Como A es normal, entonces L_A es normal, luego existe una base ortonormal $\mathcal{B}$ de $\mathbb{C}^n$ tal que $[L_A]_\mathcal{B} = D$, siendo D una matriz diagonal. Sea $\mathcal{C}$ la base canónica de $\mathbb{C}^n$ y $Q = {}_\mathcal{C}[\mathrm{Id}]_\mathcal{B}$. Entonces

$$A = [L_A]_\mathcal{C} = {}_\mathcal{C}[\mathrm{Id}]_\mathcal{B}\, [L_A]_\mathcal{B}\, {}_\mathcal{B}[\mathrm{Id}]_\mathcal{C} = QDQ^{-1}.$$

Como C y $\mathcal{B}$ son bases ortonormales de $\mathbb{C}^n$ y $Q = {}_\mathcal{C}[\mathrm{Id}]_\mathcal{B}$, entonces Q es unitaria y $A = QDQ^*$.

($\Leftarrow$): Sean Q y D en $M_n(\mathbb{C})$, con Q unitaria y D diagonal. Consideremos $A = QDQ^*$. Observar que vale $A^* = (QDQ^*)^* = Q^{**}D^*Q^* = Q\overline{D}Q^*$, luego

$$AA^* = (QDQ^*)\big(Q\overline{D}Q^*\big) = QD\overline{D}Q^* \quad \text{y} \quad A^*A = \big(Q\overline{D}Q^*\big)(QDQ^*) = Q\overline{D}DQ^*.$$

Como D es diagonal, vale $D\overline{D} = \overline{D}D$, luego la fórmula de arriba implica $AA^* = A^*A$.

2. ($\Rightarrow$): Es la misma idea que en 1. Si $A \in M_n(\mathbb{R})$ es simétrica, entonces $L_A \in \mathcal{L}(\mathbb{R}^n)$ es autoadjunta y por lo tanto es diagonalizable en una base ortonormal. El resto sigue igual.

($\Leftarrow$): Sea $A = QDQ^t$, con Q ortogonal y D diagonal. Es $A^t = \big(QDQ^t\big)^t = Q^{tt}D^tQ^t = QDQ^t = A$, luego $A^t = A$. $\square$

Observación 3.4.19. Respecto al enunciado del teorema anterior, notar que si $A \in M_n(\mathbb{R})$ es ortogonalmente equivalente a una matriz diagonal D, entonces A y D son semejantes. Luego las entradas diagonales de D son los valores propios de A. Lo mismo ocurre en el caso complejo.

Ejemplo 3.4.20. Sea $A = \begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{pmatrix}$. Como A es una matriz simétrica real, sabemos que es ortogonalmente equivalente a una matriz diagonal. Vamos a hallar una matriz ortogonal Q y una matriz diagonal D tales que $A = QDQ^t$.
Es un ejercicio el verificar que vale $\chi_A(t) = -(t-2)^2(t-8)$ y que los subespacios propios de A son

$$E_2 = [(-1,1,0),(-1,0,1)], \quad E_8 = [(1,1,1)].$$

Aplicando Gram-Schmidt a la base $\{(-1,1,0),(-1,0,1)\}$ de E_2 obtenemos

$$E_2 = [(-1/2,1,-1/2),(-1,0,1)] = [(-1,2,-1),(-1,0,1)]\,.$$

Luego $\{(-1,2,-1),(-1,0,1),(1,1,1)\}$ es una base ortogonal de $\mathbb{R}^3$. Normalizando obtenemos

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

que es una base ortonormal de $\mathbb{R}^3$ formada por vectores propios de A.
Entonces si $D = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix}$ y $Q = \begin{pmatrix} -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix}$, es $A = QDQ^t$; es decir, vale

$$\begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{pmatrix} = \begin{pmatrix} -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & \frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \end{pmatrix} \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 8 \end{pmatrix} \begin{pmatrix} -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} & -\frac{1}{\sqrt{6}} \\ -\frac{1}{\sqrt{2}} & 0 & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{pmatrix}.$$

Una aplicación del teorema anterior es el siguiente resultado.

Proposición 3.4.21. *Sean $A, B \in M_n(\mathbb{R})$ matrices simétricas. Si A y B son semejantes, entonces son ortogonalmente equivalentes.*

Dem. Si A y B son semejantes, entonces tienen los mismos valores propios $\lambda_1 \ldots, \lambda_n$ (pueden haber repetidos). Como A y B son simétricas, entonces como vimos en el teorema anterior vale que A y B son ortogonalmente equivalentes a la matriz diagonal $D = \operatorname{diag}(\lambda_1 \ldots, \lambda_n)$. Luego A y B son ortogonalmente equivalentes entre sí. $\square$

Observación 3.4.22. El resultado anterior es falso si se quita la condición de que las matrices sean simétricas. Por ejemplo, los valores propios de $A = \left(\begin{smallmatrix} 2 & 1 \\ 0 & 3 \end{smallmatrix}\right)$ son 2 y 3, luego A es semejante a $B = \left(\begin{smallmatrix} 2 & 0 \\ 0 & 3 \end{smallmatrix}\right)$. Sin embargo, como B es diagonal, entonces es fácil de probar que para toda matriz $Q \in M_2(\mathbb{R})$ vale que QBQ^t es una matriz simétrica. Como A no es simétrica, entonces A no puede ser ortogonalmente equivalente con B.

Isometrías en $\mathbb{R}^2$ y $\mathbb{R}^3$. Finalizamos esta sección clasificando las isometrías de $\mathbb{R}^2$ y $\mathbb{R}^3$ (como siempre, el producto interno es el producto escalar). Recordar que si $T \in \mathcal{L}(\mathbb{R}^n)$ es una isometría, entonces $\det T = \pm 1$ y los valores propios de T, en caso de existir, valen ± 1.

Proposición 3.4.23. *Sea $W \subset \mathbb{R}^3$ un subespacio de dimensión 2 (es decir, un plano que pasa por el origen). Si $T \in \mathcal{L}(W)$ es una isometría, entonces T es una rotación si $\det T = 1$ o una simetría axial si $\det T = -1$.*

Dem. Sea $\mathcal{C} = \{u_1, u_2\}$ una base ortonormal de W. Entonces $[T]_\mathcal{C} = \left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ es una matriz ortogonal. Supongamos $\det T = 1$. Sabemos que las columnas de $[T]_\mathcal{C}$ forman una base ortonormal de $\mathbb{R}^2$, luego (b, d) es ortogonal con (a, c). Como $(-c, a)$ es no nulo y también es ortogonal con (a, c), deducimos que existe $k \in \mathbb{R}$ tal que $(b, d) = k(-c, a)$. La condición $\det T = 1$ junto con $1 = \|(a, c)\|^2 = a^2 + c^2$ implican $k = 1$, luego $(b, d) = (-c, a)$. Como $a^2 + c^2 = 1$, entonces existe un único $\theta \in [0, 2\pi)$ tal que $a = \cos\theta$ y $c = \operatorname{sen}\theta$. Luego

$$[T]_\mathcal{C} = \begin{pmatrix} a & -c \\ c & a \end{pmatrix} = \begin{pmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{pmatrix} \quad \Rightarrow \quad \begin{cases} T(u_1) = (\cos\theta)u_1 + (\operatorname{sen}\theta)u_2, \\ T(u_2) = -(\operatorname{sen}\theta)u_1 + (\cos\theta)u_2 \end{cases}.$$

Esto nos dice que T es una rotación en W de centro en el origen y ángulo θ.
Si $\det T = -1$, es $\chi_T(t) = t^2 - (a + d)t - 1$. El discriminante de la ecuación $t^2 - (a + d)t - 1 = 0$ es $\Delta = (a+d)^2 + 4 \geq 4$. Como Δ es positivo, entonces $\chi_T(t)$ tiene dos raíces reales cuyo producto es -1. Como los valores propios de A solo pueden ser ± 1, deducimos que existe $\mathcal{B} = \{w_1, w_2\}$ base ortonormal de W que verifica

$$T(w_1) = w_1, \quad T(w_2) = -w_2.$$

Luego T es la simetría axial en W de eje la recta $[w_1]$. □

Identificando $\mathbb{R}^2$ con el plano $W = \{(x, y, z) \in \mathbb{R}^3 : \ z = 0\}$, obtenemos el siguiente resultado.

Corolario 3.4.24. *Si T es una isometría de $\mathbb{R}^2$, entonces T es una rotación si $\det T = 1$ o una simetría axial si $\det T = -1$.* □

A continuación clasificaremos las isometrías en $\mathbb{R}^3$. Para eso necesitamos introducir las siguientes isometrías.

1. Una *rotación* es un operador $T \in \mathcal{L}(\mathbb{R}^3)$ tal que existe una base ortonormal $\mathcal{B} = \{w_1, w_2, w_3\}$ de $\mathbb{R}^3$ y un número real θ, tal que

   $$T(w_1) = (\cos\theta)w_1 + (\operatorname{sen}\theta)w_2, \quad T(w_2) = -(\operatorname{sen}\theta)w_1 + (\cos\theta)w_2, \quad T(w_3) = w_3.$$

   Notar que T deja fija a la recta $l := [w_3]$ y restringida a $W := [w_1, w_2] = [w_3]^\perp$ es una rotación del plano W. En este caso diremos que T es una rotación de eje la recta l y ángulo θ.

2. Una *simetría especular* es un operador $T \in \mathcal{L}(\mathbb{R}^3)$ tal que existe una base ortonormal $\mathcal{B} = \{w_1, w_2, w_3\}$ de $\mathbb{R}^3$ tal que

   $$T(w_1) = w_1, \quad T(w_2) = w_2, \quad T(w_3) = -w_3.$$

   Notar que T restringida al plano $W := [w_1, w_2]$ es la identidad y restringida $W^\perp = [w_3]$ es menos la identidad. En este caso diremos que T es una simetría especular respecto al plano W.

3. Una *simetría rotacional* es la composición de una simetría especular respecto a un plano W con una rotación de eje una recta l y ángulo θ, de forma tal que l sea perpendicular con W. Notar que en este caso existe una base ortonormal $\mathcal{B} = \{w_1, w_2, w_3\}$ de $\mathbb{R}^3$ tal que

   $$T(w_1) = (\cos\theta)w_1 + (\operatorname{sen}\theta)w_2, \quad T(w_2) = -(\operatorname{sen}\theta)w_1 + (\cos\theta)w_2, \quad T(w_3) = -w_3.$$

Observación 3.4.25. Las matrices asociadas a una rotación, simetría especular y simetría rotacional en las bases ortonormales anteriores son, repectivamente,

$$\begin{pmatrix} \cos\theta & -\operatorname{sen}\theta & 0 \\ \operatorname{sen}\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{pmatrix}, \quad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{pmatrix}, \quad \begin{pmatrix} \cos\theta & -\operatorname{sen}\theta & 0 \\ \operatorname{sen}\theta & \cos\theta & 0 \\ 0 & 0 & -1 \end{pmatrix}.$$

Notar que estas matrices son ortogonales, lo cual prueba que los operadores son isometrías.

Proposición 3.4.26. *Si T es una isometría de $\mathbb{R}^3$, entonces T es una rotación si $\det T = 1$, o una simetría especular o rotacional si $\det T = -1$.*

Dem. Como el polinomio característico de T es de grado 3, entonces el teorema de Bolzano implica que tiene alguna raíz real λ. Además, como λ es un valor propio de T, es $\lambda = \pm 1$. Sea v_0 un vector propio de T correspondiente a λ y $W = [v_0]^\perp$. Veamos que W es T invariante.

$$\begin{aligned}\text{Si } w \in W \;&\Rightarrow\; 0 = \langle w, v_0\rangle = \langle T(w), T(v_0)\rangle = \langle T(w), \lambda v_0\rangle = \langle T(w), \pm v_0\rangle = \pm\langle T(w), v_0\rangle \\ &\Rightarrow\; \langle T(w), v_0\rangle = 0 \;\Rightarrow\; T(w) \in W.\end{aligned}$$

Notar que es $\mathbb{R}^3 = [v_0] \oplus W$, y que $[v_0]$ y W son T-invariantes. Si $\mathcal{C} = \{w_1, w_2\}$ es una base de W, entonces $\mathcal{B} = \{v_0, w_1, w_2\}$ es una base de $\mathbb{R}^3$ y[3]

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda & 0 & 0 \\ 0 & a & b \\ 0 & c & d \end{pmatrix}, \quad [T|_W]_{\mathcal{C}} = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

esto implica $\det T = \lambda \det(T|_W)$. Como T es una isometría sabemos que valen $\lambda = \pm 1$ y $\det T = \pm 1$. Discutiremos según estas cuatro posibilidades, teniendo en cuenta la proposición 3.4.23.

1. Si $\det T = 1$ y $\lambda = 1$, entonces $\det(T|_W) = 1$. Luego $T|_W$ es una rotación en el plano W, y como además T deja fija a la recta $[v_0]$, deducimos que T es una rotación en el espacio de eje $[v_0]$.

2. Si $\det T = 1$ y $\lambda = -1$, entonces $\det(T|_W) = -1$. Luego $T|_W$ es una simetría axial en W. Esto implica que existe una base ortonormal $\{u_1, u_2\}$ de W tal que $T(u_1) = u_1$ y $T(u_2) = -u_2$. Como vale $T(v_0) = -v_0$, obtenemos que T deja fija a la recta $[u_1]$ y en el plano $[v_0, u_2]$ es una rotación de ángulo π. Luego T es una rotación en $\mathbb{R}^3$ de eje $[u_1]$ y ángulo π.

3. Si $\det T = -1$ y $\lambda = 1$, entonces $\det(T|_W) = -1$. Luego $T|_W$ es una simetría axial en W y T deja fija a la recta $[v_0]$. Tomando una base ortonormal $\{u_1, u_2\}$ de W como en el caso anterior, obtenemos $T(v_0) = v_0$, $T(u_1) = u_1$ y $T(u_2) = -u_2$. Luego T es una simetría especular respecto al plano $[v_0, u_1]$.

4. Si $\det T = -1$ y $\lambda = -1$, entonces $\det(T|_W) = 1$. Luego $T|_W$ es una rotación en el plano W y por lo tanto T es una rotación en el espacio de eje $[v_0]$ compuesta con una simetría especular respecto a W, lo cual es (por definición) una simetría rotacional. $\square$

[3]Notar que si cambiamos el orden escribiendo $\mathcal{B} = \{w_1, w_2, v_0\}$, entonces $[T]_{\mathcal{B}}$ queda con el bloque 2×2 al principio, de la misma forma que las matrices de la observación 3.4.25.

Capítulo 4

Formas bilineales simétricas

En este capítulo trabajaremos siempre en espacios vectoriales reales de dimensión finita[1]. Además asumiremos que las bases están ordenadas.

4.1. Formas bilineales

Sea V un espacio vectorial. Una *forma bilineal* en V es una función $\varphi : V \times V \to \mathbb{R}$ que es lineal en cada una de sus variables, es decir, que verifica

$$\varphi(au+v,w) = a\varphi(u,w) + \varphi(v,w), \quad \varphi(w,au+v) = a\varphi(w,u) + \varphi(w,v),$$

para todo $a \in \mathbb{R}$, $u, v, w \in V$. Al conjunto de las formas bilineales en V lo escribiremos $\operatorname{Bil}(V)$.

Es un ejercicio el verificar que el conjunto de las formas bilineales es un subespacio del espacio de las funciones con dominio $V \times V$ y codominio $\mathbb{R}$, en el cual las operaciones se definen punto a punto:

$$(f+g)(u,v) = f(u,v) + g(u,v), \quad (af)(u,v) = af(u,v), \quad \forall a \in \mathbb{R},\ u,v \in V.$$

Luego $\operatorname{Bil}(V)$ con las operaciones anteriores es un $\mathbb{R}$-espacio vectorial.

Ejemplo 4.1.1. Un producto interno en un espacio vectorial real $\langle\ ,\ \rangle : V \times V \to \mathbb{R}$ es una forma bilineal. Notar que un producto interno en un espacio vectorial complejo no es una forma bilineal.

Definición 4.1.2. Sea $A \in M_n(\mathbb{R})$. Definimos $\beta_A : \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ por $\beta_A(x,y) = x^t A y$, siendo $x, y \in \mathbb{R}^n$ vectores columna (es decir estamos pensando $\mathbb{R}^n = M_{n\times 1}(\mathbb{R})$).

Las propiedades del producto de matrices implican directamente que β_A es una forma bilineal. En coordenadas, si escribimos

$$x = \begin{pmatrix} x_1 \\ \vdots \\ x_n \end{pmatrix}, \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix}, \quad A = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & & \vdots \\ a_{n1} & \cdots & a_{nn} \end{pmatrix},$$

entonces

$$\beta_A\big((x_1,\cdots,x_n),(y_1,\cdots,y_n)\big) = \sum_{i,j=1}^n a_{ij} x_i y_j, \quad \forall (x_1,\cdots,x_n),(y_1,\cdots,y_n) \in \mathbb{R}^n. \tag{4.1}$$

[1] Para la teoría general no es necesario que el cuerpo sean los reales, aunque en ese caso hay que poner restricciones sobre el mismo (que tenga característica distinta de 2). Pero como ciertos resultados son específicos para los reales, y son los que más nos interesan, para simplificar asumiremos desde el principio que el cuerpo es $\mathbb{R}$.

Ejemplo 4.1.3. Si consideramos la matriz $A = \left(\begin{smallmatrix} 2 & 3 \\ 4 & -1 \end{smallmatrix}\right)$ entonces $\beta_A : \mathbb{R}^2 \times \mathbb{R}^2 \to \mathbb{R}$ está definida por

$$\begin{aligned} &\beta_A\big((x_1,x_2),(y_1,y_2)\big) = \\ &= (x_1,x_2)\begin{pmatrix} 2 & 3 \\ 4 & -1 \end{pmatrix}\begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = (x_1,x_2)\begin{pmatrix} 2y_1+3y_2 \\ 4y_1-y_2 \end{pmatrix} = 2x_1y_1+3x_1y_2+4x_2y_1-x_2y_2. \end{aligned}$$

Luego $\beta_A\big((x_1,x_2),(y_1,y_2)\big) = 2x_1y_1+3x_1y_2+4x_2y_1-x_2y_2$, para todo $(x_1,x_2),(y_1,y_2) \in \mathbb{R}^2$.

El siguiente resultado permite construir nuevas formas bilineales a partir de una forma dada. La prueba es directa y queda como ejercicio.

Proposición 4.1.4. *Sea V un espacio vectorial y $\varphi : V \times V \to \mathbb{R}$ una forma bilineal. Sea W otro espacio y $T, S : W \to V$ dos transformaciones lineales. Si definimos $\psi : W \times W \to \mathbb{R}$ mediante*

$$\psi(w_1,w_2) = \varphi\big(T(w_1), S(w_2)\big), \quad \forall w_1, w_2 \in W,$$

Entonces ψ es una forma bilineal en W. □

De ahora en adelante V es un espacio de dimensión n.

Observación 4.1.5. Sea $\mathcal{B}$ una base de V y $A \in M_n(\mathbb{R})$. El mapa coordenadas $\operatorname{coord}_{\mathcal{B}} : V \to \mathbb{R}^n$ es un isomorfismo lineal, luego aplicando la proposición anterior a $\varphi = \beta_A \in \operatorname{Bil}(\mathbb{R}^n)$ obtenemos que el mapa $\psi : V \times V \to \mathbb{R}$ definido por

$$\psi(u,v) = \operatorname{coord}_{\mathcal{B}}(u)^t A \operatorname{coord}_{\mathcal{B}}(v), \quad \forall u, v \in V,$$

es una forma bilineal en V. A continuación probaremos que toda forma bilineal en V se puede escribir de esa manera. Para eso necesitamos la siguiente definición.

Definición 4.1.6. Sea $\varphi \in \operatorname{Bil}(V)$ y $\mathcal{B} = \{v_1, \ldots, v_n\}$ una base de V. La *matriz asociada* a la forma bilineal φ en la base $\mathcal{B}$ es la matriz $M_{\mathcal{B}}(\varphi) \in M_n(\mathbb{R})$ definida por

$$M_{\mathcal{B}}(\varphi) := \begin{pmatrix} \varphi(v_1,v_1) & \cdots & \varphi(v_1,v_n) \\ \vdots & & \vdots \\ \varphi(v_n,v_1) & \cdots & \varphi(v_n,v_n) \end{pmatrix}.$$

Llamaremos *representación matricial* de φ toda matriz asociada a φ en alguna base de V.

Observación 4.1.7. Si $A \in M_n(\mathbb{R})$ y $\varphi = \beta_A \in \operatorname{Bil}(\mathbb{R}^n)$, entonces $A = M_{\mathcal{C}}(\varphi)$, siendo $\mathcal{C}$ la base canónica de $\mathbb{R}^n$. Luego toda matriz cuadrada se puede pensar como la representación matricial de alguna forma bilineal.

Proposición 4.1.8. *Sea $\varphi \in \operatorname{Bil}(V)$ y $\mathcal{B}$ una base de V. Entonces*

$$\varphi(u,v) = \operatorname{coord}_{\mathcal{B}}(u)^t M_{\mathcal{B}}(\varphi) \operatorname{coord}_{\mathcal{B}}(v), \quad \forall u, v \in V.$$

Además, si una matriz $A \in M_n(\mathbb{R})$ verifica

$$\varphi(u,v) = \operatorname{coord}_{\mathcal{B}}(u)^t A \operatorname{coord}_{\mathcal{B}}(v), \quad \forall u, v \in V, \tag{4.2}$$

entonces $A = M_{\mathcal{B}}(\varphi)$.

Dem. Sean $u = \sum_{i=1}^n x_i v_i$ y $v = \sum_{j=1}^n y_j v_j$. Entonces por la bilinealidad de φ es

$$\varphi(u,v) = \varphi\left(\sum_{i=1}^n x_i v_i, \sum_{j=1}^n y_j\, v_j\right) = \sum_{i,j=1}^n x_i y_j \varphi(v_i, v_j) = \operatorname{coord}_{\mathcal{B}}(u)^t M_{\mathcal{B}}(\varphi)\operatorname{coord}_{\mathcal{B}}(v), \quad \forall u, v \in V.$$

Esto prueba la primera afirmación. Para la segunda, si $A = (a_{ij}) \in M_n(\mathbb{R})$ verifica la fórmula (4.2) y $\mathcal{B} = \{v_1, \ldots, v_n\}$, entonces

$$\varphi(v_i, v_j) = \operatorname{coord}_{\mathcal{B}}(v_i)^t A \operatorname{coord}_{\mathcal{B}}(v_j) = e_i^t A e_j = a_{ij}, \quad \forall i, j,$$

siendo $\{e_1, \ldots, e_n\}$ la base canónica de $\mathbb{R}^n$. Luego $A = M_{\mathcal{B}}(\varphi)$. □

Corolario 4.1.9. *Si $\varphi \in \operatorname{Bil}(\mathbb{R}^n)$ entonces existe una única matriz $A \in M_n(\mathbb{R})$ tal que $\varphi = \beta_A$.*

Dem. Sea $\mathcal{C}$ la base canónica de $\mathbb{R}^n$ y $A = M_{\mathcal{C}}(\varphi)$. Entonces

$$\varphi(u,v) = \operatorname{coord}_{\mathcal{C}}(u)^t M_{\mathcal{C}}(\varphi)\operatorname{coord}_{\mathcal{C}}(v) = u^t A v = \beta_A(u,v), \quad \forall u, v \in V.$$

La unicidad de A se deduce de la unicidad en la proposición 4.1.8. □

Proposición 4.1.10. *Sea $\mathcal{B}$ una base de V. Definimos $\Theta : \operatorname{Bil}(V) \to M_n(\mathbb{R})$, mediante $\Theta(\varphi) = M_{\mathcal{B}}(\varphi)$, para toda $\varphi \in \operatorname{Bil}(V)$. Entonces $\Theta : \operatorname{Bil}(V) \to M_n(\mathbb{R})$ es un isomorfismo. Luego* $\dim \operatorname{Bil}(V) = n^2$.

Dem. Veamos que Θ es una transformación lineal. Sean $\varphi, \psi \in \operatorname{Bil}(V)$ y $a \in \mathbb{R}$. Entonces

$$\begin{aligned}(\varphi + a\psi)(u,v) &= \varphi(u,v) + a\psi(u,v) = \operatorname{coord}_{\mathcal{B}}(u)^t M_{\mathcal{B}}(\varphi)\operatorname{coord}_{\mathcal{B}}(v) + a \cdot \operatorname{coord}_{\mathcal{B}}(u)^t M_{\mathcal{B}}(\psi)\operatorname{coord}_{\mathcal{B}}(v)\\ &= \operatorname{coord}_{\mathcal{B}}(u)^t \big(M_{\mathcal{B}}(\varphi) + aM_{\mathcal{B}}(\psi)\big)\operatorname{coord}_{\mathcal{B}}(v), \quad \forall u, v \in V.\end{aligned}$$

Esto implica $M_{\mathcal{B}}(\varphi + a\psi) = M_{\mathcal{B}}(\varphi) + aM_{\mathcal{B}}(\psi)$, luego Θ es lineal. Además, por lo que vimos anteriormente, dada $A \in M_n(\mathbb{R})$, si definimos $\varphi : V \times V \to \mathbb{R}$ mediante (4.2), entonces $\varphi \in \operatorname{Bil}(V)$ y $M_{\mathcal{B}}(\varphi) = A$. Esto implica que Θ es sobreyectiva. La inyectividad de Θ se deduce de la primera parte de la proposición 4.1.8. □

El siguiente resultado relaciona las matrices asociadas a una misma forma bilineal en bases distintas.

Proposición 4.1.11. *Sean $\mathcal{B}$ y $\mathcal{C}$ dos bases de V y $\varphi \in \operatorname{Bil}(V)$. Entonces*

$$M_{\mathcal{C}}(\varphi) = \big({}_{\mathcal{B}}[\operatorname{Id}]_{\mathcal{C}}\big)^t M_{\mathcal{B}}(\varphi)\, {}_{\mathcal{B}}[\operatorname{Id}]_{\mathcal{C}}.$$

Dem. Sean $u, v \in V$, entonces usando la fórmula de cambio de coordenadas de $\mathcal{C}$ a $\mathcal{B}$ obtenemos

$$\begin{aligned}\varphi(u,v) &= \operatorname{coord}_{\mathcal{B}}(u)^t M_{\mathcal{B}}(\varphi)\operatorname{coord}_{\mathcal{B}}(v) = \big({}_{\mathcal{B}}[\operatorname{Id}]_{\mathcal{C}}\operatorname{coord}_{\mathcal{C}}(u)\big)^t M_{\mathcal{B}}(\varphi)\big({}_{\mathcal{B}}[\operatorname{Id}]_{\mathcal{C}}\operatorname{coord}_{\mathcal{C}}(v)\big)\\ &= \operatorname{coord}_{\mathcal{C}}(u)^t \big({}_{\mathcal{B}}[\operatorname{Id}]_{\mathcal{C}}\big)^t M_{\mathcal{B}}(\varphi)\, {}_{\mathcal{B}}[\operatorname{Id}]_{\mathcal{C}} \operatorname{coord}_{\mathcal{C}}(v).\end{aligned}$$

Como esto vale para todo $u, v \in V$, entonces la tesis se deduce aplicando la proposición 4.1.8. □

Definición 4.1.12. Dos matrices A y B en $M_n(\mathbb{R})$ se dicen *congruentes* si existe una matriz invertible Q en $M_n(\mathbb{R})$ tal que $A = Q^t B Q$.

Proposición 4.1.13. *La congruencia es una relación de equivalencia en $M_n(\mathbb{R})$.*

Dem. La afirmación se deduce de los cálculos siguientes:

$$A = I^t A I; \quad \text{si } A = Q^t B Q \;\Rightarrow\; B = \big(Q^t\big)^{-1} A Q^{-1} = \big(Q^{-1}\big)^t A Q^{-1};$$
$$\text{si } A = Q^t B Q \text{ y } B = P^t C P \;\Rightarrow\; A = Q^t \big(P^t C P\big) Q = Q^t P^t C P Q = (PQ)^t C P Q. \quad \square$$

Proposición 4.1.14. *Sea $\varphi \in \mathrm{Bil}(V)$ y $\mathcal{B}$ una base de V. Si $A \in M_n(\mathbb{R})$ es congruente con $M_{\mathcal{B}}(\varphi)$, entonces existe $\mathcal{C}$ base de V tal que $M_{\mathcal{C}}(\varphi) = A$.*

Dem. Sea $Q = (q_{ij}) \in M_n(\mathbb{R})$ invertible tal que $A = Q^t M_{\mathcal{B}}(\varphi)\, Q$. Si $\mathcal{B} = \{v_1, \ldots, v_n\}$, definimos $\mathcal{C} = \{w_1, \ldots, w_n\}$ mediante

$$w_j = \sum_{i=1}^{n} q_{ij} v_i, \quad j = 1, \ldots, n.$$

Como Q es invertible, resulta que $\mathcal{C}$ es una base de V y ${}_{\mathcal{B}}[\mathrm{Id}]_{\mathcal{C}} = Q$. Luego

$$A = Q^t M_{\mathcal{B}}(\varphi)\, Q = \left({}_{\mathcal{B}}[\mathrm{Id}]_{\mathcal{C}}\right)^t M_{\mathcal{B}}(\varphi)\, {}_{\mathcal{B}}[\mathrm{Id}]_{\mathcal{C}} = M_{\mathcal{C}}(\varphi). \quad \square$$

Corolario 4.1.15. *Dos matrices son congruentes si y solo si representan a una misma forma bilineal.*

Dem. Sean $A, A' \in M_n(\mathbb{R})$ dos matrices congruentes. Consideramos la forma bilineal $\beta_{A'} \in \mathrm{Bil}(\mathbb{R}^n)$. Si $\mathcal{B}$ es la base canónica de $\mathbb{R}^n$, entonces es $A' = M_{\mathcal{B}}(\beta_{A'})$. Luego A es congruente con $M_{\mathcal{B}}(\beta_{A'})$ y la proposición anterior implica que es $A = M_{\mathcal{C}}(\beta_{A'})$ para alguna base $\mathcal{C}$ de $\mathbb{R}^n$. El recíproco es la Proposición 4.1.11. $\square$

Observaciones 4.1.16.

1. Si $A, B \in M_n(\mathbb{R})$ y $Q \in M_n(\mathbb{R})$ es una matriz invertible tal que $A = Q^t B Q$, entonces $A = PBP^t$, siendo $P = Q^t$ una matriz invertible, y es claro que estas dos condiciones son equivalentes. Luego tenemos dos formas equivalentes de definir la congruencia (lo mismo sucede con la semejanza). Nosotros usaremos la primera porque se adapta mejor al trabajo con formas bilineales.

2. Hay dos relaciones de equivalencia en $M_n(\mathbb{R})$ que son similares y pueden dar lugar a confusión. Sean $A, B \in M_n(\mathbb{R})$. Por un lado, usando la observación anterior podemos decir que A y B son congruentes si existe una matriz invertible $Q \in M_n(\mathbb{R})$ tal que $A = QBQ^t$. Pero por otro lado decimos que A y B son ortogonalmente equivalentes si existe una matriz ortogonal $P \in M_n(\mathbb{R})$ tal que $A = PBP^t$ (recordar que P es ortogonal si es invertible y $P^{-1} = P^t$). Luego dos matrices ortogonalmente equivalentes son también congruentes. El recíproco no es necesariamente cierto, dado que dos matrices congruentes no tienen porqué tener los mismos valores propios. Si las matrices tienen distintos valores propios, entonces no pueden ser semejantes y por lo tanto no van a ser ortogonalmente equivalentes.

4.2. Formas bilineales simétricas

Sea V un espacio de dimensión n.

Definición 4.2.1. Una forma bilineal $\varphi : V \times V \to \mathbb{R}$ se dice *simétrica* si verifica

$$\varphi(u, v) = \varphi(v, u), \quad \forall u, v \in V.$$

Ejemplo 4.2.2. Todo producto interno real $\langle\ ,\ \rangle : V \times V \to \mathbb{R}$ es una forma bilineal simétrica.

Proposición 4.2.3. *Sea $\varphi \in \mathrm{Bil}(V)$ y $\mathcal{B}$ una base de V. Entonces la forma bilineal φ es simétrica si y solo si la matriz $M_{\mathcal{B}}(\varphi)$ es simétrica.*

Dem. Sea $\mathcal{B} = \{v_1, \ldots, v_n\}$ y $M_{\mathcal{B}}(\varphi) = (a_{ij})$. Si φ es simétrica, entonces

$$a_{ij} = \varphi(v_i, v_j) = \varphi(v_j, v_i) = a_{ji}, \quad \forall i, j = 1, \ldots, n$$

luego $M_{\mathcal{B}}(\varphi) = (a_{ij})$ es simétrica.

Recíprocamente, supongamos ahora que $M_{\mathcal{B}}(\varphi)$ es simétrica. Luego vale $\varphi(v_i,v_j)=\varphi(v_j,v_i)$, para todo i,j. Sean $u,v\in V$ arbitrarios, entonces existen escalares $a_i,b_i\in\mathbb{R}$, $i=1,\ldots,n$ tales que $u=\sum_{i=1}^n a_iv_i$ y $v=\sum_{i=1}^n b_iv_i$. Luego

$$\varphi(u,v)=\varphi\left(\sum_{i=1}^n a_iv_i,\sum_{j=1}^n b_jv_j\right)=\sum_{i=1}^n a_i\sum_{j=1}^n b_j\varphi(v_i,v_j)=\sum_{i,j=1}^n a_ib_j\varphi(v_i,v_j)$$
$$\varphi(v,u)=\varphi\left(\sum_{j=1}^n b_jv_j,\sum_{i=1}^n a_iv_i\right)=\sum_{j=1}^n b_j\sum_{i=1}^n a_i\varphi(v_j,v_i)=\sum_{j,i=1}^n b_ja_i\varphi(v_j,v_i)=\sum_{j,i=1}^n a_ib_j\varphi(v_j,v_i)\,.$$

Como es $\varphi(v_i,v_j)=\varphi(v_j,v_i)$, para todo i,j, entonces en las sumas anteriores los sumandos coinciden y por lo tanto es $\varphi(u,v)=\varphi(v,u)$, para todo $u,v\in V$. □

Al conjunto de las formas bilineales simétricas en V lo escribiremos $\mathrm{Bil}_S(V)$.

Observación 4.2.4. Si consideramos una base $\mathcal{B}$ de V y el isomorfismo $\Theta:\mathrm{Bil}(V)\to M_n(\mathbb{R})$ dado por $\varphi\mapsto M_{\mathcal{B}}(\varphi)$, entonces la proposición anterior muestra que la imagen por Θ de $\mathrm{Bil}_S(V)$ es el subespacio de $M_n(\mathbb{R})$ formado por las matrices simétricas. Esto implica que $\mathrm{Bil}_S(V)$ es un subespacio de $\mathrm{Bil}(V)$ y que la dimensión de $\mathrm{Bil}_S(V)$ es $\frac{n(n+1)}{2}$.

Ejemplo 4.2.5. Como la matriz identidad $I\in M_n(\mathbb{R})$ es simétrica, entonces $\beta_I\in\mathrm{Bil}(\mathbb{R}^n)$ es una forma bilineal simétrica. Explícitamente obtenemos

$$\beta_I\big((x_1,\ldots,x_n),(y_1,\ldots,y_n)\big)=x_1y_1+\cdots+x_ny_n.$$

Luego β_I es el producto escalar en $\mathbb{R}^n$.

Definición 4.2.6. Si $\varphi:V\times V\to\mathbb{R}$ es una forma bilineal simétrica, entonces la función $\Phi:V\to\mathbb{R}$ definida por $\Phi(v)=\varphi(v,v)$ para todo $v\in V$, se llama la *forma cuadrática* asociada a φ.

Ejemplo 4.2.7. Si V es un espacio real y $\varphi=\langle\ ,\ \rangle$ es un producto interno en V, entonces la forma cuadrática asociada es el cuadrado de la norma: $\Phi(v)=\|v\|^2$, para todo $v\in V$.

Ejemplo 4.2.8. Consideremos $\varphi\in\mathrm{Bil}(\mathbb{R}^2)$ definida por

$$\varphi\big((x,y),(x',y')\big)=2xx'+3xy'+3yx'-5yy'.$$

Observar que podemos escribir

$$\varphi\big((x,y),(x',y')\big)=\begin{pmatrix}x & y\end{pmatrix}\begin{pmatrix}2 & 3\\ 3 & -5\end{pmatrix}\begin{pmatrix}x'\\ y'\end{pmatrix}.$$

Luego es $\varphi=\beta_A$, con $A=\left(\begin{smallmatrix}2 & 3\\ 3 & -5\end{smallmatrix}\right)$. Como A es simétrica, deducimos que φ es una forma bilineal simétrica. La forma cuadrática asociada a φ es la función $\Phi:\mathbb{R}^2\to\mathbb{R}$ definida por

$$\Phi(x,y)=2x^2+6xy-5y^2,\quad\forall(x,y)\in\mathbb{R}^2.$$

Observación 4.2.9. Como las formas bilineales simétricas en V forman un subespacio del espacio de las funciones de $V\times V$ en $\mathbb{R}$, entonces se prueba fácilmente que las formas cuadráticas en V forman un subespacio del espacio de las funciones de V en $\mathbb{R}$.

Proposición 4.2.10. *Sea $\varphi \in \operatorname{Bil}_S(V)$ y $\Phi : V \to \mathbb{R}$ la forma cuadrática asociada a φ. Entonces*

1. *$\Phi(av) = a^2\Phi(v)$, para todo $a \in \mathbb{R},\ v \in V$. Luego $\Phi(0) = 0$.*

2. *$\Phi(u+v) = \Phi(u) + 2\varphi(u,v) + \Phi(v)$, para todo $u, v \in V$.*

Dem. La prueba de la primera afirmación es simple y queda como ejercicio. La segunda se deduce del cálculo siguiente

$$\Phi(u+v) = \varphi(u+v, u+v) = \varphi(u,u) + \varphi(u,v) + \varphi(v,u) + \varphi(v,v) = \Phi(u) + 2\varphi(u,v) + \Phi(v). \quad \square$$

Observación 4.2.11. Despejando $\varphi(u,v)$ en la segunda fórmula de la proposición anterior, obtenemos la *fórmula de polarización*

$$\varphi(u,v) = \frac{1}{2}\big(\Phi(u+v) - \Phi(u) - \Phi(v)\big), \quad \forall u, v \in V.$$

Luego toda forma bilineal simétrica queda determinada por su forma cuadrática asociada. En este sentido, dada una forma cuadrática Φ, diremos que la correspondiente forma bilineal simétrica φ es la *forma bilineal asociada* a Φ.

Nuestro próximo objetivo es caracterizar las formas cuadráticas en $\mathbb{R}^n$.

Proposición 4.2.12. *Las formas cuadráticas en $\mathbb{R}^n$ son las funciones $\Phi : \mathbb{R}^n \to \mathbb{R}$ del tipo*

$$\Phi(x_1, \dots, x_n) = \sum_{1 \le i \le j \le n} b_{ij} x_i x_j, \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n,$$

siendo los b_{ij} escalares arbitrarios.

Dem. Si $\Phi : \mathbb{R}^n \to \mathbb{R}$ es una forma cuadrática, entonces existe $\varphi \in \operatorname{Bil}_S(\mathbb{R}^n)$ tal que $\Phi(v) = \varphi(v,v)$, para todo $v \in \mathbb{R}^n$. Como $\varphi \in \operatorname{Bil}_S(\mathbb{R}^n)$, entonces existe una matriz simétrica $A = (a_{ij}) \in M_n(\mathbb{R})$ tal que $\varphi = \beta_A$. Luego Φ está definida por $\Phi(x_1, \dots, x_n) = \sum_{i,j=1}^n a_{ij} x_i x_j$. Como vale $a_{ij} = a_{ji}$, para todo i, j, entonces podemos escribir

$$\Phi(x_1, \dots, x_n) = \sum_{i \le j} b_{ij} x_i x_j, \quad b_{ij} := \begin{cases} a_{ii} & \text{si } i = j, \\ 2a_{ij} & \text{si } i < j. \end{cases}$$

Es claro que este proceso podemos revertirlo, definiendo los a_{ij} a partir de los b_{ij} mediante $a_{ii} := b_{ii}$, para todo i y $a_{ij} = a_{ji} := \frac{1}{2} b_{ij}$, si $i < j$. Luego toda función $\Phi : \mathbb{R}^n \to \mathbb{R}$ definida por una fórmula del tipo $\Phi(x_1, \dots, x_n) = \sum_{i \le j} b_{ij} x_i x_j$ es una forma cuadrática en $\mathbb{R}^n$. $\square$

Observación 4.2.13. Los polinomios en n variables de la forma $p = \sum_{1 \le i \le j \le n} a_{ij} x_i x_j$, siendo los a_{ij} escalares arbitrarios, se llaman polinomios *homogéneos* de grado 2. La proposición anterior muestra que las formas cuadráticas en $\mathbb{R}^n$ son las funciones que están dadas por polinomios homogéneos de grado 2.

Observación 4.2.14. Las formas cuadráticas en $\mathbb{R}^2$ y en $\mathbb{R}^3$ son las funciones del tipo

$$\Phi(x,y) = ax^2 + bxy + cy^2, \quad a, b, c \in \mathbb{R},$$
$$\Phi(x,y,z) = ax^2 + by^2 + cz^2 + dxy + exz + fyz, \quad a, b, \dots, f \in \mathbb{R}.$$

Ejemplo 4.2.15. Consideremos la forma cuadrática Φ en $\mathbb{R}^3$ definida por $\Phi(x,y,z) = 5x^2 + 3z^2 - 2xy + xz$. Recordando la prueba de la proposición anterior, obtenemos que la forma bilineal asociada está definida por

$$\varphi\big((x,y,z),(x',y',z')\big) = 5xx' + 3zz' - xy' - x'y + \frac{1}{2}xz' + \frac{1}{2}x'z.$$

4.3. Diagonalización

De ahora en adelante φ es una forma bilineal simétrica en V y Φ es su forma cuadrática asociada. Para evitar discutir casos triviales, en general asumiremos $V \neq \{0\}$.

Definición 4.3.1. 1. Dos vectores $u, v \in V$ son *φ-ortogonales* si $\varphi(u, v) = 0$.

2. Una base $\mathcal{B} = \{v_1, \dots, v_n\}$ de V se dice *φ-ortogonal* si $\varphi(v_i, v_j) = 0$ para todo $i \neq j$.

Notar que si $\mathcal{B}$ es una base de V, entonces $\mathcal{B}$ es φ-ortogonal si y solo si la matriz $M_{\mathcal{B}}(\varphi)$ es diagonal. Explícitamente, si $\mathcal{B} = \{v_1, \dots, v_n\}$ es una base φ-ortogonal de V, entonces

$$M_{\mathcal{B}}(\varphi) = \begin{pmatrix} \Phi(v_1) & 0 & \cdots & 0 \\ 0 & \Phi(v_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi(v_n) \end{pmatrix}.$$

Luego las expresiones de φ y Φ con coordenadas en esta base son

$$\varphi(u, v) = \sum_{i=1}^{n} \Phi(v_i) x_i y_i, \quad \Phi(u) = \sum_{i=1}^{n} \Phi(v_i) x_i^2,$$

siendo $u = \sum_{i=1}^{n} x_i v_i$ y $v = \sum_{i=1}^{n} y_i v_i$.

Ejemplo 4.3.2. Consideremos la forma cuadrática $\Phi : \mathbb{R}^3 \to \mathbb{R}$ definida por $\Phi(x, y, z) = x^2 - y^2$ y sea $\varphi \in \mathrm{Bil}_S(\mathbb{R}^3)$ la forma bilineal correspondiente. Si $\mathcal{B}$ es la base canónica de $\mathbb{R}^3$, es

$$M_{\mathcal{B}}(\varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Luego $\mathcal{B}$ es una base φ-ortogonal.

Definición 4.3.3. Si W es un subespacio de V, la *restricción de φ a W* es la función $\varphi|_{W \times W} : W \times W \to \mathbb{R}$, definida por $\varphi|_{W \times W}(w_1, w_2) := \varphi(w_1, w_2)$, para todo $w_1, w_2 \in W$. Claramente $\varphi|_{W \times W} \in \mathrm{Bil}_S(W)$.

Teorema 4.3.4. *Si $\varphi \in \mathrm{Bil}_S(V)$, entonces existe una base $\mathcal{B}$ de V tal que $M_{\mathcal{B}}(\varphi)$ es una matriz diagonal.*

Dem. Lo que tenemos que probar es que existe una base φ-ortogonal. La demostración es por inducción en $n = \dim V$. Notar que si $n = 1$, entonces toda base de V es (trivialmente) φ-ortogonal.

Supongamos ahora que vale la tesis cuando la dimensión del espacio es $n - 1$ y sea $\varphi \in \mathrm{Bil}_S(V)$ con $\dim V = n$. Si $\varphi = 0$, entonces toda base de V es φ-ortogonal. Supongamos ahora que es $\varphi \neq 0$. Si fuese $\Phi = 0$, entonces usando la fórmula de polarización obtendríamos $\varphi = 0$, contra lo supuesto. Luego existe $u \in V$ tal que $\Phi(u) \neq 0$. Definimos $\alpha \in V^*$ mediante $\alpha(v) = \varphi(v, u)$, para todo $v \in V$. Observar que $\alpha(u) = \varphi(u, u) = \Phi(u) \neq 0$, luego $\alpha \neq 0$ y por lo tanto $\dim \mathrm{Ker}(\alpha) = n - 1$. Aplicando la hipótesis de inducción a φ restringida a $\mathrm{Ker}(\alpha)$ tenemos que existe $\{v_1, \dots, v_{n-1}\}$ base de $\mathrm{Ker}(\alpha)$ tal que $\varphi(v_i, v_j) = 0$ si $i \neq j$. Como $u \notin \mathrm{Ker}(\alpha) = [v_1, \dots, v_{n-1}]$, entonces el conjunto $\mathcal{B} = \{v_1, \dots, v_{n-1}, u\}$ es LI, y al tener n elementos es base de V. Como $\{v_1, \dots, v_{n-1}\} \subset \mathrm{Ker}(\alpha)$, resulta que $v_1, \dots, v_{n-1}$ son φ-ortogonales con u. Luego $\mathcal{B}$ es una base φ-ortogonal de V. $\square$

Algoritmo de diagonalización Dada $\varphi \in \mathrm{Bil}_S(V)$, el teorema 4.3.4 nos asegura que siempre existe una base de V en la cual la matriz asociada a φ es diagonal. En lo que sigue vamos a ver cómo hallarla.

Sea $\varphi \in \mathrm{Bil}_S(V)$. Empezamos tomando una base cualquiera $\mathcal{C}$ de V y considerando $A = M_{\mathcal{C}}(\varphi) \in M_n(\mathbb{R})$. Si $Q \in M_n(\mathbb{R})$ es una matriz invertible cualquiera y definimos $D = Q^t A Q$, entonces la proposición 4.1.14 nos dice que existe una base $\mathcal{B}$ de V tal que $D = M_{\mathcal{B}}(\varphi)$. Luego para obtener una base φ-ortogonal de V, alcanza con encontrar una matriz invertible Q de forma tal que $D = Q^t A Q$ sea una matriz diagonal. La idea para esto es ir transformando A con matrices elementales $Q_1, \ldots, Q_l$ hasta llegar a una matriz diagonal D,

$$A \sim Q_1^t A Q_1 \sim Q_2^t Q_1^t A Q_1 Q_2 = (Q_1 Q_2)^t A (Q_1 Q_2) \sim \cdots \sim (Q_1 \cdots Q_l)^t A (Q_1 \cdots Q_l) = D.$$

Notar que si Q es una matriz elemental de tipo I, II o III, entonces AQ ($Q^t A$) es la matriz que se obtiene realizando la operación elemental correspondiente en las columnas (filas) de A. Luego $Q^t A Q$ es la matriz que se obtiene realizando la operación elemental correspondiente en las filas y columnas de A.

Resumiendo, lo que necesitamos es realizar operaciones elementales en las filas y columnas de A hasta llegar a una matriz diagonal D. En ese caso va a ser $D = Q^t A Q$, $Q = Q_1 \cdots Q_l$, siendo $Q_1, \ldots, Q_l$ las matrices elementales correspondientes a las operaciones elementales anteriores. Notar que $Q = Q_1 \cdots Q_l$ se obtiene realizando en las columnas de la matriz identidad I, las mismas operaciones elementales que hicimos en las columnas de A.

Mostraremos el método mediante un ejemplo. Consideremos en $\mathbb{R}^3$ la forma cuadrática Φ definida por

$$\Phi(x, y, z) = -7x^2 - 2y^2 + 7z^2 + 8xy + 2xz - 4yz, \quad \forall (x, y, z) \in \mathbb{R}^3.$$

Sea $\varphi \in \mathrm{Bil}_S(\mathbb{R}^3)$ la forma bilineal simétrica asociada a Φ. Si $\mathcal{C}$ es la base canónica de $\mathbb{R}^3$, es

$$A = M_{\mathcal{C}}(\varphi) = \begin{pmatrix} -7 & 4 & 1 \\ 4 & -2 & -2 \\ 1 & -2 & 7 \end{pmatrix}.$$

Queremos encontrar una matriz diagonal D y una matriz invertible Q tales que $D = Q^t A Q$. El método para obtener la matriz Q es el siguiente: escribimos a la izquierda la matriz A y a la derecha la matriz identidad I, luego realizamos operaciones elementales en A y cada vez que hacemos una operación elemental en las colummnas de A, también la hacemos en las columnas de I, pero cuando hacemos operaciones elementales en las filas de A no hacemos nada en la matriz I. Seguimos realizando este proceso hasta obtener del lado izquierdo una matriz diagonal D, en cuyo caso lo que aparece a la derecha es la matriz de congruencia Q. Lo mostraremos en el ejemplo anterior. En lo que sigue, la notación $C_3 - C_2$ quere decir que a la columna

3 le restamos la columna 2, $F_1 + 2F_2$ que a la fila 1 le sumamos la fila 2 multiplicada por 2, etc.

$$\begin{array}{rrr|rrr|c}
-7 & 4 & 1 & 1 & 0 & 0 & \\
4 & -2 & -2 & 0 & 1 & 0 & \\
1 & -2 & 7 & 0 & 0 & 1 & \\
\hline
-7 & 4 & -3 & 1 & 0 & 0 & \\
4 & -2 & 0 & 0 & 1 & -1 & C_3 - C_2 \\
1 & -2 & 9 & 0 & 0 & 1 & \\
\hline
-7 & 4 & -3 & & & & \\
4 & -2 & 0 & & & & F_3 - F_2 \\
-3 & 0 & 9 & & & & \\
\hline
1 & 4 & -3 & 1 & 0 & 0 & \\
0 & -2 & 0 & 2 & 1 & -1 & C_1 + 2C_2 \\
-3 & 0 & 9 & 0 & 0 & 1 & \\
\hline
1 & 0 & -3 & & & & \\
0 & -2 & 0 & & & & F_1 + 2F_2 \\
-3 & 0 & 9 & & & & \\
\hline
1 & 0 & 0 & 1 & 0 & 3 & \\
0 & -2 & 0 & 2 & 1 & 5 & C_3 + 3C_1 \\
-3 & 0 & 0 & 0 & 0 & 1 & \\
\hline
1 & 0 & 0 & & & & \\
0 & -2 & 0 & & & & F_3 + 3F_1 \\
0 & 0 & 0 & & & &
\end{array}$$

Luego $D = Q^t A Q$, siendo $D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ y $Q = \begin{pmatrix} 1 & 0 & 3 \\ 2 & 1 & 5 \\ 0 & 0 & 1 \end{pmatrix}$.

Finalmente, si definimos $\mathcal{B} = \{(1,2,0),\ (0,1,0),\ (3,5,1)\}$, entonces $\mathcal{B}$ es una base de $\mathbb{R}^3$ tal que ${}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}} = Q$. Luego $\mathcal{B}$ es una base φ-ortogonal de $\mathbb{R}^3$ y $M_{\mathcal{B}}(\varphi) = D$.

Bases φ-ortonormales.

Definición 4.3.5. Una base $\mathcal{B} = \{v_1, \ldots, v_n\}$ de V se dice *φ-ortonormal* si es una base φ-ortogonal y además $\Phi(v_i) \in \{-1, 0, 1\}$, para todo $i = 1, \ldots, n$.

Observación 4.3.6. Si $\mathcal{B} = \{v_1, \ldots, v_n\}$ es una base φ-ortonormal de V, entonces la matriz $M_{\mathcal{B}}(\varphi)$ es diagonal

$$M_{\mathcal{B}}(\varphi) = \begin{pmatrix} \Phi(v_1) & 0 & \cdots & 0 \\ 0 & \Phi(v_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi(v_n) \end{pmatrix}$$

y sus entradas diagonales $\Phi(v_1), \ldots, \Phi(v_n)$ son 0, 1 o -1. Notar que, reordenando $v_1, \ldots, v_n$, podemos asumir genéricamente que existen $1 \leq p \leq r \leq n$ tales que

$$\Phi(v_i) = 1,\ \forall i = 1, \ldots, p; \quad \Phi(v_i) = -1,\ \forall i = p+1, \ldots, r; \quad \Phi(v_i) = 0,\ \forall i = r+1, \ldots, n.$$

Luego las expresiones de φ y Φ con coordenadas en esta base son

$$\varphi(u,v) = x_1 y_1 + \cdots + x_p y_p - x_{p+1} y_{p+1} - \cdots - x_r y_r; \quad \Phi(u) = x_1^2 + \cdots + x_p^2 - x_{p+1}^2 - \cdots - x_r^2.$$

siendo $u = \sum_{i=1}^n x_i v_i$ y $v = \sum_{i=1}^n y_i v_i$.

Ejemplo 4.3.7. Si consideramos la forma bilineal $\varphi \in \mathrm{Bil}_S(\mathbb{R}^3)$ asociada a la forma cuadrática $\Phi : \mathbb{R}^3 \to \mathbb{R}$ definida por $\Phi(x, y, z) = x^2 - y^2$, entonces la base canónica es una base φ-ortonormal de $\mathbb{R}^3$ (ejemplo 4.3.2).

Proposición 4.3.8. *Si $\varphi \in \mathrm{Bil}_S(V)$, entonces existe una base φ-ortonormal de V.*

Dem. Sea $\mathcal{B} = \{v_1, \dots, v_n\}$ una base φ-ortogonal de V que suponemos ordenada de forma tal que existen $1 \leq p \leq r \leq n$ tales que

$$\Phi(v_i) > 0,\ \forall i = 1, \dots, p; \quad \Phi(v_i) < 0,\ \forall i = p+1, \dots, r; \quad \Phi(v_i) = 0,\ \forall i = r+1, \dots, n.$$

Luego si definimos $w_1, \dots, w_n$ por

$$w_i = \frac{1}{\sqrt{\Phi(v_i)}} v_i,\ \forall i = 1, \dots, p; \quad w_i = \frac{1}{\sqrt{-\Phi(v_i)}} v_i,\ \forall i = p+1, \dots, r; \quad w_i = v_i,\ \forall i = r+1, \dots, n,$$

entonces es claro que w_i es φ-ortogonal con w_j si $i \neq j$, y vale

$$\Phi(w_i) = 1,\ \forall i = 1, \dots, p; \quad \Phi(w_i) = -1,\ \forall i = p+1, \dots, r; \quad \Phi(w_i) = 0,\ \forall i = r+1, \dots, n.$$

Luego $\mathcal{C} = \{w_1, \dots, w_n\}$ es una base φ-ortonormal de V. $\square$

Ejemplo 4.3.9. En la sección anterior vimos que si consideramos $\Phi : \mathbb{R}^3 \to \mathbb{R}$ definida por

$$\Phi(x, y, z) = -7x^2 - 2y^2 + 7z^2 + 8xy + 2xz - 4yz, \quad \forall (x, y, z) \in \mathbb{R}^3,$$

entonces $\mathcal{B} = \{(1, 2, 0),\ (0, 1, 0),\ (3, 5, 1)\}$ es una base φ-ortogonal y $\Phi(1, 2, 0) = 1$, $\Phi(0, 1, 0) = -2$ y $\Phi(3, 5, 1) = 0$. Luego "normalizando" el segundo vector obtenemos que $\mathcal{B}' = \left\{(1, 2, 0),\ \left(0, \frac{1}{\sqrt{2}}, 0\right),\ (3, 5, 1)\right\}$ es una base φ-ortonormal de $\mathbb{R}^3$ tal que

$$M_{\mathcal{B}'}(\varphi) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Observación 4.3.10. Las bases φ-ortonormales son útiles para fines teóricos, pero a nivel práctico no tienen grandes ventajas sobre las bases φ-ortogonales. Por ejemplo, no hay un análogo de la fórmula $u = \sum_{i=1}^n \langle u, v_i \rangle v_i$, para todo $u \in V$, que vale en productos internos cuando $\{v_1, \dots, v_n\}$ es una base ortonormal.

4.4. Rango, degeneramiento y signatura

En esta sección φ es una forma bilineal simétrica en un espacio V y Φ es la forma cuadrática asociada.

En un producto interno en un espacio V, si $u \in V$ verifica $\langle u, v \rangle = 0$ para todo $v \in V$, entonces $u = 0$. Pero para una forma bilineal simétrica arbitraria esto puede no ocurrir. Por ejemplo si $\varphi \in \mathrm{Bil}(\mathbb{R}^2)$ esta definida por $\varphi\big((x_1, x_2), (y_1, y_2)\big) = x_1y_1 + x_1y_2 + x_2y_1 + x_2y_2$, entonces $\varphi\big((1, -1), (y_1, y_2)\big) = 0$, para todo $(y_1, y_2) \in \mathbb{R}^2$. Esto da lugar a la siguiente definición

Definición 4.4.1. Decimos que φ es *no degenerada* si $\varphi(u, v) = 0$ para todo $v \in V$, implica $u = 0$. En ese caso también decimos que Φ es no degenerada.

Proposición 4.4.2. *Si $\mathcal{B} = \{v_1, \dots, v_n\}$ es una base φ-ortogonal de V, entonces Φ es* no degenerada *si y solo si $\Phi(v_i) \neq 0$, para todo $i = 1, \dots, n$.*

Dem. Como $\mathcal{B} = \{v_1, \ldots, v_n\}$ es una base φ-ortogonal, vale

$$M_{\mathcal{B}}(\varphi) = \begin{pmatrix} \Phi(v_1) & 0 & \cdots & 0 \\ 0 & \Phi(v_2) & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Phi(v_n) \end{pmatrix}.$$

Lo cual implica

$$\varphi(u, v) = \sum_{i=1}^{n} \Phi(v_i) x_i y_i, \quad \Phi(u) = \sum_{i=1}^{n} \Phi(v_i) x_i^2, \tag{4.3}$$

siendo $u = \sum_{i=1}^{n} x_i v_i$ y $v = \sum_{i=1}^{n} y_i v_i$. Si existe algún i tal que $\Phi(v_i) = 0$, entonces $\varphi\Big(v_i, \sum_{j=1}^{n} y_j v_j\Big) = \Phi(v_i) y_i = 0$, para todo $y_1, \ldots, y_n \in \mathbb{R}$. Luego $v_i \neq 0$ verifica $\varphi(v_i, v) = 0$, para todo $v \in V$.

Supongamos ahora que es $\Phi(v_i) \neq 0$, para todo i. Sea $u = \sum_{i=1}^{n} x_i v_i$ tal que $\varphi(u, v) = 0$, para todo $v \in V$. Entonces para cada i, es $0 = \varphi(u, v_i) = \Phi(v_i) x_i$. Como es $\Phi(v_i) \neq 0$, deducimos $x_i = 0$; como esto vale para todo i, concluimos $u = 0$. □

Observación 4.4.3. Si $A, B, P, Q \in M_n(\mathbb{R})$ verifican que P y Q son invertibles y $A = PBQ$, entonces el rango de A coincide con el de B. Esto se prueba considerando las transformaciones lineales $L_A, L_B, L_P, L_Q \in \mathcal{L}(\mathbb{R}^n)$. Como L_P y L_Q son isomorfismos, entonces no es difícil de probar que de $L_A = L_P \circ L_B \circ L_Q$, se deduce $\dim \operatorname{Im}(L_A) = \dim \operatorname{Im}(L_B)$, lo cual equivale a $\operatorname{rango}(A) = \operatorname{rango}(B)$. Esto implica que matrices congruentes tienen el mismo rango y por lo tanto tiene sentido la siguiente definición.

Definición 4.4.4. Definimos el *rango* de φ como el rango de $M_{\mathcal{B}}(\varphi)$, siendo $\mathcal{B}$ una base arbitraria de V.

Observación 4.4.5. Es claro que si $\mathcal{B}$ es una base φ-ortogonal de V, entonces el rango de φ coincide con la cantidad de entradas diagonales no nulas de $M_{\mathcal{B}}(\varphi)$. Luego la proposición anterior implica lo siguiente.

Corolario 4.4.6. *Una forma $\varphi \in \operatorname{Bil}_S(V)$ es no degenerada si y solo si $\operatorname{rango}(\varphi) = \dim V$. Esto equivale a $\det M_{\mathcal{B}}(\varphi) \neq 0$, siendo $\mathcal{B}$ una base arbitraria de V.* □

La siguientes definiciones tiene que ver con el signo de una forma cuadrática.

Definición 4.4.7. Consideremos una forma cuadrática Φ en V.

1. Φ es *definida positiva* si vale $\Phi(v) > 0$, para todo $v \neq 0$.

2. Φ es *definida negativa* si vale $\Phi(v) < 0$, para todo $v \neq 0$.

3. Φ es *definida*, si es definida positiva o es definida negativa.

4. Φ es *semidefinida positiva* si vale $\Phi(v) \geq 0$, para todo $v \in V$ y existe $0 \neq v_0 \in V$ tal que $\Phi(v_0) = 0$.

5. Φ es *semidefinida negativa* si vale $\Phi(v) \leq 0$, para todo $v \in V$ y existe $0 \neq v_0 \in V$ tal que $\Phi(v_0) = 0$.

6. Φ es *semidefinida*, si es semidefinida positiva o es semidefinida negativa.

7. Φ es *no definida* si existen v_1 y v_2 en V tales que $\Phi(v_1) > 0$ y $\Phi(v_2) < 0$.

Estas definiciones se extienden naturalmente a las formas bilineales simétricas, diciendo que una forma bilineal simétrica es definida positiva si lo es su forma cuadrática asociada, etc.

Observación 4.4.8. Una forma bilineal simétrica definida positiva es lo mismo que un producto interno (real).

Proposición 4.4.9. *Si $\mathcal{B} = \{v_1, \ldots, v_n\}$ es una base φ-ortogonal de V, entonces vale lo siguiente.*

1. *Φ es definida positiva si y solo si $\Phi(v_i) > 0$, para todo $i = 1, \ldots, n$.*
2. *Φ es definida negativa si y solo si $\Phi(v_i) < 0$, para todo $i = 1, \ldots, n$.*
3. *Φ es semidefinida positiva si y solo si $\Phi(v_i) \geq 0$, para todo $i = 1, \ldots, n$ y existe i_0 tal que $\Phi(v_{i_0}) = 0$.*
4. *Φ es semidefinida negativa si y solo si $\Phi(v_i) \leq 0$, para todo $i = 1, \ldots, n$ y existe i_0 tal que $\Phi(v_{i_0}) = 0$.*
5. *Φ es no definida si y solo si existen i y j tales que $\Phi(v_i) > 0$ y $\Phi(v_j) < 0$.*

Dem. Todas las afirmaciones son fáciles de probar. Como muestra, probaremos la tercera dejando la demostración de las otras como ejercicio. Si Φ es semidefinida positiva, entonces es claro que vale $\Phi(v_i) \geq 0$, para todo $i = 1, \ldots, n$. Además, si fuese $\Phi(v_i) > 0$, para todo $i = 1, \ldots, n$, entonces usando la segunda fórmula en (4.3) obtenemos que $\Phi(v) = 0$ implica $v = 0$. Luego si existe $0 \neq w \in V$ tal que $\Phi(w) = 0$, entonces existe $i_0 \in \{1, \ldots, n\}$ tal que $\Phi(v_{i_0}) = 0$. El recíproco es inmediato usando la segunda fórmula en (4.3). $\square$

Combinando la proposición anterior con la proposición 4.4.2, obtenemos el siguiente resultado.

Corolario 4.4.10. *Si una forma cuadrática Φ es definida, entonces es no degenerada, y si Φ es semidefinida, entonces es degenerada.* $\square$

Ejemplo 4.4.11. Consideremos las siguientes formas cuadráticas en $\mathbb{R}^3$. Aplicando la proposición anterior obtenemos lo siguiente.

1. $\Phi(x, y, z) = x^2 + y^2 + z^2$, es definida positiva (la forma bilineal asociada es el producto escalar).
2. $\Phi(x, y, z) = -x^2 - y^2 - z^2$, es definida negativa.
3. $\Phi(x, y, z) = x^2 + y^2$, es semidefinida positiva.
4. $\Phi(x, y, z) = -x^2 - y^2$, es semidefinida negativa.
5. $\Phi(x, y, z) = x^2 - y^2$, es no definida y degenerada.
6. $\Phi(x, y, z) = x^2 - y^2 - z^2$, es no definida y no degenerada.

Ejemplo 4.4.12. Sea $\Phi : \mathbb{R}^3 \to \mathbb{R}$ definida por $\Phi(x, y, z) = 4x^2 - y^2$, para todo $(x, y, z) \in \mathbb{R}^3$. Si consideramos las bases $\mathcal{B} = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ y $\mathcal{C} = \{(1, 1, 0), (1, 4, 0), (0, 0, 1)\}$, entonces

$$M_{\mathcal{B}}(\varphi) = \begin{pmatrix} 4 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \quad \text{y} \quad M_{\mathcal{C}}(\varphi) = \begin{pmatrix} 3 & 0 & 0 \\ 0 & -12 & 0 \\ 0 & 0 & 0 \end{pmatrix}.$$

Notar que si bien las matrices $M_{\mathcal{B}}(\varphi)$ y $M_{\mathcal{C}}(\varphi)$ son distintas, en sus diagonales principales ambas tienen la misma cantidad de ceros, de entradas positivas y de entradas negativas. Vamos a probar que esto siempre ocurre.

Lema 4.4.13. *Sea Φ una forma cuadrática en V y φ su forma bilineal asociada. Sea $\mathcal{B} = \{v_1, \ldots, v_n\}$ una base φ-ortogonal de V, que podemos considerar ordenada de forma tal que*

$$\Phi(v_i) > 0, \ \forall i = 1, \ldots, p; \quad \Phi(v_i) < 0, \ \forall i = p + 1, \ldots, r; \quad \Phi(v_i) = 0, \ \forall i = r + 1, \ldots, n,$$

para ciertos[2] $1 \leq p \leq r \leq n$. *Entonces*

[2]Podría no existir ningún i tal que $\Phi(v_i) > 0$. En ese caso el lema prueba que si Φ es definida positiva en W, entonces $W = \{0\}$. Tampoco se afecta la prueba si ningún $\Phi(v_i)$ es negativo o nulo.

1. *La forma cuadrática Φ es definida positiva en $V_+ := [v_1, \ldots, v_p]$.*

2. *Si W es un subespacio de V tal que Φ es definida positiva en W, entonces $\dim W \leq \dim V_+ = p$.*

Dem. La primera afirmación se deduce directamente de la proposición anterior.

Para la segunda, observar primero que si definimos $V_{\leq 0} := [v_{p+1}, \ldots, v_n]$, entonces es $V = V_+ \oplus V_{\leq 0}$. Además, usando (4.3) obtenemos que si $v = \sum_{i=p+1}^{n} x_i v_i \in V_{\leq 0}$, entonces $\Phi(v) = \sum_{i=p+1}^{r} \Phi(v_i) x_i^2 \leq 0$; luego $\Phi(v) \leq 0$, para todo $v \in V_{\leq 0}$. Consideremos ahora W como en la hipótesis. Sea $w \in W \cap V_{\leq 0}$. Como Φ es definida positiva en W, entonces $\Phi(w) \geq 0$. Por otro lado, $w \in V_{\leq 0}$ implica $\Phi(w) \leq 0$. Luego es $\Phi(w) = 0$ y al ser Φ definida positiva en W, concluimos que es $w = 0$. Esto implica $W \cap V_{\leq 0} = \{0\}$. Luego W y $V_{\leq 0}$ son subespacios independientes y por lo tanto

$$\dim W + \dim V_{\leq 0} = \dim(W_+ \oplus V_{\leq 0}) \leq \dim V = \dim V_+ + \dim V_{\leq 0} \quad \Rightarrow \quad \dim W \leq \dim V_+. \quad \square$$

Teorema 4.4.14 (Ley de inercia de Sylvester)**.** *La cantidad de entradas diagonales positivas, negativas y nulas de una matriz diagonal asociada a una forma cuadrática, no depende de la representación diagonal.*

Dem. Sea Φ una forma cuadrática en V y φ su forma bilineal asociada. Sean $\mathcal{B} = \{v_1, \ldots, v_n\}$ y $\mathcal{C} = \{w_1, \ldots, w_n\}$ dos bases φ-ortogonales de V. Lo primero es observar que la cantidad de entradas diagonales no nulas de $M_{\mathcal{B}}(\varphi)$ concide con las de $M_{\mathcal{C}}(\varphi)$ y es el rango r de φ. Esto implica que la cantidad de entradas diagonales nulas de $M_{\mathcal{B}}(\varphi)$ concide con las de $M_{\mathcal{C}}(\varphi)$. Siempre podemos suponer que las bases $\mathcal{B}$ y $\mathcal{C}$ están ordenadas de forma tal que

$$\begin{array}{lll} \Phi(v_i) > 0, \ \forall i = 1, \ldots, p; & \Phi(v_i) < 0, \ \forall i = p+1, \ldots, r; & \Phi(v_i) = 0, \ \forall i = r+1, \ldots, n, \\ \Phi(w_i) > 0, \ \forall i = 1, \ldots, q; & \Phi(w_i) < 0, \ \forall i = q+1, \ldots, r; & \Phi(w_i) = 0, \ \forall i = r+1, \ldots, n, \end{array}$$

para ciertos p, q. Sean

$$V_+ = [v_1, \ldots, v_p] \quad \text{y} \quad W_+ = [w_1, \ldots, w_q].$$

La primera parte del lema anterior aplicada a la base $\mathcal{C}$, nos dice que Φ es definida positiva en W_+. Luego la segunda parte del lema aplicada a la base $\mathcal{B}$ implica $\dim W_+ \leq \dim V_+$. Intercambiando los roles de las bases $\mathcal{B}$ y $\mathcal{C}$, obtenemos $\dim V_+ \leq \dim W_+$; luego $\dim W_+ = \dim V_+$. Entones probamos que la cantidad de entradas diagonales positivas de $M_{\mathcal{B}}(\varphi)$ concide con las de $M_{\mathcal{C}}(\varphi)$. Pero como en ambos casos la suma de la cantidad de entradas diagonales positivas y negativas da el rango de φ, entonces concluimos que también la cantidad de entradas diagonales negativas de $M_{\mathcal{B}}(\varphi)$ concide con las de $M_{\mathcal{C}}(\varphi)$. $\square$

Definición 4.4.15. Sea Φ una forma cuadrática en V y φ su forma bilineal simétrica asociada. Sea $\mathcal{B}$ una base φ-ortogonal de V. Llamaremos *signatura* de Φ a la terna (n_+, n_-, n_0), siendo n_+ la cantidad de entradas diagonales positivas de $M_{\mathcal{B}}(\varphi)$, n_- la cantidad de entradas diagonales negativas y n_0 la cantidad de entradas diagonales nulas.

Observación 4.4.16. Notar que con las notaciones de la definición anterior, es $\operatorname{rango}(\Phi) = n_+ + n_-$ y $n_+ + n_- + n_0 = \dim V$; luego $n_0 = \dim V - \operatorname{rango}(\Phi)$.

Observación 4.4.17. Los nombres anteriores no están muy estandarizados. Por ejemplo algunos autores llaman *signatura* a la terna (n_0, n_+, n_-) y otros llaman *signatura* a la resta $n_+ - n_-$. Nosotros convendremos que la signatura es (n_+, n_-, n_0), dado que ese orden coincide con el que hemos estado usando en las representaciones matriciales diagonales.

Ejemplo 4.4.18. Si consideramos la forma $\Phi : \mathbb{R}^3 \to \mathbb{R}$ definida por $\Phi(x, y, z) = 4x^2 - y^2$, que vimos en el ejemplo 4.4.12, entonces su signatura es $(1, 1, 1)$.

Terminamos este capítulo con algunos resultados que se obtienen a partir de que sabemos que toda matriz simétrica real es ortogonalmente equivalente a una matriz diagonal (teorema 3.4.18).

Teorema 4.4.19. *Sea $\varphi \in \mathrm{Bil}_S(V)$. Entonces las cantidades de entradas diagonales positivas, negativas y nulas de una representación matricial diagonal arbitraria de φ coinciden respectivamente con las cantidades de valores propios positivos, negativos y nulos de una representación matricial arbitraria de φ.*

Dem. Sea $\mathcal{B}$ una base arbitraria de V y $A = M_{\mathcal{B}}(\varphi)$. La matriz A es simétrica real, luego existen matrices reales D, Q tales que D es diagonal, Q es ortogonal y $D = Q^{-1}AQ = Q^tAQ$.
Como D es congruente con $A = M_{\mathcal{B}}(\varphi)$, entonces existe $\mathcal{C}$ base de V tal que $D = M_{\mathcal{C}}(\varphi)$. Por otro lado como A y D son semejantes, entonces tienen los mismos valores propios, los cuales coinciden con las entradas diagonales de D. Luego encontramos una base $\mathcal{C}$ tal que $D = M_{\mathcal{C}}(\varphi)$ es diagonal y las entradas diagonales de D son los valores propios de A. Finalmente la tesis se obtiene aplicando la ley de inercia de Sylvester. $\square$

Corolario 4.4.20. *Sea $A \in M_n(\mathbb{R})$ una matriz simétrica. Entonces las cantidades de entradas diagonales positivas, negativas y nulas de cualquier matriz diagonal congruente con A, coinciden respectivamente con las cantidades de valores propios positivos, negativos y nulos de A.*

Dem. La matriz A es una representación matricial de $\beta_A \in \mathrm{Bil}_S(\mathbb{R}^n)$ (en la base canónica), luego cualquier matriz diagonal que sea congruente con A es también una representación matricial de β_A y por lo tanto se aplica el teorema anterior. $\square$

Observación 4.4.21. Combinando este corolario con el algoritmo de la sección 4.3, se obtiene una forma simple de conocer la cantidad de valores propios positivos, negativos y nulos de una matriz simétrica real.

Teorema 4.4.22. *Sea $\varphi \in \mathrm{Bil}_S(V)$, siendo V un espacio vectorial real con producto interno. Entonces existe una base ortonormal $\mathcal{B}$ de V tal que $M_{\mathcal{B}}(\varphi)$ es diagonal.*

Dem. Sea $\mathcal{C}$ una base ortonormal cualquiera de V y $A = M_{\mathcal{C}}(\varphi)$. La matriz A es simétrica real, luego existen matrices $D, Q \in M_n(\mathbb{R})$ tales que D es diagonal, Q es ortogonal y $D = Q^{-1}AQ = Q^tAQ$.
Como Q es invertible, entonces existe $\mathcal{B}$ base de V tal que $Q = {}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}}$. Luego

$$D = Q^tAQ = \left({}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}}\right)^t M_{\mathcal{C}}(\varphi) {}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}} = M_{\mathcal{B}}(\varphi).$$

Además, como ${}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}}$ es ortogonal y la base $\mathcal{C}$ es ortonormal, entonces $\mathcal{B}$ es una base ortonormal. En resumen, encontramos una base ortonormal $\mathcal{B}$ tal que $M_{\mathcal{B}}(\varphi) = D$ es una matriz diagonal. $\square$

Observación 4.4.23. El teorema anterior es un resultado de diagonalización simultánea. En esencia lo que nos dice es que si en un espacio vectorial real de dimensión finita tenemos dos formas bilineales simétricas y una de ellas es definida positiva, entonces existe una base en la cual ambas se diagonalizan.

Capítulo 5

Operadores en espacios de dimensión finita

En el capítulo 1 estudiamos la diagonalización de operadores en espacios de dimensión finita. En este capítulo y los siguientes profundizamos en el tema y nos adentramos en el caso no diagonalizable.

En ete capítulo V es un $\Bbbk$-espacio vectorial no nulo de dimensión finita.

5.1. Subespacios invariantes

Sea $T \in \mathcal{L}(V)$. Recordemos que un subespacio W de V se dice *T-invariante* si verifica $T(W) \subset W$. En este caso escribimos $T|_W : W \to W$ a la restricción de T a W, es decir a la función definida por $T|_W(w) = T(w)$ para todo $w \in W$. Notar que si W es T-invariante, entonces $T|_W \in \mathcal{L}(W)$.

Definición 5.1.1. Dado $T \in \mathcal{L}(V)$ y $n = 0, 1, \dots$, definimos $T^n \in \mathcal{L}(V)$ por recurrencia mediante

$$T^0 = \mathrm{Id}, \quad T^n = T \circ T^{n-1},\ n \geq 1.$$

Notar que vale $T^n = \underbrace{T \circ \cdots \circ T}_{n}$, para todo $n = 1, 2, \dots$.

Observación 5.1.2. Es facil de probar que vale $T^n \circ T^m = T^{n+m}$, para todo $n, m \in \mathbb{N}$.

Proposición 5.1.3. *Los subespacios $\{0\}$, V, $\mathrm{Ker}(T)$ e $\mathrm{Im}(T)$ son T-invariantes. Si $\lambda \in \Bbbk$ es un valor propio de T, entonces el subespacio propio $E_\lambda = \mathrm{Ker}(T - \lambda \mathrm{Id})$ es T-invariante.*

Dem. Ejercicio. $\square$

Proposición 5.1.4. *Supongamos que tenemos una descomposición $V = W_1 \bigoplus \cdots \bigoplus W_h$ en que cada W_i es T-invariante. Si $\mathcal{B}_i$ es una base de W_i, $i = 1, \dots, h$, y definimos $\mathcal{B} = \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_h$, entonces $[T]_\mathcal{B}$ es una matriz en bloques de la forma*

$$[T]_\mathcal{B} = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_h \end{pmatrix}, \quad A_i = [T|_{W_i}]_{\mathcal{B}_i}, \ \forall i = 1, \dots, h. \quad \square$$

Dem. Ejercicio. $\square$

Ejemplo 5.1.5. Consideremos $T \in \mathcal{L}(\mathbb{R}^3)$ definida por $T(x,y,z) = (x-z, 2y, x+z)$, para todo $(x,y,z) \in \mathbb{R}^3$. Notar que valen $T(x,0,z) = (x-z, 0, x+z)$ y $T(0,y,0) = (0,2y,0)$, luego $W_1 = \{(x,y,z) \in \mathbb{R}^3 : \ y = 0\}$ y $W_2 = \{(x,y,z) \in \mathbb{R}^3 : \ x = z = 0\}$ son subespacios T-invariantes. Si $\{e_1, e_2, e_3\}$ es la base canónica de $\mathbb{R}^3$, entonces es claro que $\mathcal{B}_1 = \{e_1, e_3\}$ es base de W_1 y $\mathcal{B}_2 = \{e_2\}$ es base de W_2. Si consideramos $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 = \{e_1, e_3, e_2\}$, entonces obtenemos

$$[T]_{\mathcal{B}} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}, \quad [T|_{W_1}]_{\mathcal{B}_1} = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}, \quad [T|_{W_2}]_{\mathcal{B}_2} = \begin{pmatrix} 2 \end{pmatrix}$$

Notar que es $\chi_T(t) = -(t-2)\big(t^2 - 2t + 2\big)$ y este polinomio no se escinde; luego T no es diagonalizable.

5.2. Polinomios evaluados en operadores

Sea $T \in \mathcal{L}(V)$. Si $p(t) = \sum_{i=0}^n a_i\, t^i \in \Bbbk[t]$, definimos $p(T) \in \mathcal{L}(V)$ mediante

$$p(T) := \sum_{i=0}^{n} a_i T^i = a_n\, T^n + a_{n-1}\, T^{n-1} + \cdots + a_1\, T + a_0 \mathrm{Id}.$$

Ejemplo 5.2.1. Si consideramos los polinomios $p(t) = 1$ y $q(t) = t$, entonces $p(T) = \mathrm{Id}$ y $q(T) = T$.

Proposición 5.2.2. *Sean $\lambda \in \Bbbk$ y $p(t), q(t) \in \Bbbk[t]$. Si $r(t) = \lambda\, p(t) + q(t)$ y $s(t) = p(t)q(t)$, entonces*

$$r(T) = \lambda\, p(T) + q(T), \qquad s(T) = p(T) \circ q(T).$$

Dem. Probaremos solo la segunda igualdad, la prueba de la primera es análoga y más fácil, así que queda como ejercicio. Además, para simplificar la notación, vamos a suponer que $p(t)$ y $q(t)$ tienen grado uno, aunque la prueba vale en general. Sean entonces $p(t) = a_0 + a_1 t$ y $q(t) = b_0 + b_1 t$. Entonces

$$s(t) = (a_0 + a_1 t)(b_0 + b_1 t) = a_0 b_0 + (a_0 b_1 + a_1 b_0)t + a_1 b_1 t^2 \quad \Rightarrow \quad s(T) = a_0 b_0 \mathrm{Id} + (a_0 b_1 + a_1 b_0)T + a_1 b_1 T^2.$$

Por otro lado es $p(T) = a_0 \mathrm{Id} + a_1 T$ y $q(T) = b_0 \mathrm{Id} + b_1 T$, luego

$$\begin{aligned} p(T) \circ q(T) &= (a_0 \mathrm{Id} + a_1 T) \circ (b_0 \mathrm{Id} + b_1 T) = (a_0 \mathrm{Id}) \circ (b_0 \mathrm{Id} + b_1 T) + (a_1 T) \circ (b_0 \mathrm{Id} + b_1 T) \\ &= (a_0 \mathrm{Id}) \circ (b_0 \mathrm{Id}) + (a_0 \mathrm{Id}) \circ (b_1 T) + (a_1 T) \circ (b_0 \mathrm{Id}) + (a_1 T) \circ (b_1 T) \\ &= a_0 b_0 \mathrm{Id} + a_0 b_1 T + a_1 b_0 T + a_1 b_1 T^2 = a_0 b_0 \mathrm{Id} + (a_0 b_1 + a_1 b_0) T + a_1 b_1 T^2. \end{aligned}$$

Comparando con la fórmula de $s(t)$ de arriba, deducimos $s(T) = p(T) \circ q(T)$. $\square$

Como el producto de polinomios es conmutativo, la proposición anterior implica el siguiente resultado.

Corolario 5.2.3. *Si $p(t), q(t) \in \Bbbk[t]$, entonces $p(T) \circ q(T) = q(T) \circ p(T)$.* $\square$

Proposición 5.2.4. *Si $p(t) \in \Bbbk[t]$, entonces $\operatorname{Ker} p(T)$ e $\operatorname{Im} p(T)$ son subespacios T-invariantes.*

Dem. Sea $v \in \operatorname{Ker} p(T)$, entonces

$$p(T)(T(v)) = (p(T) \circ T)(v) = (T \circ p(T))(v) = T(p(T)(v)) = T(0) = 0 \ \Rightarrow \ T(v) \in \operatorname{Ker} p(T).$$

Sea $u = p(T)(w) \in \operatorname{Im} p(T)$, entonces

$$T(u) = T(p(T)(w)) = (T \circ p(T))\,(w) = (p(T) \circ T)\,(w) = p(T)(T(w)) \ \Rightarrow \ T(u) \in \operatorname{Im} p(T). \quad \square$$

Observación 5.2.5. Si $T \in \mathcal{L}(V)$, $\lambda \in \Bbbk$ y $k \in \mathbb{N}$, entonces aplicando la proposición anterior al polinomio $p(t) = (t-\lambda)^k$, obtenemos que los subespacios $\operatorname{Ker}(T - \lambda \mathrm{Id})^k$ y $\operatorname{Im}(T - \lambda \mathrm{Id})^k$ son T-invariantes.. Esto lo aplicaremos con frecuencia más adelante.

Polinomios evaluados en matrices. Como es de esperar, todo lo que vimos anteriormente para operadores se aplica también a matrices. A continuación desarrollamos brevemente la versión matricial.

Si $A \in M_n(\Bbbk)$ y $p = \sum_{i=0}^n a_i\, t^i \in \Bbbk[t]$, entonces definimos $p(A) \in M_n(\Bbbk)$ mediante

$$p(A) = \sum_{i=0}^{n} a_i A^i = a_n\, A^n + a_{n-1}\, A^{n-1} + \cdots + a_1\, A + a_0 I.$$

Proposición 5.2.6. *Sea $T \in \mathcal{L}(V)$ y $\mathcal{B}$ una base de V. Entonces $[p(T)]_\mathcal{B} = p([T]_\mathcal{B})\,,\ \forall p \in \Bbbk[t]$.*

Dem. Si $p = \sum_{i=0}^n a_i\, t^i$, es $[p(T)]_\mathcal{B} = \left[\sum_{i=0}^n a_i\, T^i\right]_\mathcal{B} = \sum_{i=0}^n a_i\, ([T]_\mathcal{B})^i = p([T]_\mathcal{B})\,.$ ☐

Corolario 5.2.7. *Si $A \in M_n(\Bbbk)$, es $p(L_A) = L_{p(A)} \in \mathcal{L}(\Bbbk^n)$ para todo $p \in \Bbbk[t]$.*

Dem. Si $\mathcal{B}$ es la base canónica de $\Bbbk^n$, es $A = [L_A]_\mathcal{B}$. Aplicando la proposición anterior obtenemos $[p(L_A)]_\mathcal{B} = p(A)$, luego $p(L_A) = L_{p(A)}$. ☐

Ejemplo 5.2.8. Sea $T \in \mathcal{L}\big(\mathbb{R}^2\big)$ definida por $T(x, y) = (x, 2x + y)$ y consideremos $p = t^3 + t^2 - t + 2 \in \mathbb{R}[t]$. Es $T = L_A$, siendo $A = \left(\begin{smallmatrix} 1 & 0 \\ 2 & 1 \end{smallmatrix}\right)$. Luego

$$p(A) = A^3 + A^2 - A + 2I = \begin{pmatrix} 3 & 0 \\ 8 & 3 \end{pmatrix} \quad \Rightarrow \quad p(T)(x, y) = (3x,\, 8x + 3y),\ \forall (x, y) \in \mathbb{R}^2.$$

Observación 5.2.9. La proposición 5.2.6 y el corolario 5.2.7 nos permiten deducir propiedades de polinomios aplicados a operadores, a partir de propiedades de polinomios aplicados a matrices, y viceversa. La proposición siguiente se obtiene aplicando lo anterior.

Proposición 5.2.10. *Sea $A \in M_n(\Bbbk)$.*

1. *Dados $\lambda \in \Bbbk$ y $p(t), q(t) \in \Bbbk[t]$, si definimos $r(t) = \lambda\, p(t) + q(t)$ y $s(t) = p(t)q(t)$, entonces valen $r(A) = \lambda\, p(A) + q(A)$ y $s(A) = p(A)\, q(A)$.*
2. *Vale $p(A)\, q(A) = q(A)\, p(A)$, para todo $p(t), q(t) \in \Bbbk[t]$.* ☐

5.3. El teorema de Cayley-Hamilton

Sea $T \in \mathcal{L}(V)$. En esta sección probaremos que el polinomio característico de T se anula en T.

Proposición 5.3.1. *Sea W un subespacio T-invariante de V. Si $\mathcal{B}_W$ es una base de W y la completamos a una base $\mathcal{B}$ de V, entonces $[T]_\mathcal{B} = \left(\begin{smallmatrix} A & B \\ 0 & D \end{smallmatrix}\right)$, siendo $A = [T|_W]_{\mathcal{B}_W}$.*

Dem. Ejercicio. ☐

Proposición 5.3.2. *Si $W \subset V$ es un subespacio T-invariante, entonces $\chi_{T|_W}(t)$ divide a $\chi_T(t)$.*

Dem. Sea $\mathcal{B}$ una base de V como en la proposición anterior. Entonces es $[T]_\mathcal{B} = \left(\begin{smallmatrix} A & B \\ 0 & D \end{smallmatrix}\right)$, siendo $A = [T|_W]_{\mathcal{B}_W}$. Luego aplicando la proposición 8.2.1 obtenemos

$$\chi_T(t) = \begin{vmatrix} A - t\, I & B \\ 0 & D - t\, I \end{vmatrix} = |A - t\, I| \times |D - t\, I| = \chi_{T|_W}(t)\, p(t).$$

Luego $\chi_T(t) = \chi_{T|_W}(t)\, p(t)$, siendo $p(t) = |D - t\, I|$. ☐

Ejemplo 5.3.3. Sea $T \in \mathcal{L}(\mathbb{R}^4)$ definida por

$$T(x,y,z,t) = (x+y+2z-t,\ y+t,\ 2z-t,\ z+t).$$

Consideremos el subespacio $W = \{(x,y,0,0) :\ x,y \in \mathbb{R}\}$. Observar que $T(x,y,0,0) = (x+y,\ y,\ 0,\ 0) \in W$, luego W es T-invariante. Si $\mathcal{B} = \{e_1,e_2,e_3,e_4\}$ es la base canónica de $\mathbb{R}^4$, entonces $\mathcal{B}_W = \{e_1,e_2\}$ es base de W y obtenemos

$$[T|_W]_{\mathcal{B}_W} = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix}, \quad [T]_{\mathcal{B}} = \left(\begin{array}{cc|cc} 1 & 1 & 2 & -1 \\ 0 & 1 & 0 & 1 \\ \hline 0 & 0 & 2 & -1 \\ 0 & 0 & 1 & 1 \end{array}\right).$$

Luego $\chi_{T|_W}(t) = (1-t)^2$ y $\chi_T(t) = (1-t)^2\big(t^2-3t+3\big)$.

Definición 5.3.4. Sea $v \in V$. Llamamos *subespacio T-cíclico* generado por v a

$$S_{v,T} := \big[v,\ T(v),\ T^2(v),\ \ldots\big] = \left\{\sum_{i=0}^{n} a_i\, T^i(v) :\ a_i \in \Bbbk,\ i=0,\ldots,n,\ n \in \mathbb{N}\right\} = \{p(T)(v) :\ p(t) \in \Bbbk[t]\}.$$

Proposición 5.3.5. *El subespacio $S_{v,T}$ es el menor subespacio T-invariante de V que contiene a v.*

Dem. Es claro que $v \in S_{v,T}$. Por otro lado $T\big(\sum_{i=0}^n a_i\, T^i(v)\big) = \sum_{i=0}^n a_i\, T^{i+1}(v)$, luego $S_{v,T}$ es T-invariante.

Sea W un subespacio T-invariante que contiene a v. Como W es T-invariante y $v \in W$, entonces $T(v) \in W$. Luego $T^2(v) = (T \circ T)(v) = T(T(v)) \in W$ y por inducción se prueba que $T^n(v) \in W$, para todo $n \in \mathbb{N}$. Como W es un subespacio, esto implica $S_{v,T} = \big[v,\ T(v),\ T^2(v),\ \ldots\big] \subset W$. $\square$

Proposición 5.3.6. *Sea $0 \neq v \in V$, $W = S_{v,T}$ y $h = \dim W$. Entonces:*

1. *El conjunto $\big\{v, T(v), \ldots, T^{h-1}(v)\big\}$ es base de W.*
2. *Si escribimos $T^h(v) = a_0\, v + a_1\, T(v) + \cdots + a_{h-1} T^{h-1}(v)$, entonces*

$$\chi_{T|_W}(t) = (-1)^h\Big(t^h - a_{h-1}\, t^{h-1} - \cdots - a_1\, t - a_0\Big).$$

Dem. Sea j el menor entero positivo tal que $\big\{v, T(v), \ldots, T^j(v)\big\}$ es LD; como $v \neq 0$, es $j \geq 1$. Observar que $\big\{v, T(v), \ldots, T^{j-1}(v)\big\}$ es LI y $\big\{v, T(v), \ldots, T^{j-1}(v), T^j(v)\big\}$ es LD, luego $T^j(v) \in \big[v, T(v), \ldots, T^{j-1}(v)\big]$.

Afirmación: Para todo $p \in \mathbb{N}$ se cumple que $T^{j+p}(v) \in \big[v, T(v), \ldots, T^{j-1}(v)\big]$.

Lo probaremos por inducción en p. Sabemos que es cierto para $p=0$. Supongamos que vale $T^{j+p}(v) \in \big[v, T(v), \ldots, T^{j-1}(v)\big]$. Entonces podemos escribir

$$T^j(v) = \sum_{i=0}^{j-1} a_i T^i(v), \quad T^{j+p}(v) = \sum_{i=0}^{j-1} b_i T^i(v),$$

para ciertos escalares a_i, b_i. Luego

$$\begin{aligned} T^{j+p+1}(v) &= T\big(T^{j+p}(v)\big) = \sum_{i=0}^{j-1} b_i T^{i+1}(v) = \sum_{i=1}^{j} b_{i-1} T^i(v) = b_{j-1} T^j(v) + \sum_{i=1}^{j-1} b_{i-1} T^i(v) \\ &= b_{j-1}\left(\sum_{i=0}^{j-1} a_i T^i(v)\right) + \sum_{i=1}^{j-1} b_{i-1} T^i(v) = b_{j-1} a_0 v + \sum_{i=1}^{j-1} (b_{j-1} a_i + b_{i-1})\, T^i(v). \end{aligned}$$

Esto termina la inducción. Luego vale $T^n(v) \in \big[v, T(v), T^2(v), \dots, T^{j-1}(v)\big]$, para todo $n \in \mathbb{N}$ y por lo tanto $W = \big[v, T(v), \dots, T^{j-1}(v)\big]$, $j = h = \dim W$ y el conjunto $\mathcal{B} = \big\{v, T(v), T^2(v), \dots, T^{h-1}(v)\big\}$ es una base de W. Esto prueba la primer afirmación. Para la segunda, observar que si escribimos $T^h(v) = a_0\, v + a_1\, T(v) + \cdots + a_{h-1} T^{h-1}(v)$, entonces

$$[T|_W]_{\mathcal{B}} = \begin{pmatrix} 0 & 0 & \cdots & 0 & a_0 \\ 1 & 0 & \cdots & 0 & a_1 \\ 0 & 1 & \cdots & 0 & a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & 0 & a_{h-2} \\ 0 & 0 & \cdots & 1 & a_{h-1} \end{pmatrix} \quad \Rightarrow \quad \chi_{T|_W}(t) = \begin{vmatrix} -t & 0 & \cdots & 0 & a_0 \\ 1 & -t & \cdots & 0 & a_1 \\ 0 & 1 & \cdots & 0 & a_2 \\ \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & \cdots & -t & a_{h-2} \\ 0 & 0 & \cdots & 1 & a_{h-1} - t \end{vmatrix}.$$

Usando esta última fórmula se obtiene la fórmula para $\chi_{T|_W}(t)$ de la tesis (la misma se puede probar por inducción en h, desarrollando el determinante por la primera fila o columna). $\square$

Teorema 5.3.7 (Cayley-Hamilton)**.** *El polinomio característico de T se anula en T,* i. e. $\chi_T(T) = 0$.

Dem. Lo que tenemos que probar es que $\chi_T(T) : V \to V$ verifica $\chi_T(T)\,(v) = 0$ para todo v en V.

Sea $v \in V$ arbitrario fijo. Si $v = 0$ el resultado es obvio. Supongamos ahora $v \neq 0$ y sea $W = S_{v,T}$. Como W es T-invariante, entonces $\chi_{T|_W}$ divide a χ_T. Sea $p(t) \in \mathbb{k}[t]$ tal que $\chi_T(t) = p(t)\, \chi_{T|_W}(t)$. Sean $h = \dim W$ y $T^h(v) = a_0\, v + a_1\, T(v) + \cdots + a_{h-1}\, T^{h-1}(v)$. Sabemos por la proposición anterior que vale $\chi_{T|_W}(t) = (-1)^h \big(t^h - a_{h-1}\, t^{h-1} - a_1\, t - \cdots - a_0\big)$. Luego

$$\chi_{T|_W}(T)\,(v) = (-1)^h \Big(T^h(v) - a_{h-1}\, T^{h-1}(v) - \cdots a_1\, T(v) - a_0\, v\Big) = 0.$$

Entonces

$$\chi_T(T)\,(v) = \big(p(T) \circ \chi_{T|_W}(T)\big)\ (v) = p(T)\,\big(\chi_{T|_W}(T)\,(v)\big) = p(T)\,(0) = 0. \quad \square$$

Corolario 5.3.8. *Si $A \in M_n(\mathbb{k})$, entonces $\chi_A(A) = 0$.*

Dem. Sea $\mathcal{C}$ la base canónica de $\mathbb{k}^n$. Si consideramos $T = L_A \in \mathcal{L}(\mathbb{k}^n)$, es $[T]_{\mathcal{C}} = A$ y $\chi_T(t) = \chi_A(t)$. Luego

$$\chi_A(A) = \chi_A([T]_{\mathcal{C}}) = [\chi_A(T)]_{\mathcal{C}} = [\chi_T(T)]_{\mathcal{C}} = [0]_{\mathcal{C}} = 0. \quad \square$$

5.4. Subespacios propios generalizados.

Sea $T \in \mathcal{L}(V)$. Para cada escalar λ definimos el *subespacio propio generalizado* W_λ mediante

$$W_\lambda := \Big\{v \in V :\ (T - \lambda \mathrm{Id})^k(v) = 0,\ \text{para algún } k \geq 1\Big\} = \bigcup_{k=1}^{\infty} \operatorname{Ker}(T - \lambda \mathrm{Id})^k. \tag{5.1}$$

Observación 5.4.1. Si λ es un valor propio de T, entonces el subespacio propio correspondiente es $E_\lambda = \operatorname{Ker}(T - \lambda \mathrm{Id})$. Luego E_λ está contenido en W_λ.

Observación 5.4.2. Si $T \in \mathcal{L}(V)$, entonces

$$\operatorname{Ker}(T) \subset \operatorname{Ker}\big(T^2\big) \subset \operatorname{Ker}\big(T^3\big) \subset \cdots$$

Esto se debe a que si $v \in \operatorname{Ker}\big(T^k\big)$, entonces $T^{k+1}(v) = T\big(T^k(v)\big) = T(0) = 0$, luego $v \in \operatorname{Ker}\big(T^{k+1}\big)$.

Proposición 5.4.3. *Sean $T \in \mathcal{L}(V)$ y $\lambda \in \mathbb{k}$. Entonces*

1. $\operatorname{Ker}(T - \lambda \mathrm{Id}) \subset \operatorname{Ker}(T - \lambda \mathrm{Id})^2 \subset \operatorname{Ker}(T - \lambda \mathrm{Id})^3 \subset \cdots \subset W_\lambda$.
2. *El subespacio propio generalizado W_λ es un subespacio T-invariante.*
3. *El escalar λ es un valor propio de T si y solo si $W_\lambda \neq \{0\}$.*

Dem. La primera afirmación se deduce de la observación anterior aplicada a $T - \lambda \mathrm{Id} \in \mathcal{L}(V)$. Probaremos ahora que W_λ es un subespacio. Es claro que el vector nulo está en W_λ. Sean $u, v \in W_\lambda$ y $a \in \Bbbk$. Sabemos que existen $k, h \geq 1$ tales que $u \in \operatorname{Ker}(T - \lambda \mathrm{Id})^k$ y $v \in \operatorname{Ker}(T - \lambda \mathrm{Id})^h$. Podemos suponer $k \geq h$, luego u, v están en el subespacio $\operatorname{Ker}(T - \lambda \mathrm{Id})^k$ y por lo tanto $au + v \in \operatorname{Ker}(T - \lambda \mathrm{Id})^k \subset W_\lambda$. Esto prueba que W_λ es un subespacio. Para ver que W_λ es T-invariante, alcanza con observar que es $W_\lambda = \bigcup_{k=1}^\infty \operatorname{Ker}(T - \lambda \mathrm{Id})^k$ y recordar que cada $\operatorname{Ker}(T - \lambda \mathrm{Id})^k$ es T-invariante (observación 5.2.5).

Si λ es un valor propio de T, entonces $\{0\} \neq \operatorname{Ker}(T - \lambda \mathrm{Id}) \subset W_\lambda$, luego $W_\lambda \neq \{0\}$. Veamos el recíproco. Sea λ tal que $W_\lambda \neq \{0\}$. Entonces existe $0 \neq v \in V$ y $m \geq 1$ tales que $(T - \lambda \mathrm{Id})^m(v) = 0$. Sea p el menor entero positivo tal que $(T - \lambda \mathrm{Id})^p(v) = 0$. Consideramos $w := (T - \lambda \mathrm{Id})^{p-1}(v)$. Entonces $w \neq 0$ y

$$(T - \lambda \mathrm{Id})(w) = (T - \lambda \mathrm{Id})\big((T - \lambda \mathrm{Id})^{p-1}(v)\big) = (T - \lambda \mathrm{Id})^p(v) = 0.$$

Esto implica que w es un vector propio de T con valor propio λ. $\square$

Proposición 5.4.4. *Si $T(v) = \lambda v$ para ciertos $\lambda \in \Bbbk$ y $v \in V$, entonces $T^k(v) = \lambda^k v$, para todo k.*

Dem. La prueba es por inducción en k. Por hipótesis la fórmula vale cuando $k = 1$. Si asumimos que es $T^k(v) = \lambda^k v$, entonces $T^{k+1}(v) = T\big(T^k(v)\big) = T\big(\lambda^k v\big) = \lambda^k T(v) = \lambda^k \lambda v = \lambda^{k+1} v$. $\square$

Proposición 5.4.5. *Sean $\lambda, \mu \in \Bbbk$ con $\lambda \neq \mu$. Entonces $(T - \mu \mathrm{Id})|_{W_\lambda} \in \mathcal{L}(W_\lambda)$ es un isomorfismo.*

Dem. Primero observar que W_λ es T-invariante y por lo tanto es $(T - \mu \mathrm{Id})$-invariante; luego tiene sentido considerar $(T - \mu \mathrm{Id})|_{W_\lambda} \in \mathcal{L}(W_\lambda)$. Sea $v \in W_\lambda$ tal que $(T - \mu \mathrm{Id})(v) = 0$. Entonces

$$T(v) = \mu v \quad \Rightarrow \quad (T - \lambda \mathrm{Id})(v) = T(v) - \lambda v = \mu v - \lambda v = (\mu - \lambda) v.$$

Luego $(T - \lambda \mathrm{Id})(v) = (\mu - \lambda) v$. Entonces aplicando la proposición anterior a $T - \lambda \mathrm{Id} \in \mathcal{L}(V)$ obtenemos que vale $(T - \lambda \mathrm{Id})^n(v) = (\mu - \lambda)^n v$, para todo n. Como $v \in W_\lambda$, entonces existe $k \geq 1$ tal que $(T - \lambda \mathrm{Id})^k(v) = 0$. Luego $(\mu - \lambda)^k v = 0$. Como es $\lambda \neq \mu$, concluimos $v = 0$. Hemos probado $\operatorname{Ker}(T - \mu \mathrm{Id})|_{W_\lambda} = \{0\}$. Luego $(T - \mu \mathrm{Id})|_{W_\lambda} \in \mathcal{L}(W_\lambda)$ es inyectiva y por lo tanto es un isomorfismo (dado que W_λ tiene dimensión finita). $\square$

Proposición 5.4.6. *Supongamos que el polinomio característico $\chi_T(t)$ se escinde. Sea λ un valor propio de T con multiplicidad algebraica m. Entonces*

1. $\dim W_\lambda \leq m$.
2. $W_\lambda = \operatorname{Ker}(T - \lambda \mathrm{Id})^m$.

Dem. Sea μ un valor propio de T, con $\mu \neq \lambda$. Notar $\operatorname{Ker}(T|_{W_\lambda} - \mu \mathrm{Id}_{W_\lambda}) = \operatorname{Ker}(T - \mu \mathrm{Id})|_{W_\lambda} = \{0\}$ (por la proposición anterior). Luego μ no es un valor propio de $T|_{W_\lambda}$. Como $\chi_T(t)$ se escinde, entonces es $\chi_T(t) = \pm(t - \lambda_1)^{n_1} \cdots (t - \lambda_h)^{n_h}$, siendo $\lambda_i \neq \lambda_j$ si $i \neq j$. Podemos suponer $\lambda = \lambda_1$ y por lo tanto $m = n_1$. Como W_λ es T-invariante, entonces $\chi_{T|_{W_\lambda}}(t)$ divide a $\chi_T(t)$. Pero por lo que vimos recién, $\lambda_2, \ldots, \lambda_k$ no son raíces de $\chi_{T|_{W_\lambda}}(t)$. Luego $\chi_{T|_{W_\lambda}}(t) = \pm(t - \lambda)^r$, con $r \leq m$ y por lo tanto $\dim W_\lambda = \operatorname{grado}\Big(\chi_{T|_{W_\lambda}}(t)\Big) = r \leq m$.

Consideremos la segunda afirmación. Por un lado es $W_\lambda = \bigcup_{k=1}^\infty \operatorname{Ker}(T - \lambda \mathrm{Id})^k$, luego $\operatorname{Ker}(T - \lambda \mathrm{Id})^m \subset W_\lambda$. Por otro lado, el teorema de Cayley-Hamilton aplicado a $T|_{W_\lambda}$ implica $0 = (T|_{W_\lambda} - \lambda \mathrm{Id})^r = (T - \lambda \mathrm{Id})^r|_{W_\lambda}$, y al ser $r \leq m$, es $(T - \lambda \mathrm{Id})^m|_{W_\lambda} = 0$. Luego vale $(T - \lambda \mathrm{Id})^m(v) = 0$ para todo $v \in W_\lambda$, lo cual implica $W_\lambda \subset \operatorname{Ker}(T - \lambda \mathrm{Id})^m$. Así que tenemos la doble inclusión y por lo tanto $W_\lambda = \operatorname{Ker}(T - \lambda \mathrm{Id})^m$. $\square$

Teorema 5.4.7. *Si los valores propios de T son $\lambda_1, \dots, \lambda_h$ y el polinomio característico $\chi_T(t)$ se escinde, entonces $V = \sum_{i=1}^h W_{\lambda_i}$.*

Dem. La prueba es por inducción en h. Si $h = 1$, entonces $\chi_T(t) = (-1)^n(t-\lambda_1)^n$, siendo $n = \dim V$. Aplicando el teorema de Cayley-Hamilton obtenemos $(T - \lambda_1 \mathrm{Id})^n = 0$; luego $V = \operatorname{Ker}(T - \lambda_1\mathrm{Id})^n = W_{\lambda_1}$.

Razonando inductivamente, supongamos que tenemos probado el teorema cuanto el operador tiene $h-1$ valores propios y sea $T \in \mathcal{L}(V)$ como en la hipótesis. Consideremos el valor propio λ_h. Si la multiplicidad algebraica de λ_h es n_h, entonces $W_{\lambda_h} = \operatorname{Ker}(T-\lambda_h\mathrm{Id})^{n_h}$. Sea $U = \operatorname{Im}(T - \lambda_h \mathrm{Id})^{n_h}$. Notar que U es un subespacio T-invariante. Esto implica que $\chi_{T|_U}(t)$ divide a $\chi_T(t)$. Entonces $\chi_{T|_U}(t)$ se escinde y sus raíces son también raíces de $\chi_T(t)$; luego los valores propios de $T|_U$ están contenidos en los valores propios de T. Probaremos que $\lambda_1, \dots, \lambda_{h-1}$ son valores propios de $T|_U$. Sea $i \in \{1, \dots, h-1\}$. La proposición 5.4.5 implica que $(T - \lambda_h\mathrm{Id})|_{W_{\lambda_i}} \in \mathcal{L}(W_{\lambda_i})$ es un isomorfismo, luego $(T-\lambda_h\mathrm{Id})^{n_h}|_{W_{\lambda_i}} \in \mathcal{L}(W_{\lambda_i})$ es un isomorfismo. Entonces

$$W_{\lambda_i} = (T-\lambda_h\mathrm{Id})^{n_h}(W_{\lambda_i}) \subset (T-\lambda_h\mathrm{Id})^{n_h}(V) = U \quad \Rightarrow \quad E_{\lambda_i} \subset W_{\lambda_i} \subset U, \quad \forall i = 1, \dots, h-1.$$

Esto prueba que $\lambda_1, \dots, \lambda_{h-1}$ son valores propios de $T|_U$. Veamos ahora que λ_h no es valor propio de $T|_U$. Sea $v \in U$ tal que $T(v) = \lambda_h v$. Como $v \in U$, entonces existe $w \in V$ tal que $v = (T-\lambda_h\mathrm{Id})^{n_h}(w)$. Luego

$$0 = (T-\lambda_h\mathrm{Id})(v) = (T-\lambda_h\mathrm{Id})^{n_h+1}(w) \quad \Rightarrow \quad w \in W_{\lambda_h} = \operatorname{Ker}(T-\lambda_h\mathrm{Id})^{n_h} \quad \Rightarrow \quad v = (T-\lambda_h\mathrm{Id})^{n_h}(w) = 0.$$

En resumen, $\chi_{T|_U}(t)$ se escinde y sus valores propios son $\lambda_1, \dots, \lambda_{h-1}$. Entonces aplicando la hipótesis de inducción a $T|_U \in \mathcal{L}(U)$ obtenemos $U = \sum_{i=1}^{h-1} W_{\lambda_i}(T|_U)$, siendo $W_{\lambda_i}(T|_U)$ los subespacios propios generalizados correspondientes a $T|_U$. Observar que es $W_{\lambda_i}(T|_U) = W_{\lambda_i} \cap U$, pero recién probamos $W_{\lambda_i} \subset U$, luego $W_{\lambda_i}(T|_U) = W_{\lambda_i}$, para todo $i = 1, \dots, h-1$. Así que vale $U = \sum_{i=1}^{h-1} W_{\lambda_i}$.
Sea $v \in V$. Entonces $(T-\lambda_h\mathrm{Id})^{n_h}(v) \in U$. Luego existen $v_i \in W_{\lambda_i}$ tales que $(T-\lambda_h\mathrm{Id})^{n_h}(v) = \sum_{i=1}^{h-1} v_i$. Como cada $(T-\lambda_h\mathrm{Id})^{n_h}|_{W_{\lambda_i}}$ es un isomorfismo, entonces existen $u_i \in W_{\lambda_i}$ tales que $v_i = (T-\lambda_h\mathrm{Id})^{n_h}(u_i)$ y por lo tanto

$$(T-\lambda_h\mathrm{Id})^{n_h}(v) = \sum_{i=1}^{h-1} v_i = \sum_{i=1}^{h-1}(T-\lambda_h\mathrm{Id})^{n_h}(u_i) \quad \Rightarrow \quad (T-\lambda_h\mathrm{Id})^{n_h}\left(v - \sum_{i=1}^{h-1} u_i\right) = 0.$$

Luego existe $u_h \in \operatorname{Ker}(T-\lambda_h\mathrm{Id})^{n_h} = W_{\lambda_h}$ tal que $v - \sum_{i=1}^{h-1} u_i = u_h$. Entonces $v = \sum_{i=1}^h u_i$, con $u_i \in W_{\lambda_i}$, para todo $i = 1, \dots, h$, lo cual prueba la tesis. $\square$

Proposición 5.4.8. *Si $\lambda_1, \dots, \lambda_h$ son valores propios distintos de T, entonces $W_{\lambda_1}, \dots, W_{\lambda_h}$ son subespacios independientes.*

Dem. La prueba es por inducción en h. Si $h = 1$, entonces el resultado vale trivialmente. Supongamos que es $h \geq 2$ y que el resultado vale cuando hay $h-1$ valores propios. Sean $v_i \in W_{\lambda_i}$ tales que $v_1 + \cdots + v_h = 0$. Como $v_h \in W_{\lambda_h}$, entonces existe $k \geq 1$ tal que $(T-\lambda_h\mathrm{Id})^k(v_h) = 0$. Luego

$$-v_h = \sum_{i=1}^{h-1} v_i \quad \Rightarrow \quad 0 = (T-\lambda_h\mathrm{Id})^k(-v_h) = \sum_{i=1}^{h-1}(T-\lambda_h\mathrm{Id})^k(v_i).$$

Como cada W_{λ_i} es T-invariante, deducimos $(T-\lambda_h\mathrm{Id})^k(v_i) \in W_{\lambda_i}$, para todo $i = 1, \dots, h-1$. Luego por la hipótesis de inducción es $(T-\lambda_h\mathrm{Id})^k(v_i) = 0$, para todo $i = 1, \dots, h-1$. Como es $\lambda_h \neq \lambda_i$, sabemos que $(T-\lambda_h\mathrm{Id})^k|_{W_{\lambda_i}}$ es un isomorfismo y por lo tanto es $v_i = 0$, para todo $i = 1, \dots, h-1$. Esto a su vez implica $v_h = 0$, lo cual concluye la prueba. $\square$

Corolario 5.4.9. *Supongamos* $\chi_T(t) = (-1)^n (t-\lambda_1)^{n_1} \cdots (t-\lambda_h)^{n_h}$*, con* $\lambda_i \neq \lambda_j$ *si* $i \neq j$*. Entonces*

$$V = \bigoplus_{i=1}^{h} \operatorname{Ker}(T - \lambda_i \mathrm{Id})^{n_i}. \tag{5.2}$$

Además, $\dim \operatorname{Ker}(T-\lambda_i \mathrm{Id})^{n_i} = n_i$*, para todo* $i = 1, \ldots, h$*.*

Dem. La fórmula (5.2) se obtiene aplicando del teorema y proposición anteriores, y la proposición 5.4.6. Además en la proposición 5.4.6 vimos que vale $\dim \operatorname{Ker}(T-\lambda_i \mathrm{Id})^{n_i} \leq n_i$, para todo $i = 1, \ldots, h$. Luego usando (5.2) y que vale $\sum_{i=1}^{h} n_i = \dim V$, obtenemos

$$\dim V = \sum_{i=1}^{h} \dim \operatorname{Ker}(T-\lambda_i \mathrm{Id})^{n_i} \leq \sum_{i=1}^{h} n_i = \dim V.$$

Entonces necesariamente es $\dim \operatorname{Ker}(T-\lambda_i \mathrm{Id})^{n_i} = n_i$, para todo $i = 1, \ldots, h$. □

Observación 5.4.10. En la fórmula (5.2) los subespacios son T-invariantes, luego se puede aplicar la proposición 5.1.4 para obtener una base $\mathcal{B}$ de V tal que $[T]_{\mathcal{B}}$ es una matriz diagonal en bloques.

Capítulo 6

Forma de Jordan

En el primer capítulo vimos que un operador es diagonalizable si y solo si su polinomio característico se escinde y para cada uno de sus valores propios la multiplicidad geométrica coincide con la algebraica. En este capítulo estudiaremos los operadores cuyo polinomio característico se escinde, pero que no necesariamente verifican la segunda condición. Esto cubre todos los operadores en espacios complejos. Trabajaremos siempre en un espacio de dimensión finita $V \neq \{0\}$ sobre un cuerpo arbitrario $\mathbb{k}$. Las bases serán siempre bases ordenadas. Si $A \in M_n(\mathbb{k})$, entonces diremos que n es el *orden* de la matriz A.

6.1. Forma de Jordan

Recordemos que un operador $T \in \mathcal{L}(V)$ es diagonalizable si y sólo si existe una base $\mathcal{B}$ de V formada por vectores propios de T. En ese caso, si $\mathcal{B} = \{v_1, \ldots, v_n\}$ y $T(v_i) = \lambda_i v_i$, para todo $i = 1, \ldots, n$, entonces

$$[T]_{\mathcal{B}} = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_n \end{pmatrix}.$$

También vimos que no todo operador es diagonalizable, aún si el polinomio característico se escinde. En este capítulo probaremos que vale el siguiente resultado.

Teorema 6.1.1 (Jordan)**.** *Si $T \in \mathcal{L}(V)$ es tal que $\chi_T(t)$ se escinde en $\mathbb{k}$, entonces existe una base $\mathcal{B}$ de V tal que*

$$[T]_{\mathcal{B}} = \begin{pmatrix} J_1 & & & \\ & J_2 & & \\ & & \ddots & \\ & & & J_h \end{pmatrix}, \quad J_i = \begin{pmatrix} \lambda_i & 1 & & & \\ & \lambda_i & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda_i & 1 \\ & & & & \lambda_i \end{pmatrix} \in M_{p_i}(\mathbb{k}), \ i = 1, \ldots, h. \tag{6.1}$$

para ciertos $\lambda_1, \ldots, \lambda_h \in \mathbb{k}$, $p_1, \ldots, p_h \geq 1$ y $h \geq 1$. □

La prueba de se verá más adelante (corolario 6.1.16). Respecto a las fórmulas en (6.1), notar primero que $[T]_{\mathcal{B}}$ es triangular superior, luego los escalares $\lambda_1, \ldots, \lambda_h$ son los valores propios de T. Cada matriz J_i es un *bloque de Jordan* correspondiente al valor propio λ_i (pueden haber varios bloques correspondientes al mismo valor propio). Cuando ordenamos los bloques que forman $[T]_{\mathcal{B}}$ de forma tal que ponemos juntos los bloques correspondientes a los mismos valores propios y estos los ordenamos en órdenes decrecientes, entonces a $[T]_{\mathcal{B}}$ le llamamos la *forma de Jordan*[1] de T y decimos que $\mathcal{B}$ es una *base de Jordan* de V correspondiente a T.

[1]Se puede probar que la forma de Jordan es única a menos de reordenar los valores propios. Esto justifica el referirnos a $[T]_{\mathcal{B}}$ como **la** forma de Jordan de T.

Observación 6.1.2. Como en $\mathbb{C}$ todo polinomio se escinde, del teorema anterior se deduce que todo operador en un espacio vectorial complejo admite una base de Jordan.

Observación 6.1.3. Si $T \in \mathcal{L}(V)$ es tal que existe una base de Jordan correspondiente a T, entonces $\chi_T(t)$ se escinde en $\Bbbk$. Esto se debe a que si existe una base $\mathcal{B}$ de V tal que $[T]_\mathcal{B}$ es como en (6.1), entonces

$$\chi_T(t) = \prod_{i=1}^{h} \det(J_i - t\mathrm{Id}) = \prod_{i=1}^{h} (\lambda_i - t)^{p_i} = (-1)^n \prod_{i=1}^{h} (t - \lambda_i)^{p_i}\,, \text{ siendo } n = \dim V.$$

Luego para que exista una base de Jordan correspondiente a T, es necesario que el polinomio característico de T se escinda.

Ejemplo 6.1.4. Una forma de Jordan de un operador, es una matriz del tipo

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

en el cual estamos omitiendo los ceros fuera de los bloques de Jordan de A. En este caso hay dos bloques de Jordan correspondientes al valor propio 2, uno correspondiente a 3 y otro correspondiente a 0.

Observación 6.1.5. Sea $T \in \mathcal{L}(V)$ y supongamos que tenemos una base de Jordan $\mathcal{B}$ como en el teorema 6.1.1. Mantenemos las notaciones del teorema. Separando la base $\mathcal{B}$ en partes correspondientes a cada bloque J_i obtenemos una descomposición $\mathcal{B} = \mathcal{C}_1 \sqcup \cdots \sqcup \mathcal{C}_h$ en que cada subconjunto $\mathcal{C}_i$ tiene p_i elementos. Para simplificar las notaciones, omitiendo el subíndice i podemos escribir cada subconjunto $\mathcal{C}_i$ de la forma $\mathcal{C} = \{v_1, v_2, \ldots, v_p\}$, y si escribimos el bloque de Jordan correspondiente de la forma

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix},$$

entonces vale

$$T(v_1) = \lambda v_1, \quad T(v_2) = v_1 + \lambda v_2, \quad \cdots \quad , T(v_{p-1}) = v_{p-2} + \lambda v_{p-1}, \quad T(v_p) = v_{p-1} + \lambda v_p. \tag{6.2}$$

Estas fórmulas se pueden reescribir de la forma siguiente

$$(T - \lambda\mathrm{Id})(v_1) = 0, \quad v_1 = (T - \lambda\mathrm{Id})(v_2), \quad \cdots, \quad v_{p-2} = (T - \lambda\mathrm{Id})(v_{p-1}), \quad v_{p-1} = (T - \lambda\mathrm{Id})(v_p).$$

La última fórmula permite obtener v_{p-1} en función de v_p, sustituyendo ese valor en la penúltima fórmula obtenemos v_{p-2} en función de v_p, y siguiendo así obtenemos $v_1, \ldots, v_{p-1}$ en función de v_p:

$$v_1 = (T - \lambda\mathrm{Id})^{p-1}(v_p), \quad \cdots, \quad v_{p-2} = (T - \lambda\mathrm{Id})^2(v_p), \quad v_{p-1} = (T - \lambda\mathrm{Id})(v_p). \tag{6.3}$$

Notar que a $(T - \lambda\mathrm{Id})(v_1) = 0$ le corresponde $(T - \lambda\mathrm{Id})^p(v_p) = 0$. Luego escribiendo $v = v_p$, obtenemos

$$\mathcal{C} = \left\{(T - \lambda\mathrm{Id})^{p-1}(v),\, (T - \lambda\mathrm{Id})^{p-2}(v),\, \ldots, (T - \lambda\mathrm{Id})^2(v),\, (T - \lambda\mathrm{Id})(v),\, v\right\}, \text{ con } (T - \lambda\mathrm{Id})^p(v) = 0.$$

Recíprocamente, si $\mathcal{C}$ es como arriba, entonces escribiendo $v_p = v$ y definiendo $v_1, v_2, \ldots, v_{p-1}$ mediante (6.3) obtenemos vectores $v_1, v_2, \ldots, v_p$ que verifican (6.2).

Definición 6.1.6. Sea $T \in \mathcal{L}(V)$ y $\lambda \in \Bbbk$. Un *ciclo* de T correspondiente a λ es un conjunto ordenado de la forma

$$\mathcal{C} = \left\{(T-\lambda\mathrm{Id})^{p-1}(v),\, (T-\lambda\mathrm{Id})^{p-2}(v),\, \ldots,(T-\lambda\mathrm{Id})^{2}(v),\, (T-\lambda\mathrm{Id})(v),\, v\right\} \tag{6.4}$$

siendo $v \in \mathrm{Ker}(T-\lambda\mathrm{Id})^p \setminus \mathrm{Ker}(T-\lambda\mathrm{Id})^{p-1}$ para cierto $p \geq 1$. La *longitud* de $\mathcal{C}$ es p y su *vector inicial* es $(T-\lambda\mathrm{Id})^{p-1}(v)$.

Por ejemplo, un ciclo de longitud 1 correspondiente a λ, es un conjunto de la forma $\mathcal{C} = \{v\}$, en el cual v es un vector propio de T con valor propio λ. Un ciclo de longitud 2 es un conjunto de la forma $\mathcal{C} = \{(T-\lambda\mathrm{Id})(v),\, v\}$ en el cual $(T-\lambda\mathrm{Id})^2(v) = 0$ y $(T-\lambda\mathrm{Id})(v) \neq 0$.

Observación 6.1.7. Una base de Jordan es una base obtenida como unión disjunta de ciclos de T. Para poder hallarla tenemos que saber cómo encontrar esos ciclos. Los siguientes resultados van en esa dirección.

En lo que sigue $T \in \mathcal{L}(V)$ es un operador arbitrario fijo.

Proposición 6.1.8. *Sea $\lambda \in \Bbbk$ y $\mathcal{C}$ un ciclo correspondiente a λ. Entonces*

1. *El escalar λ es un valor propio de T y el vector inicial de $\mathcal{C}$ es un vector propio correspondiente a λ.*
2. *Si $\mathcal{C}$ tiene longitud p, entonces $\mathcal{C}$ está contenido en $\mathrm{Ker}(T-\lambda\mathrm{Id})^p$ y por lo tanto está contenido en el subespacio propio generalizado W_λ.*

Dem. Escribimos $\mathcal{C}$ como en (6.4). La condición $v \in \mathrm{Ker}(T-\lambda\mathrm{Id})^p \setminus \mathrm{Ker}(T-\lambda\mathrm{Id})^{p-1}$, implica que si escribimos $v_1 = (T-\lambda\mathrm{Id})^{p-1}(v)$, entonces $v_1 \neq 0$ y $(T-\lambda\mathrm{Id})(v_1) = (T-\lambda\mathrm{Id})^p(v) = 0$. Luego $T(v_1) = \lambda v_1$, con $v_1 \neq 0$. Además, para cada $k \geq 0$ es

$$(T-\lambda\mathrm{Id})^p\big((T-\lambda\mathrm{Id})^k(v)\big) = (T-\lambda\mathrm{Id})^k\big((T-\lambda\mathrm{Id})^p(v)\big) = (T-\lambda\mathrm{Id})^k(0) = 0 \Rightarrow (T-\lambda\mathrm{Id})^k(v) \in \mathrm{Ker}(T-\lambda\mathrm{Id})^p.$$

Luego $\mathcal{C} \subset \mathrm{Ker}(T-\lambda\mathrm{Id})^p \subset W_\lambda$. $\square$

Proposición 6.1.9. *Si $\mathcal{C}_1$ y $\mathcal{C}_2$ son ciclos de T correspondientes a un mismo valor propio λ tales que sus vectores iniciales son distintos, entonces $\mathcal{C}_1$ y $\mathcal{C}_2$ son disjuntos.*

Dem. Supongamos

$$\mathcal{C}_1 = \left\{(T-\lambda\mathrm{Id})^{p-1}(v),\, \ldots,(T-\lambda\mathrm{Id})(v),\, v\right\}, \quad \mathcal{C}_2 = \left\{(T-\lambda\mathrm{Id})^{q-1}(w),\, \ldots,(T-\lambda\mathrm{Id})(w),\, w\right\}$$

tales que $v \in \mathrm{Ker}(T-\lambda\mathrm{Id})^p \setminus \mathrm{Ker}(T-\lambda\mathrm{Id})^{p-1}$ y $w \in \mathrm{Ker}(T-\lambda\mathrm{Id})^q \setminus \mathrm{Ker}(T-\lambda\mathrm{Id})^{q-1}$. Si $\mathcal{C}_1 \cap \mathcal{C}_2 \neq \emptyset$, entonces existen $k \leq p-1$ y $h \leq q-1$ tales que $(T-\lambda\mathrm{Id})^k(v) = (T-\lambda\mathrm{Id})^h(w)$. Aplicando $(T-\lambda\mathrm{Id})^{p-1-k}$ a ambos lados obtenemos $(T-\lambda\mathrm{Id})^{p-1}(v) = (T-\lambda\mathrm{Id})^r(w)$, siendo $r = h+p-1-k$. Si es $r \geq q$, entonces obtendríamos $(T-\lambda\mathrm{Id})^{p-1}(v) = 0$, lo cual contradice $v \notin \mathrm{Ker}(T-\lambda\mathrm{Id})^{p-1}$. Luego $r \leq q-1$. Si es $r \leq q-2$, entonces obtendríamos $0 = (T-\lambda\mathrm{Id})^p(v) = (T-\lambda\mathrm{Id})^{r+1}(w)$, con $r+1 \leq q-1$, lo cual contradice $w \notin \mathrm{Ker}(T-\lambda\mathrm{Id})^{p-1}$. Entonces necesariamente es $r = q-1$, lo cual nos lleva a $(T-\lambda\mathrm{Id})^{p-1}(v) = (T-\lambda\mathrm{Id})^{q-1}(w)$, lo cual tampoco es posible porque por hipótesis estos vectores son distintos. $\square$

Observación 6.1.10. Si

$$\mathcal{C} = \left\{(T-\lambda\mathrm{Id})^{p-1}(v),\, \ldots,(T-\lambda\mathrm{Id})^{2}(v),\, (T-\lambda\mathrm{Id})(v),\, v\right\}$$

es un ciclo de longitud p, entonces

$$\hat{\mathcal{C}} = \left\{(T-\lambda\mathrm{Id})^{p-1}(v),\, \ldots,(T-\lambda\mathrm{Id})^{2}(v),\, (T-\lambda\mathrm{Id})(v)\right\}.$$

es un ciclo de longitud $p-1$. Esto se debe a que si definimos $w = (T-\lambda\mathrm{Id})(v)$, entonces

$$\hat{\mathcal{C}} = \left\{(T-\lambda\mathrm{Id})^{p-2}(w),\, \ldots,(T-\lambda\mathrm{Id})(v),\, w\right\},$$

y $v \in \mathrm{Ker}(T-\lambda\mathrm{Id})^p \setminus \mathrm{Ker}(T-\lambda\mathrm{Id})^{p-1}$ implica $w \in \mathrm{Ker}(T-\lambda\mathrm{Id})^{p-1} \setminus \mathrm{Ker}(T-\lambda\mathrm{Id})^{p-2}$.

Teorema 6.1.11. *Sean $\mathcal{C}_1, \cdots, \mathcal{C}_h$ ciclos de T correspondientes a un mismo valor propio λ. Si sus vectores iniciales son linealmente independientes, entonces $\mathcal{C}_1, \cdots, \mathcal{C}_h$ son disjuntos*[2] *y $\mathcal{C}_1 \cup \cdots \cup \mathcal{C}_h$ es LI.*

Dem. Notar que el ser disjuntos es consecuencia de la proposición anterior. La prueba de que la unión es LI será hecha por inducción en p, siendo p el máximo de las longitudes de $\mathcal{C}_1, \cdots, \mathcal{C}_h$ y $h \geq 1$ arbitrario.

Si $p = 1$, entonces es $\mathcal{C}_1 = \{v_1\}, \ldots, \mathcal{C}_h = \{v_h\}$, siendo $v_1, \ldots, v_h$ vectores propios correspondientes a λ tal que $\{v_1, \ldots, v_h\}$ LI. Entonces obviamente $\mathcal{C}_1 \cup \cdots \cup \mathcal{C}_h = \{v_1, \ldots, v_h\}$ es LI.

Supongamos que tenemos probada la proposición cuando hay cierta cantidad de ciclos tales que el máximo de sus longitudes es $p-1$ y que tenemos ahora ciclos $\mathcal{C}_1, \cdots, \mathcal{C}_h$ tales que el máximo de sus longitudes es p. Supongamos que la longitud de cada ciclo es m_i y escribimos

$$\mathcal{C}_i = \left\{(T - \lambda \mathrm{Id})^{m_i - 1}(v_i),\, \ldots, (T - \lambda \mathrm{Id})(v_i),\, v_i\right\}, \quad i = 1, \ldots, h$$

Nuestra hipótesis es que $\left\{(T - \lambda \mathrm{Id})^{m_1-1}(v_1), \ldots, (T - \lambda \mathrm{Id})^{m_h-1}(v_h)\right\}$ es LI. Consideremos una combinación lineal de todos los vectores de $\mathcal{C}_1 \cup \cdots \cup \mathcal{C}_h$ igualada al vector nulo

$$\sum_{l_1=0}^{m_1-1} a_{1,l_1}(T - \lambda \mathrm{Id})^{l_1}(v_1) + \cdots + \sum_{l_h=0}^{m_h-1} a_{h,l_h}(T - \lambda \mathrm{Id})^{l_h}(v_h) = 0. \tag{6.5}$$

Si aplicamos $T - \lambda \mathrm{Id}$ a una de las sumatorias anteriores, obtenemos

$$(T - \lambda \mathrm{Id})\left(\sum_{l_i=0}^{m_i-1} a_{i,l_i}(T - \lambda \mathrm{Id})^{l_i}(v_i)\right) = \sum_{l_i=0}^{m_i-1} a_{i,l_i}(T - \lambda \mathrm{Id})^{l_i+1}(v_i) = \sum_{l_i=0}^{m_i-2} a_{i,l_i}(T - \lambda \mathrm{Id})^{l_i+1}(v_i).$$

Observar que la segunda suma llega solo hasta $m_i - 2$, dado que $(T - \lambda \mathrm{Id})^{m_i}(v_i) = 0$. Luego aplicando $T - \lambda \mathrm{Id}$ a (6.5) obtenemos

$$\sum_{l_1=0}^{m_1-2} a_{1,l_1}(T - \lambda \mathrm{Id})^{l_1+1}(v_1) + \cdots + \sum_{l_h=0}^{m_h-2} a_{h,l_h}(T - \lambda \mathrm{Id})^{l_h+1}(v_h) = 0. \tag{6.6}$$

Para cada $i = 1, \ldots, h$, consideremos

$$\hat{\mathcal{C}}_i = \left\{(T - \lambda \mathrm{Id})^{m_i-1}(v_i),\, \ldots, (T - \lambda \mathrm{Id})(v_i)\right\},$$

que es el conjunto de los vectores involucrados en la sumatoria correspondiente dentro de (6.6). Cada conjunto $\hat{\mathcal{C}}_i$ es un ciclo[3] de longitud $m_i - 1$ y su vector inicial es el mismo que el de $\mathcal{C}_i$, luego $\hat{\mathcal{C}}_1, \cdots, \hat{\mathcal{C}}_h$ verifican las hipótesis de inducción y por lo tanto $\hat{\mathcal{C}}_1 \cup \cdots \cup \hat{\mathcal{C}}_h$ es LI. Esto implica que todos los coeficientes en (6.6) son nulos, es decir

$$a_{i,0} = a_{i,1} = \cdots = a_{i,m_i-2} = 0, \quad \forall i = 1, \ldots, h.$$

Sustituyendo estos valores en (6.5) obtenemos

$$a_{1,m_1-1}(T - \lambda \mathrm{Id})^{m_1-1}(v_1) + \cdots + a_{h,m_h-1}(T - \lambda \mathrm{Id})^{m_h-1}(v_h) = 0.$$

Finalmente, como por hipótesis el conjunto $\left\{(T - \lambda \mathrm{Id})^{m_1-1}(v_1), \ldots, (T - \lambda \mathrm{Id})^{m_h-1}(v_h)\right\}$ es LI, deducimos $a_{i,m_i-1} = 0$, para todo $i = 1, \ldots, h$. Esto termina la prueba de que $\mathcal{C}_1 \cup \cdots \cup \mathcal{C}_h$ es LI. □

[2] Por disjuntos nos referimos a que son disjuntos dos a dos, es decir, $\mathcal{C}_i \cap \mathcal{C}_j = \emptyset$, si $i \neq j$.

[3] Acá hay cierta sutileza, ver la primera parte de la observación siguiente.

Observaciones 6.1.12.

1. Cuando en la prueba anterior estamos haciendo el razonamiento inductivo, podría suceder que algunos de los ciclos $\mathcal{C}_1, \cdots, \mathcal{C}_h$ tengan longitud 1. Si $\mathcal{C}_i = \{v_i\}$ es un ciclo de longitud $m_i = 1$, entonces $(T - \lambda \mathrm{Id})(v_i) = 0$ y por lo tanto no aparece la sumatoria correspondiente en (6.6). Luego puede pasar que algunos de los ciclos en $\hat{\mathcal{C}}_1, \cdots, \hat{\mathcal{C}}_h$ no aparezcan. Eso no afecta la prueba, pero es es la razón por la cual en la inducción estamos considerando que la cantidad h de ciclos es arbitraria.

2. El razonamiento de la prueba anterior es válido aun si tenemos un solo ciclo $\mathcal{C}$, dado que el vector inicial v de $\mathcal{C}$ es un vector propio y por lo tanto $\{v\}$ es LI. Luego los ciclos son conjuntos LI.

Definición 6.1.13. Un operador $T \in \mathcal{L}(V)$ se dice *nilpotente* si existe un entero positivo k tal que $T^k = 0$. El *orden* de nilpotencia de T es el menor k que verifica $T^k = 0$. Luego T es nilpotente de orden p si y solo si $T^p = 0$ y $T^{p-1} \neq 0$.

Observaciones 6.1.14. Sea $T \in \mathcal{L}(V)$ un operador nilpotente de orden p.

1. Sabemos $T^{p-1} \neq 0$, luego existe $v \in V$ tal que $T^{p-1}(v) \neq 0$. Entonces $0 = T^p(v) = T\big(T^{p-1}(v)\big)$ y por lo tanto $0 \neq T^{p-1}(v) \in \operatorname{Ker} T$. Luego $\operatorname{Ker} T \neq \{0\}$.

2. Si λ es un valor propio de T, entonces existe $v \neq 0$ tal que $T(v) = \lambda v$. Esto implica $T^p(v) = \lambda^p v$, pero como es $T^p = 0$, obtenemos $\lambda = 0$. Luego el único valor propio de T es $\lambda = 0$.

3. Un ciclo de T (correspondiente a 0) de longitud p es un conjunto de la forma $\{T^{p-1}(v), \cdots, T(v), v\}$, siendo $v \in V$ tal que $T^p(v) = 0$ y $T^{p-1}(v) \neq 0$.

Teorema 6.1.15. *Si $T \in \mathcal{L}(V)$ es un operador nilpotente, entonces existe una base de Jordan de V correspondiente a T.*

Dem. La prueba es por inducción en $n = \dim V$. El caso $n = 1$ es trivial (necesariamente es $T = 0$ y si elegimos un vector $v \in V$ tal que $v \neq 0$, entonces $\{v\}$ es un ciclo y es una base de V).
Razonando inductivamente, supongamos que vale la tesis cuando la dimensión del espacio es menor que n y sea $T \in \mathcal{L}(V)$ un operador nilpotente con $\dim V = n > 1$.
Si $T = 0$ (el operador nulo), entonces toda base de V es una base de Jordan correspondiente a T.
Supongamos ahora $T \neq 0$ y consideremos $W = \operatorname{Im}(T) \neq \{0\}$. Como T es nilpotente, entonces $\operatorname{Ker} T \neq \{0\}$. Luego usando $\dim \operatorname{Ker}(T) + \dim \operatorname{Im}(T) = \dim V$, obtenemos $\dim W < \dim V$. Claramente W es T-invariante, y como T es nilpotente, concluimos que $T|_W \in \mathcal{L}(W)$ es nilpotente. Entonces por la hipótesis de inducción aplicada a $T|_W \in \mathcal{L}(W)$, sabemos que existe $\mathcal{B}_0$ base de W de la forma $\mathcal{B}_0 = \mathcal{C}_1 \cup \cdots \cup \mathcal{C}_k$, siendo $\mathcal{C}_1, \ldots, \mathcal{C}_k$ ciclos de T disjuntos. Sea $\mathcal{C}_i$ uno de estos ciclos. Supongamos que la longitud de $\mathcal{C}_i$ es p_i y que escribimos

$$\mathcal{C}_i = \left\{T^{p_i-1}(v_i),\, \ldots, T(v_i),\, v_i\right\}.$$

Como es $v_i \in \mathcal{C}_i \subset W = \operatorname{Im}(T)$, entonces existe $u_i \in V$ tal que $v_i = T(u_i)$. Luego si consideramos

$$\hat{\mathcal{C}}_i := \mathcal{C}_i \cup \{u_i\} = \left\{T^{p_i-1}(v_i),\, \ldots, T(v_i),\, v_i,\, u_i\right\} = \left\{T^{p_i}(u_i),\, \ldots, T^2(u_i),\, T(u_i),\, u_i\right\},$$

entonces $\hat{\mathcal{C}}_i$ es un ciclo de longitud $p_i + 1$ cuyo vector inicial es el mismo que el de $\mathcal{C}_i$.
Luego hemos probado que existen vectores $u_1, \ldots, u_k \in V$ tales que si definimos $\hat{\mathcal{C}}_i := \mathcal{C}_i \cup \{u_i\}$, entonces $\hat{\mathcal{C}}_1, \ldots, \hat{\mathcal{C}}_k$ son ciclos de T cuyos vectores iniciales forman un conjunto LI (por ser parte de $\mathcal{B}_0$). Esto implica que $\hat{\mathcal{C}}_1, \ldots, \hat{\mathcal{C}}_k$ son ciclos disjuntos y que $\hat{\mathcal{C}}_1 \cup \cdots \cup \hat{\mathcal{C}}_k$ es un conjunto LI.

Consideremos $\mathcal{A} = \left\{T^{p_1-1}(v_1), \ldots, T^{p_k-1}(v_k)\right\}$ el conjunto formado por los vectores iniciales de $\mathcal{C}_1, \ldots, \mathcal{C}_k$. Como los vectores iniciales de los ciclos son vectores propios (correspondientes al valor propio 0), entonces

$\mathcal{A} \subset \operatorname{Ker}(T)$. Además como $\mathcal{A}$ está contenido en $\mathcal{B}_0$ y $\mathcal{B}_0$ es LI, obtenemos que $\mathcal{A}$ es LI. Luego existen $w_1, \ldots, w_h \in \operatorname{Ker}(T)$ tales que

$$\mathcal{B}_1 = \left\{T^{p_1-1}(v_1), \ldots, T^{p_k-1}(v_k), w_1, \ldots, w_h\right\}$$

es una base de $\operatorname{Ker}(T)$. Notar que cada conjunto unitario $\{w_i\}$ es un ciclo de longitud 1. Consideremos

$$\mathcal{B} := \hat{\mathcal{C}}_1 \cup \cdots \cup \hat{\mathcal{C}}_k \cup \{w_1, \ldots, w_h\}.$$

Como los vectores iniciales de los ciclos $\hat{\mathcal{C}}_1, \ldots, \hat{\mathcal{C}}_k, \{w_1\}, \ldots, \{w_h\}$ son LI, entonces estos ciclos son disjuntos y $\mathcal{B}$ es LI. Vamosa probar que $\mathcal{B}$ es una base de V.
Como los ciclos que forman $\mathcal{B}$ son disjuntos, entonces la cantidad de elementos de $\mathcal{B}$ es

$$\#\mathcal{B} = h + \sum_{i=1}^{k} \#\hat{\mathcal{C}}_i = h + \sum_{i=1}^{k}(p_i + 1) = h + k + \sum_{i=1}^{k} p_i.$$

Por otro lado, como $\mathcal{B}_1$ es una base de $\operatorname{Ker} T$, entonces $\dim \operatorname{Ker} T = k + h$, y como $\mathcal{B}_0$ es una base de $\operatorname{Im} T$, entonces $\dim \operatorname{Im} T = \sum_{i=1}^{k} p_i$. Luego $\dim V = \dim \operatorname{Ker}(T) + \dim \operatorname{Im}(T) = h + k + \sum_{i=1}^{k} p_i$.
Entonces $\mathcal{B} \subset V$ es un conjunto LI con la misma cantidad de vectores que la dimensión de V y por lo tanto es base de V. Luego $\mathcal{B}$ es una base de V formada por una unión disjunta de ciclos de T. □

Corolario 6.1.16. *Si el polinomio característico de $T \in \mathcal{L}(V)$ se escinde, entonces existe una base de Jordan de V correspondiente a T.*

Dem. Supongamos $\chi_T(t) = (-1)^n (t - \lambda_1)^{n_1} \cdots (t - \lambda_h)^{n_h}$, con $\lambda_i \neq \lambda_j$, si $i \neq j$. Entonces el corolario 5.4.9 implica

$$V = \bigoplus_{i=1}^{h} \operatorname{Ker}(T - \lambda_i \mathrm{Id})^{n_i}. \tag{6.7}$$

Para cada i, consideremos $W_{\lambda_i} = \operatorname{Ker}(T - \lambda_i \mathrm{Id})^{n_i}$ (el subespacio propio generalizado) y $S_i = (T - \lambda_i \mathrm{Id})|_{W_{\lambda_i}}$. Entonces $S_i^{n_i} = (T - \lambda_i \mathrm{Id})^{n_i}|_{W_{\lambda_i}} = 0$ y por lo tanto $S_i \in \mathcal{L}(W_{\lambda_i})$ es nilpotente. Luego el teorema anterior nos dice que existe una base $\mathcal{B}_i$ de W_{λ_i} formada por ciclos de S_i. Un ciclo de S_i es un conjunto de la forma

$$\mathcal{C} = \left\{S_i^{p-1}(v), \ldots, S_i(v), v\right\}$$

con $v \in \operatorname{Ker}(S_i^p) \setminus \operatorname{Ker}\left(S_i^{p-1}\right)$. Como es $S_i = (T - \lambda_i \mathrm{Id})|_{W_i}$, entonces

$$\mathcal{C} = \left\{(T - \lambda_i \mathrm{Id})^{p-1}(v), \ldots, (T - \lambda_i \mathrm{Id})(v), v\right\}$$

con $v \in \operatorname{Ker}(T - \lambda_i \mathrm{Id})^p \setminus \operatorname{Ker}(T - \lambda_i \mathrm{Id})^{p-1}$. Luego los ciclos de S_i (correspondientes al valor propio 0) son ciclos de T correspondientes al valor propio λ_i. Entonces $\mathcal{B}_i$ es una base de W_i formada por ciclos de T correspondientes al valor propio λ_i. Como la suma (6.7) es directa, entonces $\mathcal{B} := \mathcal{B}_1 \cup \cdots \cup \mathcal{B}_k$ es una base de V formada por ciclos de T. □

Semejanza de matrices y forma de Jordan. Sin referirnos a transformaciones lineales, un *bloque de Jordan* es una matriz de la forma

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix},$$

siendo $\lambda \in \Bbbk$ un escalar arbitrario, y una *matriz de Jordan* es una matriz diagonal en bloques de la forma $A = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{pmatrix}$, siendo $J_1, \dots, J_k$ bloques de Jordan arbitrarios.

Definición 6.1.17. Si $A \in M_n(\Bbbk)$ es tal que su polinomio característico $\chi_A(t) \in \Bbbk[t]$ se escinde, entonces definimos la *forma de Jordan* de A como la de $L_A \in \mathcal{L}(\Bbbk^n)$.

Observar que si J es la forma de Jordan de A y $\mathcal{B} = \{v_1, \dots, v_n\}$ es la base de Jordan para L_A correspondiente, entonces es

$$A = QJQ^{-1}, \quad Q = [v_1|\cdots|v_n].$$

Del teorema de Jordan se deduce el siguiente.

Corolario 6.1.18. *Si $A \in M_n(\Bbbk)$ es tal que $\chi_A(t) \in \Bbbk[t]$ se escinde, entonces A es semejante a una matriz de Jordan.* □

Observación 6.1.19. Sean $A, B \in M_n(\mathbb{C})$. El corolario anterior implica que si A y B tienen la misma forma de Jordan, entonces son semejantes. Como comentamos al principio del capítulo, la forma de Jordan de un operador es única (a menos del orden de los bloques). Además, sabemos que si A y B son semejantes entonces representan a un mismo operador (en bases distintas) y por lo tanto tienen la misma forma de Jordan. Luego $A, B \in M_n(\mathbb{C})$ son semejantes si y solo si tienen la misma forma de Jordan[4].

6.2. Cálculo de la forma de Jordan

En esta sección veremos algunos resultados que en dimensiones bajas permiten determinar la forma de Jordan. En particular esto facilita el hallar la base de Jordan.

Sea $T \in \mathcal{L}(V)$ tal que $\chi_T(t) = (-1)^n (t-\lambda_1)^{n_1} \cdots (t-\lambda_h)^{n_h}$, con $\lambda_i \neq \lambda_j$, si $i \neq j$. Entonces

$$V = \bigoplus_{i=1}^{h} W_{\lambda_i}, \quad W_{\lambda_i} = \operatorname{Ker}(T - \lambda_i \mathrm{Id})^{n_i}.$$

Recordar $\dim W_{\lambda_i} = n_i$, para todo i (corolario 5.4.9). Luego si $\mathcal{B}_i$ es una base de W_{λ_i}, entonces $\mathcal{B} = \bigcup_{i=1}^{h} \mathcal{B}_i$ es una base de V tal que

$$[T]_{\mathcal{B}} = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_h \end{pmatrix}, \quad A_i = \left[T|_{W_{\lambda_i}}\right]_{\mathcal{B}_i} \in M_{n_i}(\Bbbk),\ \forall i = 1, \dots, h.$$

A su vez para cada valor propio λ_i podemos elegir $\mathcal{B}_i$ de la forma $\mathcal{B}_i = \bigcup_{j=1}^{k} \mathcal{C}_j$, siendo $\mathcal{C}_1, \dots, \mathcal{C}_k$ ciclos disjuntos correspondientes a λ_i. Calculando $A_i = \left[T|_{W_{\lambda_i}}\right]_{\mathcal{B}_i}$ en esta base obtenemos

$$A_i = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{pmatrix},$$

[4] Sucede lo mismo en el caso real, usando la forma de Jordan *real* (que se deduce de la forma de Jordan compleja), pero nosotros no vamos a entrar en ese tema.

siendo $J_1,\dots,J_k$ los bloques de Jordan correspondientes al valor propio λ_i. Notar que a cada ciclo $\mathcal{C}_l$ de longitud p le corresponde un bloque de Jordan J_l de orden p.

Vamos a probar que para cada λ_i, la cantidad k de bloques de Jordan correspondientes a λ_i coincide con la multiplicidad geométrica de λ_i. Para eso necesitamos el siguiente resultado.

Lema 6.2.1. *Sea $T \in \mathcal{L}(V)$ y λ un valor propio de T. Consideremos $\mathcal{C} = \{v_1, v_2, \dots, v_p\}$ un ciclo correspondiente a λ y U el subespacio generado por $\mathcal{C}$. Sea $E_\lambda = \operatorname{Ker}(T - \lambda \mathrm{Id})$ el subespacio propio correspondiente a λ. Si $v \in U \cap E_\lambda$, entonces existe $a \in \Bbbk$ tal que $v = av_1$.*

Dem. Sea $v \in U \cap E_\lambda$. Como $\mathcal{C}$ es base de U, entonces existen $a_1,\dots,a_p \in \Bbbk$ tales que $v = \sum_{i=1}^p a_i v_i$. Dado que $\mathcal{C}$ es un ciclo, vale

$$T(v_1) = \lambda v_1, \quad T(v_2) = v_1 + \lambda v_2, \quad \cdots \quad , T(v_{p-1}) = v_{p-2} + \lambda v_{p-1}, \quad T(v_p) = v_{p-1} + \lambda v_p.$$

Usando estas fórmulas obtenemos $T(v) = \lambda a_p v_p + \sum_{i=1}^{p-1}(\lambda a_i + a_{i+1})v_i$. Como $v \in E_\lambda$, es $T(v) = \lambda v$. Luego

$$\lambda a_p v_p + \sum_{i=1}^{p-1}(\lambda a_i + a_{i+1})v_i = \sum_{i=1}^{p} \lambda a_i v_i \quad \Rightarrow \quad \lambda a_i + a_{i+1} = \lambda a_i,\ 1 \le i \le p-1 \quad \Rightarrow \quad a_2 = \cdots = a_p = 0.$$

Luego $v = a_1 v_1$. □

Proposición 6.2.2. *Sean $T \in \mathcal{L}(V)$ y λ un valor propio de T. Sean W_λ el subespacio propio generalizado correspondiente y $\mathcal{B}$ una base[5] de W_λ fomada por ciclos disjuntos $\mathcal{C}_1,\dots,\mathcal{C}_k$. Para cada $i = 1,\dots,k$, sea w_i el vector inicial de $\mathcal{C}_i$ y consideremos $\mathcal{B}_0 = \{w_1,\dots,w_k\}$. Entonces $\mathcal{B}_0$ es una base del subespacio propio E_λ y por lo tanto k es la multiplicidad geométrica de λ.*

Dem. Para cada i, sea U_i el subespacio generado por $\mathcal{C}_i$. Como es $\mathcal{B} = \bigcup_{i=1}^k \mathcal{C}_i$, entonces $W_\lambda = \bigoplus_{i=1}^k U_i$. Los vectores iniciales de los ciclos son vectores propios de T, luego $\mathcal{B}_0 = \{w_1,\dots,w_k\}$ está contenido en E_λ. Notar que $\mathcal{B}_0$ está contenido en $\mathcal{B}$, luego $\mathcal{B}_0$ es LI. Probaremos que $\mathcal{B}_0$ es un generador de E_λ. Si $v \in E_\lambda$, como es $E_\lambda \subset W_\lambda$, entonces existen únicos $u_i \in U_i$ tales que $v = \sum_{i=1}^k u_i$. Luego

$$T(v) = \sum_{i=1}^k T(u_i) \quad \Rightarrow \quad \lambda v = \sum_{i=1}^k T(u_i) \quad \Rightarrow \quad \sum_{i=1}^k \lambda u_i = \sum_{i=1}^k T(u_i).$$

Como cada U_i es T-invariante, deducimos $T(u_i) = \lambda u_i$, para todo i. Luego $u_i \in U_i \cap E_\lambda$ y por lo tanto existe $a_i \in \Bbbk$ tal que $u_i = a_i w_i$ (lema anterior). Entonces $v = \sum_{i=1}^k a_i w_i$. Luego $\mathcal{B}_0$ es una base de E_λ. □

Observación 6.2.3. Supongamos que $T \in \mathcal{L}(V)$ es tal que su polinomio característico se escinde y que sus valores propios son $\lambda_1,\dots,\lambda_h$. Sea $\mathcal{B}$ una base de Jordan correspondiente a T. Luego

$$[T]_\mathcal{B} = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_h \end{pmatrix}, \quad A_i = \begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{pmatrix}, \ i = 1,\dots,h.$$

siendo $J_1,\dots,J_k$ los bloques de Jordan correspondientes a λ_i. Entonces para cada $i = 1,\dots,h$, sabemos:

- el orden del bloque A_i es la multiplicidad algebraica de λ_i;
- la cantidad k de bloques de Jordan contenidos en A_i es la multiplicidad geométrica de λ_i;

[5] Notar que $T - \lambda \mathrm{Id}|_{W_\lambda}$ es nilpotente, luego W_λ admite una base de Jordan correspondiente a T.

- la suma de los órdenes de los bloques $J_1, \ldots, J_k$ coincide con el orden de A_i.

La primera afirmación ya la sabíamos. La segunda se deduce de la proposición anterior, dado que cada bloque de Jordan corresponde a un ciclo de la base de Jordan. La tercera afirmación es obvia.

Ejemplo 6.2.4. Sea $T \in \mathcal{L}(\mathbb{R}^3)$ definida por $T(x,y,z) = (3x+y-2z, -x+5z, -x-y+4z)$. Notar que es $T = L_A$, siendo

$$A = \begin{pmatrix} 3 & 1 & -2 \\ -1 & 0 & 5 \\ -1 & -1 & 4 \end{pmatrix}.$$

El polinomio característico de T es $\chi_T(t) = -(t-3)(t-2)^2$, luego los valores propios de T son 2 y 3 y la forma de Jordan de T tiene un bloque A_1 de orden 1 correspondiente al valor propio 3 y un bloque A_2 de orden 2 correspondiente al valor propio 2. Vamos a determinar A_2. Es

$$A - 2I = \begin{pmatrix} 1 & 1 & -2 \\ -1 & -2 & 5 \\ -1 & -1 & 2 \end{pmatrix} \text{ y } \operatorname{rango}(A-2I) = 2.$$

Luego $\mathrm{MG}(2) = 1$. Esto implica A_2 tiene un solo bloque de Jordan, luego la forma de Jordan de T es

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

Ahora obtendremos una base de Jordan $\mathcal{B}$. Como en J hay dos bloques de Jordan, entonces $\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2$, siendo $\mathcal{B}_1$ un ciclo de longitud 1 correspondiente al valor propio 3 y $\mathcal{B}_2$ un ciclo de longitud 2 correspondiente al valor propio 2.
Para el valor propio 3 es $\operatorname{Ker}(T - 3\mathrm{Id}) = [(-1,2,1)]$, luego elegimos $\mathcal{B}_1 = \{(-1,2,1)\}$.
Para el valor propio 2 es

$$(A-2I)^2 = \begin{pmatrix} 2 & 1 & -1 \\ -4 & -2 & 2 \\ -2 & -1 & 1 \end{pmatrix}.$$

Conociendo $A - 2I$ y $(A-2I)^2$ obtenemos

$$\operatorname{Ker}(T - 2\mathrm{Id}) = [(-1,3,1)], \quad \operatorname{Ker}(T - 2\mathrm{Id})^2 = [(1,0,2),(0,1,1)].$$

Sabemos que $\mathcal{B}_2$ un ciclo de longitud 2, luego tenemos que hallar un vector $v \in \operatorname{Ker}(T-2\mathrm{Id})^2 \setminus \operatorname{Ker}(T-2\mathrm{Id})$. Si elegimos $v = (0,1,1)$, entonces $(T - 2\mathrm{Id})(v) = (-1,3,1)$. Luego

$$\mathcal{B}_2 = \{(T-2\mathrm{Id})(v),\, v\} = \{(-1,3,1),(0,1,1)\}.$$

Entonces

$$\mathcal{B} = \mathcal{B}_1 \cup \mathcal{B}_2 = \{(-1,2,1),(-1,3,1),(0,1,1)\}$$

es una base de Jordan tal que $[T]_{\mathcal{B}} = J$.

Ejemplo 6.2.5. Sea $T \in \mathcal{L}(\mathbb{R}_2[x])$ definida por $T(p(x)) = p(x) + p'(x)$.

Si $\mathcal{C} = \{x^2, x, 1\}$, entonces

$$A = [T]_{\mathcal{C}} = \begin{pmatrix} 1 & 0 & 0 \\ 2 & 1 & 0 \\ 0 & 1 & 1 \end{pmatrix}.$$

Luego $\chi_T(t) = -(t-1)^3$ y el único valor propio de T es $\lambda = 1$ con $\mathrm{MA}(1) = 3$. Además,

$$MG(1) = 3 - \operatorname{rango}(A - I) = 3 - \operatorname{rango}\begin{pmatrix} 0 & 0 & 0 \\ 2 & 0 & 0 \\ 0 & 1 & 0 \end{pmatrix} = 1,$$

luego hay un solo bloque de Jordan correspondiente al valor propio 1 y entonces la forma de Jordan de T es

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

Esto nos dice que la base de Jordan es un ciclo de longitud 3 de la forma $\mathcal{B} = \big\{(T - \mathrm{Id})^2(v), (T - \mathrm{Id})(v), v\big\}$, siendo $v \in \operatorname{Ker}(T - \mathrm{Id})^3 \setminus \operatorname{Ker}(T - \mathrm{Id})^2$. Observar que vale

$$(T - \mathrm{Id})\big(p(x)\big) = p'(x), \quad (T - \mathrm{Id})^2\big(p(x)\big) = p''(x), \quad (T - \mathrm{Id})^3\big(p(x)\big) = 0, \quad \forall p(x) \in \mathbb{R}_2[x].$$

Luego $\operatorname{Ker}(T - \mathrm{Id})^3 = \mathbb{R}_2[x]$ y $\operatorname{Ker}(T - \mathrm{Id})^2 = \{ax + b : \ a, b \in \mathbb{R}\}$. Entonces $x^2 \in \operatorname{Ker}(T - \mathrm{Id})^3 \setminus \operatorname{Ker}(T - \mathrm{Id})^2$ y por lo tanto $\mathcal{B} = \{2, 2x, x^2\}$ es una base de Jordan para T y $J = [T]_{\mathcal{B}}$.

Ejemplo 6.2.6. Sea $A = \begin{pmatrix} 2 & 0 & 0 \\ 0 & -1 & 3 \\ 0 & -3 & 5 \end{pmatrix} \in M_3(\mathbb{R})$.

Veremos de hallar su forma de Jordan y la matriz de semejanza correspondiente. Es $\chi_A(t) = -(t-2)^3$, luego 2 es el único valor propio de A. Operando obtenemos $MG(2) = 3 - \operatorname{rango}(A - 2I) = 2$, por lo cual A tiene dos bloques de Jordan. Como la suma de los órdenes de los dos bloques tiene que dar 3, deducimos que la forma de Jordan de A es

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

Luego la base de Jordan correspondiente está formada por dos ciclos de longitudes 2 y 1, es decir tiene la forma $\mathcal{B} = \{(A - 2\mathrm{Id})(u), u, v\}$ siendo $u \in \operatorname{Ker}(A - 2\mathrm{Id})^2 \setminus \operatorname{Ker}(A - 2\mathrm{Id})$ y $v \in \operatorname{Ker}(A - 2\mathrm{Id})$. Observar que es

$$A - 2I = \begin{pmatrix} 0 & 0 & 0 \\ 0 & -3 & 3 \\ 0 & -3 & 3 \end{pmatrix}, \quad (A - 2I)^2 = 0,$$

luego

$$\operatorname{Ker}(A - 2\mathrm{Id}) = [(1,0,0), (0,1,1)], \quad \operatorname{Ker}(A - 2\mathrm{Id})^2 = \mathbb{R}^3$$

Así que es $u \in \mathbb{R}^3 \setminus \operatorname{Ker}(A - 2\mathrm{Id})$. Podemos tomar como u cualquier vector que no esté en $\operatorname{Ker}(A - 2\mathrm{Id})$, por ejemplo $u = (0,0,1)$. Es $(A - 2\mathrm{Id})(u) = (0,3,3) \in \operatorname{Ker}(A - 2\mathrm{Id})$. Ahora tenemos que encontrar un vector en $v \in \operatorname{Ker}(A - 2\mathrm{Id})$ que no sea colineal con $(0,3,3)$, por ejemplo $v = (1,0,0)$. Entonces el teorema 6.1.11 implica que $\mathcal{B} = \{(0,3,3), (0,0,1), (1,0,0)\}$ es LI; luego $\mathcal{B}$ es una base de Jordan tal que $[L_A]_{\mathcal{B}} = J$. Finalmente usando las fórmulas de cambio de base deducimos $A = QJQ^{-1}$, siendo

$$Q = {}_{\mathcal{C}}[\mathrm{Id}]_{\mathcal{B}} = \begin{pmatrix} 0 & 0 & 1 \\ 3 & 0 & 0 \\ 3 & 1 & 0 \end{pmatrix}.$$

Observación 6.2.7. Cuando en la forma de Jordan de un operador los bloques de Jordan correspondientes a un mismo valor propio tienen orden menor que 4, entonces con lo visto anteriormente alcanza para determinar la forma de Jordan. Para órdenes mayores es útil la próxima proposición, que requiere el siguiente resultado. Recordar que si $T \in \mathcal{L}(V)$, entonces

$$\operatorname{Ker}(T) \subset \operatorname{Ker}\big(T^2\big) \subset \operatorname{Ker}\big(T^3\big) \subset \cdots$$

Notar que no puede suceder que todas esas inclusiones sean estrictas, dado que estamos en dimensión finita.

Lema 6.2.8. *Sea $T \in \mathcal{L}(V)$. Si para un cierto número natural k se verifica $\operatorname{Ker}\big(T^k\big) = \operatorname{Ker}\big(T^{k+1}\big)$, entonces vale $\operatorname{Ker}\big(T^k\big) = \operatorname{Ker}\big(T^{k+l}\big)$, para todo $l \geq 1$.*

Dem. La prueba es por inducción en l. El caso $l = 1$ vale por hipótesis. Supongamos $\operatorname{Ker}\big(T^k\big) = \operatorname{Ker}\big(T^{k+l}\big)$, para algún $l \geq 1$, entonces

$$\begin{aligned} v \in \operatorname{Ker}\Big(T^{k+l+1}\Big) &\Rightarrow\ 0 = T^{k+l+1}(v) = T^{k+l}\big(T(v)\big) \ \Rightarrow\ T(v) \in \operatorname{Ker}\Big(T^{k+l}\Big) = \operatorname{Ker}\Big(T^k\Big) \\ &\Rightarrow\ T^{k+1}(v) = T^k(T(v)) = 0 \Rightarrow\ v \in \operatorname{Ker}\Big(T^{k+1}\Big) = \operatorname{Ker}\Big(T^k\Big). \end{aligned}$$

Esto implica $\operatorname{Ker}\big(T^{k+l+1}\big) \subset \operatorname{Ker}\big(T^k\big)$ y por lo tanto $\operatorname{Ker}\big(T^{k+l+1}\big) = \operatorname{Ker}\big(T^k\big)$. ☐

Proposición 6.2.9. *Sea $T \in \mathcal{L}(V)$ tal que su polinomio característico se escinde y λ un valor propio de T con multiplicidad algebraica m. Entonces existe k, con $1 \leq k \leq m$, tal que*

$$\{0\} \subsetneq \operatorname{Ker}(T - \lambda \mathrm{Id}) \subsetneq \cdots \subsetneq \operatorname{Ker}(T - \lambda \mathrm{Id})^k = \operatorname{Ker}(T - \lambda \mathrm{Id})^{k+1} = \cdots = \operatorname{Ker}(T - \lambda \mathrm{Id})^m. \tag{6.8}$$

Este número k es el máximo de las longitudes de los ciclos de T correspondientes a λ.

Dem. La primera afirmación se deduce del lema anterior aplicado a $T - \lambda \mathrm{Id} \in \mathcal{L}(V)$. Consideremos ahora la segunda afirmación. Sea h el máximo de las longitudes de los ciclos de T correspondientes a λ. Si consideramos un ciclo de longitud h

$$\mathcal{C} = \Big\{(T - \lambda \mathrm{Id})^{h-1}(v), \ldots, (T - \lambda \mathrm{Id})(v),\, v\Big\},$$

entonces $v \in \operatorname{Ker}(T - \lambda \mathrm{Id})^h \setminus \operatorname{Ker}(T - \lambda \mathrm{Id})^{h-1}$, luego $\operatorname{Ker}(T - \lambda \mathrm{Id})^{h-1} \subsetneq \operatorname{Ker}(T - \lambda \mathrm{Id})^h$. Esto implica $h \leq k$. Por otro lado, como es $\operatorname{Ker}(T - \lambda \mathrm{Id})^{k-1} \subsetneq \operatorname{Ker}(T - \lambda \mathrm{Id})^k$, tomando $w \in \operatorname{Ker}(T - \lambda \mathrm{Id})^k \setminus \operatorname{Ker}(T - \lambda \mathrm{Id})^{k-1}$, obtenemos que

$$\Big\{(T - \lambda \mathrm{Id})^{k-1}(w), \ldots, (T - \lambda \mathrm{Id})(w),\, w\Big\},$$

es un ciclo de longitud k, luego es $k \leq h$. Así que vale $k = h$. ☐

Observación 6.2.10. Dado un valor propio λ de T, si J es la forma de Jordan de T, entonces hay una correspondencia uno a uno entre los bloques de Jordan de orden l correspondientes a λ y los ciclos de T de longitud l correspondientes a λ. Luego en la forma de Jordan de T, el entero k de la proposición anterior es el tamaño del mayor bloque de Jordan correspondiente a λ.

Observación 6.2.11. Para aplicar la proposición anterior, notar que (6.8) equivale a

$$\dim \operatorname{Ker}(T - \lambda \mathrm{Id}) < \cdots < \dim \operatorname{Ker}(T - \lambda \mathrm{Id})^k = \dim \operatorname{Ker}(T - \lambda \mathrm{Id})^{k+1} \cdots = \dim \operatorname{Ker}(T - \lambda \mathrm{Id})^m.$$

A su vez, usando $\dim \operatorname{Ker}(T - \lambda \mathrm{Id})^k + \dim \operatorname{Im}(T - \lambda \mathrm{Id})^k = \dim V$, obtenemos que esto equivale a

$$\operatorname{rango}(T - \lambda \mathrm{Id}) > \cdots > \operatorname{rango}(T - \lambda \mathrm{Id})^k = \operatorname{rango}(T - \lambda \mathrm{Id})^{k+1} \cdots = \operatorname{rango}(T - \lambda \mathrm{Id})^m.$$

Luego k es el menor valor para el cual el rango de $(T - \lambda \mathrm{Id})^k$ coincide con el rango de $(T - \lambda \mathrm{Id})^{k+1}$.

Ejemplo 6.2.12. Sea $A = \begin{pmatrix} 1 & 3 & -2 & 1 \\ -1 & 4 & -1 & 1 \\ -1 & 2 & 1 & 1 \\ 0 & 1 & -1 & 2 \end{pmatrix} \in M_4(\mathbb{R})$. Es un ejercicio el verificar que es $\chi_A(t) = (t-2)^4$.

Luego su único valor propio es $\lambda = 2$. Notar $A - 2I = \begin{pmatrix} -1 & 3 & -2 & 1 \\ -1 & 2 & -1 & 1 \\ -1 & 2 & -1 & 1 \\ 0 & 1 & -1 & 0 \end{pmatrix}$. Dejamos como ejercicio el verificar que $A-2I$ tiene rango 2. Esto implica $\mathrm{MG}(2) = 2$ y por lo tanto hay dos bloques de Jordan correspondientes al valor propio 2. Como la suma de sus órdenes tienen que dar $\mathrm{MA}(2) = 4$, deducimos que los órdenes son $(3,1)$ o $(2,2)$. Operando obtenemos $(A-2I)^2 = 0$, luego $\mathrm{Ker}(A-2I)^2 = \mathbb{R}^4$. Así tenemos

$$\{0\} \subsetneq \mathrm{Ker}(A-2I) \subsetneq \mathrm{Ker}(A-2I)^2 = \mathrm{Ker}(A-2I)^3 = \mathrm{Ker}(A-2I)^4 = \mathbb{R}^4$$

Luego el tamaño del mayor bloque de Jordan correspondiente a $\lambda = 2$ es 2. Esto nos dice que la forma de Jordan de A es

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

A su vez esto implica que la base de Jordan correspondiente es del tipo

$$\mathcal{B} = \{(A-2I)(v),\, v,\, (A-2I)(w),\, w\}$$

con $v, w \in \mathrm{Ker}(A-2I)^2 \setminus \mathrm{Ker}(A-2I)$. Notar $\mathrm{Ker}(A-2I)^2 = \mathbb{R}^4$. Operando obtenemos

$$\mathrm{Ker}(A-2I) = \big\{(x,y,z,t) \in \mathbb{R}^4 :\ x = y+t,\ z = y\big\} = [(1,1,1,0),\, (1,0,0,1)].$$

Tenemos que elegir dos vectores de $\mathbb{R}^4$ que no estén en $\mathrm{Ker}(A-2I)$ (y que sean LI entre sí), por ejemplo $v = (1,0,0,0)$ y $w = (0,1,0,0)$. Entonces $(A-2I)(v) = (-1,-1,-1,0)$ y $(A-2I)(w) = (3,2,2,1)$. Como estos últimos vectores son LI, entonces la unión de los ciclos correspondientes da un conjunto LI. Luego una base de Jordan correspondiente a J es

$$\mathcal{B} = \{(-1,-1,-1,0),\, (1,0,0,0),\, (3,2,2,1),\, (0,1,0,0)\}.$$

Esto implica $A = PJP^{-1}$, siendo P la matriz cuyas columnas son los vectores de $\mathcal{B}$ escritos en forma vertical.

Observación 6.2.13. En el ejemplo anterior elegimos los vectores $v, w \in \mathrm{Ker}(A-2I)^2 \setminus \mathrm{Ker}(A-2I)$ de forma tal que sean LI, porque queremos que $\mathcal{B} = \{(A-2I)(v),\, v,\, (A-2I)(w),\, w\}$ sea una base. Hay que tener en cuenta que el que $\{v, w\}$ sea LI no implica que $\{(A-2I)(v),\, (A-2I)(w)\}$ sea LI, así que esto último siempre tenemos que verificarlo. Por ejemplo, si hubiésemos elegido $v = (1,1,0,1)$ y $w = (0,1,0,0)$, hubiésemos obtenido $(A-2I)(u) = (A-2I)(v) = (3,2,2,1)$ y por lo tanto $\mathcal{B}$ no sería una base de $\mathbb{R}^4$.
En caso de tener mala suerte y que $\{(A-2I)(v),\, (A-2I)(w)\}$ sea LD, entonces hay que elegir otro par de vectores v, w y probar de nuevo hasta que nos dé LI (el teorema de Jordan nos dice que hay solución).

Ejemplo 6.2.14. Consideremos $T \in \mathcal{L}\big(\mathbb{R}^{10}\big)$ del cual se tiene los siguientes datos.

1. Su polinomio característico es $\chi_T(t) = (t-7)^6(t-5)^4$.
2. $\mathrm{rango}(T - 7\mathrm{Id}) = 8,\ \mathrm{rango}(T - 7\mathrm{Id})^2 = 6,\ \mathrm{rango}(T - 7\mathrm{Id})^3 = \mathrm{rango}(T - 7\mathrm{Id})^4 = 4.$
3. $\mathrm{rango}(T - 5\mathrm{Id}) = 7.$

De $\chi_T(t) = (t-7)^6(t-5)^4$ deducimos que la forma de Jordan J de T verifica $J = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix} \in M_{10}(\mathbb{R})$, siendo $A_1 \in M_6(\mathbb{R})$ correspondiente al valor propio 7 y $A_2 \in M_4(\mathbb{R})$ correspondiente al valor propio 5. Para encontrar J solo nos resta hallar A_1 y A_2.

Consideremos A_1. La multiplicidad geométrica de 7 es $\mathrm{MG}(7) = 10 - \mathrm{rango}(T - 7\mathrm{Id}) = 2$, luego A_1 contiene 2 bloques de Jordan. La sucesión de rangos verifica

$$\mathrm{rango}(T - 7\mathrm{Id}) > \mathrm{rango}(T - 7\mathrm{Id})^2 > \mathrm{rango}(T - 7\mathrm{Id})^3 = \mathrm{rango}(T - 7\mathrm{Id})^4.$$

Luego el mayor orden de un bloque de Jordan correspondiente a 7 es 3. Como A_1 tiene orden 6, deducimos

$$A_1 = \begin{pmatrix} 7 & 1 & 0 & 0 & 0 & 0 \\ 0 & 7 & 1 & 0 & 0 & 0 \\ 0 & 0 & 7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 7 & 1 & 0 \\ 0 & 0 & 0 & 0 & 7 & 1 \\ 0 & 0 & 0 & 0 & 0 & 7 \end{pmatrix}.$$

Consideremos ahora A_2. La multiplicidad geométrica de 5 es $\mathrm{MG}(5) = 10 - \mathrm{rango}(T - 5\mathrm{Id}) = 3$, luego A_2 contiene 3 bloques de Jordan. Como la suma de los órdenes de los bloques tiene que dar el orden de A_2 que es 4, deducimos

$$A_2 = \begin{pmatrix} 5 & 1 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}.$$

Observaciones 6.2.15.

1. Si todos los valores propios del operador tienen multiplicidad algebraica menor que 7, entonces para obtener su forma de Jordan alcanza con aplicar lo que vimos en esta sección. Si alguno tiene multiplicidad algebraica 7, entonces hay solo un caso que no sabemos determinar, que es cuando los órdenes de los bloques de Jordan correspondientes son $(3,2,2)$ y $(3,3,1)$, todos los otros casos se obtienen sin problema.

2. Hay un método general para hallar la base de Jordan (conociendo los valores propios), pero es bastante engorroso y en dimensiones bajas lo que vimos funciona más rápido.

3. Si λ es un valor propio de T, entonces su multiplicidad geométrica nos da la cantidad de bloques de Jordan correspondientes. En la proposición 6.2.9 vimos cómo hallar el tamaño del bloque mayor; trabajando con más cuidado se obtiene un algoritmo para conocer los tamaños de todos los bloques (ver, por ejemplo, *Linear Algebra*, de S. M. Friedberg, et al.). Esto junto con la observación 6.2.3, implican que la forma de Jordan queda determinada por el operador y por lo tanto es única[6].

[6] La unicidad es a menos de reordenar los bloques de Jordan.

Capítulo 7

Polinomio minimal

En este capítulo V es un $\Bbbk$-espacio vectorial de dimensión finita.

7.1. Polinomio minimal

Recordar que un polinomio $p(t)$ es *mónico* si es no nulo y el coeficiente de su término de mayor grado es 1, es decir si es de la forma $p(t) = t^m + a_{m-1}\,t^{m-1} + \cdots + a_1\,t + a_0,$ con $m \geq 1$.

Teorema 7.1.1. *Dado $T \in \mathcal{L}(V)$, existe un único polinomio $m_T(t) \in \Bbbk[t]$ que verifica:*

1. *$m_T(T) = 0$.*
2. *$m_T(t)$ es de grado mínimo entre los polinomios no nulos que se anulan en T.*
3. *$m_T(t)$ es mónico.*

Además $m_T(t)$ divide a todo polinomio que se anule en T.

Dem. En el teorema de Cayley-Hamilton probamos que el polinomio característico $\chi_T(t)$ verifica $\chi_T(T) = 0$. Luego existe algún polinomio no nulo que se anula en T. Entonces tomando un polinomio de grado mínimo que verifique lo anterior y dividiéndolo por el coeficiente de su término de mayor grado, obtenemos un polinomio $m_T(t)$ que verifica las tres condiciones.

Probaremos ahora que $m_T(t)$ divide a todo polinomio que se anule en T. Sea $q(t) \in \Bbbk[t]$ que verifique $q(T) = 0$. Dividimos $q(t)$ por $m_T(t)$ y es $q(t) = m_T(t)\,d(t) + r(t)$ con $r(t) = 0$ o $r(t) \neq 0$ y $\operatorname{gr} r(t) < \operatorname{gr} m(t)$. Teniendo en cuenta la primera condición obtenemos:

$$r(T) = q(T) - m_T(T) \circ d(T) = 0 - 0 \circ d(T) = 0.$$

Si fuese $r(t) \neq 0$ tendríamos una contradicción con la minimalidad del grado de $m_T(t)$, luego necesariamente es $r(t) = 0$ y por lo tanto $m_T(t)$ divide a $q(t)$.

Ahora probaremos la unicidad. Supongamos ahora que $p(t)$ es otro polinomio que verifica las tres condiciones. Como $p(t)$ verifica la primer condición, entonces $m_T(t)$ divide a $p(t)$. Cambiando los roles de $p(t)$ y $m_T(t)$ obtenemos también que $p(t)$ divide a $m_T(t)$. Luego ambos se dividen mutuamente y por lo tanto existe $a \in \Bbbk$ tal que $p(t) = a\,m_T(t)$. Como $p(t)$ y $m(t)$ son mónicos, deducimos $a = 1$; luego $p(t) = m_T(t)$. $\square$

Definición 7.1.2. El polinomio $m_T(t)$ del teorema anterior se llama el *polinomio minimal* de T.

Observación 7.1.3. El teorema de Cayley-Hamilton nos dice que el polinomio característico de T se anula en T, luego el polinomio minimal de T divide al característico.

Ejemplos 7.1.4. Los siguientes son casos en que es fácil hallar el polinomio minimal de $T \in \mathcal{L}(V)$.

1. Vale $m_T(t) = t$ si y solo si $T = 0$.
2. Vale $m_T(t) = t - \lambda$, para cierto $\lambda \in \Bbbk$, si y solo si $T = \lambda \mathrm{Id}$. En particular $m_{\mathrm{Id}}(t) = t - 1$.
3. Vale $m_T(t) = t^2 - t$ si y solo si T es una proyección ($T^2 = T$) tal que $T \neq 0$ y $T \neq \mathrm{Id}$.

En forma análoga a lo anterior se prueba que, dada una matriz $A \in M_n(\Bbbk)$, existe un único polinomio $m_A(t) \in \Bbbk[t]$ llamado el *polinomio minimal* de A que verifica

1. $m_A(A) = 0$,
2. $m_A(t)$ es de grado mínimo entre los polinomios no nulos que se anulan en A,
3. $m_A(t)$ es mónico.

Además, si $q(t) \in \Bbbk[t]$ es tal que $q(A) = 0$, entonces $m_A(t)$ divide a $q(t)$; en particular $m_A(t)$ divide a $\chi_A(t)$.

Proposición 7.1.5. *Sea $\mathcal{B}$ una base de V y $A = [T]_{\mathcal{B}}$. Entonces $m_T(t) = m_A(t)$.*

Dem. Sabemos que vale $p([T]_{\mathcal{B}}) = [p(T)]_{\mathcal{B}}$, para todo $p(t) \in \Bbbk[t]$. Luego

$$0 = m_A(A) = m_A([T]_{\mathcal{B}}) = [m_A(T)]_{\mathcal{B}} \quad \Rightarrow \quad m_A(T) = 0 \quad \Rightarrow \quad m_T(t) | m_A(t),$$
$$m_T(A) = m_T([T]_{\mathcal{B}}) = [m_T(T)]_{\mathcal{B}} = [0]_{\mathcal{B}} = 0 \quad \Rightarrow \quad m_T(A) = 0 \quad \Rightarrow \quad m_A(t) | m_T(t).$$

Como ambos polinomios son mónicos, la única posibilidad es $m_T(t) = m_A(t)$. □

Usando que si $A \in M_n(\Bbbk)$, entonces $[L_A]_{\mathcal{B}} = A$, siendo $\mathcal{B}$ la base canónica de $\Bbbk^n$, se deduce lo siguiente.

Corolario 7.1.6. *Si $A \in M_n(\Bbbk)$, entonces $m_{L_A}(t) = m_A(t)$.* □

Ejemplo 7.1.7. Si consideramos una matriz escalar[1] $A = \lambda I \in M_n(\Bbbk)$, entonces es $L_A = \lambda \mathrm{Id}$. Luego $m_A(t) = t - \lambda$. Notar que también vale el recíproco: si $m_A(t) = t - \lambda$, entonces $A = \lambda I$.

Sabemos que el polinomio minimal $m_T(t)$ divide al característico $\chi_T(t)$. Esto implica que las raíces de $m_T(t)$ también son raíces de $\chi_T(t)$. A continuación veremos que vale también el recíproco.

Lema 7.1.8. *Sea $p(t) \in \Bbbk[t]$ y v un vector propio de T correspondiente a un valor propio λ, entonces*

$$p(T)\,(v) = p(\lambda)\,v.$$

Dem. En la proposición 5.4.4 vimos que vale $T^n(v) = \lambda^n v$, para todo $n \in \mathbb{N}$. Luego si $p(t) = \sum_{i=0}^n a_i\, t^i$, entonces

$$p(T)\,(v) = \left(\sum_{i=0}^n a_i\, T^i\right)(v) = \sum_{i=0}^n a_i\, T^i(v) = \sum_{i=0}^n a_i\, \lambda^i\, v = \left(\sum_{i=0}^n a_i\, \lambda^i\right) v = p(\lambda)\, v. \quad \square$$

Teorema 7.1.9. *El polinomio característico de T y el polinomio minimal de T tienen las mismas raíces.*

Dem. Como el polinomio minimal divide al polinomio característico, entonces las raíces del polinomio minimal también son raíces del característico. Veremos ahora el recíproco. Si $\lambda \in \Bbbk$ es tal que $\chi_T(\lambda) = 0$, entonces λ es un valor propio de T y por lo tanto existe $v \neq 0$ tal que $T(v) = \lambda\, v$. Por el lema anterior es $m_T(\lambda)\, v = m_T(T)\,(v) = 0(v) = 0$ y como es $v \neq 0$, deducimos $m_T(\lambda) = 0$. □

[1] Recordar que a las matrices cuadradas de la forma $A = \lambda I$ se les llama matrices *escalares*.

Observación 7.1.10. Este teorema implica que si $T \in \mathcal{L}(V)$ es tal que $\chi_T(t)$ escinde y se escribe de la forma

$$\chi_T(t) = (-1)^n (t-\lambda_1)^{n_1} \cdots (t-\lambda_h)^{n_h}\,, \text{ con } \lambda_i \neq \lambda_j \text{ si } i \neq j,$$

entonces su polinomio minimal es de la forma

$$m_T(t) = (t-\lambda_1)^{m_1} \cdots (t-\lambda_h)^{m_h}\,, \text{ con } 1 \leq m_i \leq n_i,\ i = 1, \ldots, h.$$

Observación 7.1.11. Si $A = \begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_h \end{pmatrix}$ es una matriz diagonal en bloques, entonces para cada $l \in \mathbb{N}$ vale $A^l = \begin{pmatrix} A_1^l & & \\ & \ddots & \\ & & A_h^l \end{pmatrix}$. Esto implica $p(A) = \begin{pmatrix} p(A_1) & & \\ & \ddots & \\ & & p(A_h) \end{pmatrix}$, para todo $p(t) \in \Bbbk[t]$. En particular esto último vale también para matrices diagonales (el caso en que $A_1, \ldots, A_h$ son matrices 1×1).

Lema 7.1.12. *Sea T un operador tal que su polinomio minimal es de la forma*

$$m_T(t) = (t-\lambda_1) \cdots (t-\lambda_h)\,,$$

con $\lambda_i \neq \lambda_j$ si $i \neq j$. Entonces para cada valor propio λ_i, el subespacio propio generalizado W_{λ_i} coincide con el subespacio propio E_{λ_i}.

Dem. Sean $i \in \{1, \ldots, n\}$ y $v \in W_{\lambda_i}$ arbitrarios. Sabemos $m_T(T) = 0$, luego $m_T(T)(v) = 0$. Escribamos $m_T(t) = p(t)(t-\lambda_i)$, con $p(t) = \prod_{j \neq i}(t-\lambda_j)$. Luego

$$0 = m_T(T)(v) = \big(p(T) \circ (T - \lambda_i \mathrm{Id})\big)(v) = p(T)\big((T-\lambda_i \mathrm{Id})(v)\big) \tag{7.1}$$

Como W_{λ_i} es T-invariante, para todo j, entonces $(T-\lambda_i \mathrm{Id})(v) \in W_{\lambda_i}$. Además sabemos que $(T-\lambda_j \mathrm{Id})|_{W_{\lambda_i}} \in \mathcal{L}(W_{\lambda_i})$ es un isomorfismo para todo $j \neq i$ (proposición 5.4.5), luego

$$p(T)|_{W_{\lambda_i}} = (T-\lambda_1 \mathrm{Id})|_{W_{\lambda_i}} \circ \cdots \circ (T-\lambda_{i-1} \mathrm{Id})|_{W_{\lambda_i}} \circ (T-\lambda_{i+1} \mathrm{Id})|_{W_{\lambda_i}} \circ \cdots \circ (T-\lambda_h \mathrm{Id})|_{W_{\lambda_i}}$$

es un isomorfismo. Entonces (7.1) implica $(T-\lambda_i \mathrm{Id})(v) = 0$ y por lo tanto $v \in \operatorname{Ker}(T-\lambda_i \mathrm{Id}) = E_{\lambda_i}$. Luego probamos $W_{\lambda_i} \subset E_{\lambda_i}$ y por lo tanto $W_{\lambda_i} = E_{\lambda_i}$. □

Observación 7.1.13. Dado $T \in \mathcal{L}(V)$, si $W \subset V$ es un subespacio T-invariante, entonces $m_{T|_W}(t)$ divide a $m_T(t)$. Esto se debe a que vale $m_T(T)(w) = 0$, para todo $w \in W$ y por lo tanto $m_T(T|_W) = m_T(T)|_W = 0$.

Teorema 7.1.14. *Un operador T es diagonalizable si y solo su polinomio minimal es de la forma*

$$m_T(t) = (t-\lambda_1) \cdots (t-\lambda_h)\,, \tag{7.2}$$

para ciertos $\lambda_1, \ldots, \lambda_h \in \Bbbk$ distintos entre sí.

Dem. Supongamos que $T \in \mathcal{L}(V)$ es diagonalizable. Entonces $V = \bigoplus_{i=1}^h E_{\lambda_i}$, siendo cada E_{λ_i} el subespacio propio correspondiente al valor propio λ_i. Esto implica que el polinomio característico de T es de la forma $\chi_T(t) = (-1)^n (t-\lambda_1)^{n_1} \cdots (t-\lambda_h)^{n_h}$, para ciertos $n_i \geq 1$. Sabemos que $m_T(t)$ divide a $\chi_T(t)$ y que tiene las mismas raíces, así que es $m_T(t) = (t-\lambda_1)^{m_1} \cdots (t-\lambda_h)^{m_h}$, para ciertos m_i tales que $1 \leq m_i \leq n_i$. Consideremos $p(t) = (t-\lambda_1) \cdots (t-\lambda_h)$. Para cada i, podemos escribir $m_T(t) = q(t)(t-\lambda_i)$, siendo $q(t) = \prod_{j \neq i}(t-\lambda_j)$. Esto implica $p(T) = q(T) \circ (T-\lambda_i \mathrm{Id})$. Como $T - \lambda_i \mathrm{Id}$ se anula en E_{λ_i}, entonces $p(T)$ se anula en E_{λ_i}, para cada i. Como es $V = \bigoplus_{i=1}^h E_{\lambda_i}$, concluimos que $p(T)$ se anula en todo V. Luego $p(t) = (t-\lambda_1)\cdots(t-\lambda_h)$ es el polinomio minimal de T.

Supongamos ahora que es $m_T(t) = (t-\lambda_1)\cdots(t-\lambda_h)$, con $\lambda_i \neq \lambda_j$ si $i \neq j$. Probaremos por inducción en h que T es diagonalizable. La idea es la misma que en la prueba del teorema 5.4.7.
Si $h = 1$, entonces $m_T(t) = t - \lambda_1$; esto implica $T = \lambda_1 \mathrm{Id}$ y por lo tanto $V = E_{\lambda_1} = \mathrm{Ker}(T - \lambda_1 \mathrm{Id})$.
Razonando inductivamente, supongamos que es $m_T(t) = (t-\lambda_1)\cdots(t-\lambda_h)$ y que la tesis vale cuando hay $h-1$ factores. Notar que el lema anterior implica $W_{\lambda_i} = E_{\lambda_i}$, para todo i. Consideremos el subespacio $U = \mathrm{Im}(T - \lambda_h \mathrm{Id})$. Como U es T-invariante, entonces $m_{T|_U}(t)$ divide a $m_T(t)$. Probaremos $m_{T|U}(t) = (t-\lambda_1)\cdots(t-\lambda_{h-1})$.
Para cada $i < h$ el operador $T - \lambda_h \mathrm{Id}$ restringido a $W_{\lambda_i} = E_{\lambda_i}$ es un isomorfismo (proposición 5.4.5). Luego

$$E_{\lambda_i} = (T - \lambda_h \mathrm{Id})(E_{\lambda_i}) \subset (T - \lambda_h \mathrm{Id})(V) = U \quad \Rightarrow \quad E_{\lambda_i} \subset U.$$

Entonces U contiene a los subespacios propios $E_{\lambda_1}, \ldots, E_{\lambda_{h-1}}$ y por lo tanto $\lambda_1, \ldots, \lambda_{h-1}$ son valores propios de $T|_U$, lo cual implica que $\lambda_1, \ldots, \lambda_{h-1}$ son raíces de $m_{T|U}(t)$. Si juntamos esto con que $m_{T|_U}(t)$ divide a $m_T(t) = (t-\lambda_1)\cdots(t-\lambda_h)$, obtenemos que $(t-\lambda_1)\cdots(t-\lambda_{h-1})$ divide a $m_{T|U}(t)$.
Consideremos ahora $v \in \mathrm{Ker}(T|_U - \lambda_h \mathrm{Id}_U) = \mathrm{Ker}(T - \lambda_h \mathrm{Id}) \cap U$. Como $v \in U$, entonces existe $w \in V$ tal que $v = (T - \lambda_h \mathrm{Id})(w)$. Luego

$$0 = (T - \lambda_h \mathrm{Id})(v) = (T - \lambda_h \mathrm{Id})^2(w) \quad \Rightarrow \quad w \in W_{\lambda_h} = E_{\lambda_h} \quad \Rightarrow \quad v = (T - \lambda_h \mathrm{Id})(w) = 0.$$

Entonces λ_h no es valor propio de $T|_U$ y por lo tanto $m_{T|U}(t) = (t-\lambda_1)\cdots(t-\lambda_{h-1})$. Aplicando la hipótesis inductiva a $T|_U$ obtenemos que $T|_U$ es diagonalizable. Luego $U = \bigoplus_{i=1}^{h-1} \mathrm{Ker}(T|_U - \lambda_i \mathrm{Id}_U)$. Como es $\mathrm{Ker}(T|_U - \lambda_i \mathrm{Id}_U) = E_{\lambda_i} \cap U$ y $E_{\lambda_i} \subset U$, para todo $i < h$, deducimos $U = \bigoplus_{i=1}^{h-1} E_{\lambda_i}$.
Sea ahora $v \in V$. Entonces $(T - \lambda_h \mathrm{Id})(v) \in U$, luego existen $v_i \in E_{\lambda_i}$ tales que $(T - \lambda_h \mathrm{Id})(v) = \sum_{i=1}^{h-1} v_i$. Como $T - \lambda_h \mathrm{Id}$ restringido a $W_{\lambda_i} = E_{\lambda_i}$ es un isomorfismo para todo $i < h$, deducimos que existen $u_i \in E_{\lambda_i}$ tales que $v_i = (T - \lambda_h \mathrm{Id})(u_i)$, para todo $i < h$. Luego

$$(T - \lambda_h \mathrm{Id})(v) = \sum_{i=1}^{h-1} (T - \lambda_h \mathrm{Id})(u_i) = (T - \lambda_h \mathrm{Id})\left(\sum_{i=1}^{h-1} u_i\right) \quad \Rightarrow \quad v - \sum_{i=1}^{h-1} u_i \in \mathrm{Ker}(T - \lambda_h \mathrm{Id}) = E_{\lambda_h}.$$

Entonces existe $u_h \in E_{\lambda_h}$ tal que $v - \sum_{i=1}^{h-1} u_i = u_h$ y por lo tanto $v = \sum_{i=1}^h u_i$, con $u_i \in E_{\lambda_i}$ para todo i. Esto prueba $V = \sum_{i=1}^h E_{\lambda_i}$, lo cual implica que T es diagonalizable. $\square$

Observación 7.1.15. Algo interesante del teorema anterior es que no estamos asumiendo ninguna hipótesis sobre el cuerpo $\Bbbk$ ni sobre el polinomio característico del operador. Por ejemplo, este teorema vale para $T \in \mathcal{L}(V)$, siendo V un $\mathbb{R}$-espacio vectorial o un $\mathbb{Z}_p$-espacio vectorial (p primo).

Corolario 7.1.16. *Si existe un polinomio de la forma $p(t) = a(t-\lambda_1)\cdots(t-\lambda_k)$, con $\lambda_i \neq \lambda_j$ si $i \neq j$ y $0 \neq a \in \Bbbk$ tal que $p(T) = 0$, entonces T es diagonalizable.*

Dem. Como vale $p(T) = 0$, entonces $m_T(t)$ divide a $p(t)$ y por lo tanto se aplica el teorema anterior. $\square$

Ejemplo 7.1.17. Sea $T \in \mathcal{L}(\mathbb{R}_2[x])$ definida por $T(p(x)) = p'(x)$. Observar que vale

$$T\big(ax^2 + bx + c\big) = 2ax + b, \quad T^2\big(ax^2 + bx + c\big) = 2a, \quad T^3\big(ax^2 + bx + c\big) = 0.$$

Luego $T^3 = 0$ y $T^2 \neq 0$. Esto implica $m_T(t) = t^3$ y por lo tanto T no es diagonalizable.

Ejemplo 7.1.18. Sea $T \in \mathcal{L}(\mathbb{R}^3)$ definida por $T(x,y,z) = (3x - y, 2y, x - y + 2z)$. Notar que es $T = L_A$, siendo $A = \begin{pmatrix} 3 & -1 & 0 \\ 0 & 2 & 0 \\ 1 & -1 & 2 \end{pmatrix}$. Es $\chi_T(t) = -(t-2)^2(t-3)$, luego hay solo dos posibilidades para el minimal

$$m_T(t) = \begin{cases} (t-2)(t-3) \\ (t-2)^2(t-3) \end{cases}.$$

Calculando el producto obtenemos $(A - 2I)(A - 3I) = 0$, luego $m_T(t) = (t-2)(t-3)$. Esto implica que T es diagonalizable y al ser $\chi_T(t) = -(t-2)^2(t-3)$, deducimos que existe una base $\mathcal{B}$ de $\mathbb{R}^3$ tal que $[T]_{\mathcal{B}} = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 3 \end{pmatrix}$. Para determinar $\mathcal{B}$ hay que hallar los vectores propios, como vimos en diagonalización.

Lo que sigue es la versión matricial de lo visto anteriormente, que resumimos en la siguiente proposición.

Proposición 7.1.19. *Sea $A \in M_n(\Bbbk)$.*

1. *La matriz A es diagonalizable si y solo si su polinomio minimal es de la forma $(t - \lambda_1) \cdots (t - \lambda_h)$, con $\lambda_i \neq \lambda_j$ si $i \neq j$.*

2. *Si existe un polinomio de la forma $p(t) = a(t - \lambda_1) \cdots (t - \lambda_k) \in \Bbbk[t]$, con $\lambda_i \neq \lambda_j$ si $i \neq j$ y $0 \neq a \in \Bbbk$, tal que $p(A) = 0$, entonces A es diagonalizable.*

3. *Los polinomios $m_A(t)$ y $\chi_A(t)$ tienen las mismas raíces.* □

Ejemplo 7.1.20. Sea $A \in M_n(\mathbb{R})$ tal que $A^3 = A$. Entonces vale $p(A) = 0$, siendo $p(t) = t^3 - t$. Observar que es $p(t) = t^3 - t = t(t-1)(t+1)$, luego A es diagonalizable.

Ejemplo 7.1.21. Veamos cómo determinar todas las matrices reales 2×2 que verifican $A^2 - 3A + 2I = 0$. Sea $p(t) = t^2 - 3t + 2$. Es $p(A) = 0$, luego $m_A(t)$ divide a $p(t)$. Como $p(t) = (t-1)(t-2)$, entonces tenemos las siguientes posibilidades

$$m_A(t) = t - 1, \quad m_A(t) = t - 2, \quad m_A(t) = (t-1)(t-2).$$

Si $m_A(t) = t - 1$, entonces $A = I$. Si $m_A(t) = t - 2$, entonces $A = 2I$. Si $m_A(t) = (t-1)(t-2)$, entonces A es diagonalizable con valores propios 1 y 2, y por lo tanto es de la forma

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix}^{-1},$$

siendo $\left(\begin{smallmatrix} a & b \\ c & d \end{smallmatrix}\right)$ una matriz arbitraria tal que $ad - bc \neq 0$..

7.2. Relación con la forma de Jordan

En lo que sigue veremos que el polinomio minimal también brinda información sobre la forma de Jordan.

Proposición 7.2.1. 1. *Si*

$$J = \begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix} \in M_n(\Bbbk).$$

Entonces $m_J(t) = (t - \lambda)^n$.

2. *Si*

$$A=\begin{pmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{pmatrix}\in M_n(\Bbbk),\quad J_i=\begin{pmatrix} \lambda & 1 & & & \\ & \lambda & 1 & & \\ & & \ddots & \ddots & \\ & & & \lambda & 1 \\ & & & & \lambda \end{pmatrix}\in M_{p_i}(\Bbbk),$$

con $p_1 \geq p_2 \geq \cdots \geq p_k$. *Entonces* $m_A(t) = (t-\lambda)^{p_1}$.

3. *Supongamos*

$$A=\begin{pmatrix} A_1 & & \\ & \ddots & \\ & & A_h \end{pmatrix}\in M_n(\Bbbk),$$

en que cada bloque $A_i \in M_{n_i}(\Bbbk)$ *es como en la parte anterior correspondiente a cierto escalar* λ_i. *Supongamos que los* λ_i *son distintos entre sí y que* q_i *es el tamaño del primer bloque de Jordan contenido en* A_i, *para todo* i . *Entonces*

$$m_A(t) = (t-\lambda_1)^{q_1}\cdots(t-\lambda_h)^{q_h}.$$

Dem. (1). Sea

$$\tilde{J} = J-\lambda I=\begin{pmatrix} 0 & 1 & & & \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ & & & & 0 \end{pmatrix}\in M_n(\Bbbk).$$

Si consideramos $T = L_{\tilde{J}} \in \mathcal{L}(\Bbbk^n)$, entonces es $T(x_1, x_2, \ldots, x_n) = (x_2, \ldots, x_n, 0)$, para todo $(x_1, \ldots, x_n) \in \Bbbk^n$. Luego

$$\begin{aligned} T(x_1,x_2,\ldots,x_n) &= (x_2,\ldots,x_n,0),\\ T^2(x_1,x_2,\ldots,x_n) &= (x_3,\ldots,x_n,0,0),\\ &\vdots\\ T^{n-1}(x_1,x_2,\ldots,x_n) &= (x_n,0,\ldots,0),\\ T^{n}(x_1,x_2,\ldots,x_n) &= (0,\ldots,0). \end{aligned}$$

Lo anterior implica[2] $T^n = 0$ y $T^{n-1} \neq 0$, lo cual equivale a $(J-\lambda I)^n = 0$ y $(J-\lambda I)^{n-1} \neq 0$. De $(J-\lambda I)^n = 0$ deducimos $m_J(t) = (t-\lambda)^r$, para algún $1 \leq r \leq n$. Pero como es $(J-\lambda I)^{n-1} \neq 0$, concluimos $m_J(t) = (t-\lambda)^n$.

(2). Por la parte anterior sabemos que es $m_{J_i}(t) = (t-\lambda)^{p_i}$, para todo $i = 1, \ldots, k$. Como estamos asumiendo $p_1 \geq p_2 \geq \cdots \geq p_k$, es

$$m_{J_1}(J_i) = (J_i-\lambda I)^{p_1} = (J_i-\lambda I)^{p_1-p_i}\cdot(J_i-\lambda I)^{p_i} = (J_i-\lambda I)^{p_1-p_i}\cdot 0 = 0,\quad \forall i=1,\ldots,k.$$

Luego

$$m_{J_1}(A)=\begin{pmatrix} m_{J_1}(J_1) & & \\ & \ddots & \\ & & m_{J_1}(J_k) \end{pmatrix}=0. \tag{7.3}$$

[2] Esto también se puede probar calculando las potencias de $\tilde{J}$, observando que en las potencias sucesivas la diagonal de unos va subiendo hasta desaparecer al llegar a $\tilde{J}^n$.

Entonces $m_A(t)$ divide a $m_{J_1}(t) = (t-\lambda)^{p_1}$ y por lo tanto es de la forma $m_A(t) = (t-\lambda)^l$, para cierto $l \leq p_1$. Pero si consideramos $p(t) = (t-\lambda)^l$ con $l < p_1$, entonces $p(J_1) = (J_1 - \lambda I)^l \neq 0$. Luego razonando como en (7.3) obtenemos $p(A) \neq 0$. Esto implica $m_A(t) = (t-\lambda)^{p_1}$.

(3). Como la matriz A es triangular superior, entonces $\chi_A(t) = (-1)^n (t-\lambda_1)^{n_1} \cdots (t-\lambda_h)^{n_h}$, con $\lambda_i \neq \lambda_j$ si $i \neq j$. Luego $m_A(t) = (t-\lambda_1)^{m_1} \cdots (t-\lambda_h)^{m_h}$, siendo $1 \leq m_i \leq n_i$, $\forall i = 1, \ldots, h$.

Sea $p(t) = (t-\lambda_1)^{l_1} \cdots (t-\lambda_h)^{l_h}$ un polinomio arbitrario con raíces $\lambda_1, \ldots, \lambda_h$. Es

$$p(A) = \begin{pmatrix} p(A_1) & & \\ & \ddots & \\ & & p(A_h) \end{pmatrix}.$$

Luego

$$\begin{aligned} p(A) = 0 &\Leftrightarrow p(A_i) = 0, \ \forall i = 1, \ldots, h \\ &\Leftrightarrow m_{A_i}(t) | \, p(t), \ \forall i = 1, \ldots, h \\ &\Leftrightarrow (t-\lambda_i)^{q_i} | (t-\lambda_1)^{l_1} \cdots (t-\lambda_h)^{l_h}, \ \forall i = 1, \ldots, h \\ &\Leftrightarrow q_i \leq l_i, \ \forall i = 1, \ldots, h. \end{aligned}$$

Dado que $m_A(t)$ es el polinomio de grado mínimo que verifica esta condición, deducimos que es $m_A(t) = (t-\lambda_1)^{q_1} \cdots (t-\lambda_h)^{q_h}$. □

Corolario 7.2.2. *Si el polinomio característico de un operador se escinde, entonces para cada valor propio λ, el exponente de $t-\lambda$ en la descomposición factorial del polinomio minimal coincide con el tamaño del mayor bloque de Jordan correspondiente a λ.* □

Ejemplo 7.2.3. Sea $A \in M_3(\mathbb{R})$ tal que $(A-5I)^2 = 0$, $A \neq 5I$. Esto implica $m_A(t) = (t-5)^2$ y por lo tanto $\chi_A(t) = -(t-5)^3$. Sabemos que la forma de Jordan de A tiene orden 3 y en la misma solo hay bloques correspondientes al valor propio 5, y de estos el más grande tiene orden 2. Luego la forma de Jordan es

$$J = \begin{pmatrix} 5 & 1 & 0 \\ 0 & 5 & 0 \\ 0 & 0 & 5 \end{pmatrix}.$$

Ejemplo 7.2.4. Sea $A \in M_4(\mathbb{R})$ tal que $(A-5I)^2 = 0$, $A \neq 5I$. Como antes deducimos $m_A(t) = (t-5)^2$, luego $\chi_A(t) = (t-5)^4$. Por la forma del polinomio minimal de A, sabemos que el tamaño del mayor bloque de la forma de Jordan de A es 2, luego las posibles formas de Jordan de A son

$$\begin{pmatrix} 5 & 1 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 5 & 1 \\ 0 & 0 & 0 & 5 \end{pmatrix}, \quad \begin{pmatrix} 5 & 1 & 0 & 0 \\ 0 & 5 & 0 & 0 \\ 0 & 0 & 5 & 0 \\ 0 & 0 & 0 & 5 \end{pmatrix}.$$

Para determinar la forma de Jordan necesitamos algún dato más. Por ejemplo, si sabemos el rango de $A-5I$, entonces la forma de Jordan de A es la primera si rango$(A-5I) = 2$ y es la segunda si rango$(A-5I) = 1$.

Ejemplo 7.2.5. Sea $T = L_A \in \mathcal{L}(\mathbb{R}^6)$, siendo

$$A = \begin{pmatrix} 3 & 0 & 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 1 & 0 & 0 \\ 0 & 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 8 & -9 \\ 0 & 0 & 0 & 0 & 4 & -4 \end{pmatrix}.$$

Notar que A es una matriz diagonal en bloques

$$A = \begin{pmatrix} B_1 & 0 \\ 0 & B_2 \end{pmatrix}, \quad B_1 = \begin{pmatrix} 3 & 0 & 1 & 0 \\ 0 & 3 & 0 & 1 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \quad B_2 = \begin{pmatrix} 8 & -9 \\ 4 & -4 \end{pmatrix}.$$

Luego $\chi_T(t) = \chi_A(t) = \chi_{B_1}(t)\chi_{B_2}(t) = (t-3)^4(t-2)^2$. Esto implica que si $\mathcal{B}$ es una base de Jordan correspondiente, entonces $[T]_\mathcal{B}$ es de la forma

$$[T]_\mathcal{B} = \begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}, \quad A_1 \in M_4(\mathbb{R}),\ A_2 \in M_2(\mathbb{R}),$$

siendo A_1 el bloque correspondiente al valor propio 3 y A_2 el correspondiente al valor propio 2. Calculando obtenemos $\operatorname{rango}(A-2I) = 5$ y $\operatorname{rango}(A-3I) = 4$. Luego las multiplicidades geométricas de los valores propios son $\mathrm{MG}(2) = 1$ y $\mathrm{MG}(3) = 2$. De $\mathrm{MG}(2) = 1$ deducimos que hay un solo bloque de Jordan correspondiente al valor propio 2, luego $A_2 = \left(\begin{smallmatrix} 2 & 1 \\ 0 & 2 \end{smallmatrix}\right)$. Al ser $\mathrm{MG}(3) = 2$, sabemos que hay dos bloques de Jordan correspondientes al valor propio 3, luego hay dos posibilidades para A_1 que son las siguientes

$$\begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \\ 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 3 & 1 & 0 & 0 \\ 0 & 3 & 1 & 0 \\ 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 3 \end{pmatrix}.$$

De acuerdo al corolario anterior, el polinomio minimal de T es, en el primer caso $(t-3)^2(t-2)^2$, y en el segundo $(t-3)^3(t-2)^2$. Se verifica fácilmente que vale $(B_1 - 3I)^2 = 0$, luego el minimal es el primero y por lo tanto la forma de Jordan es

$$[T]_\mathcal{B} = \begin{pmatrix} 3 & 1 & 0 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 2 \end{pmatrix}.$$

Esto implica que la base de Jordan $\mathcal{B}$ está formada por tres ciclos de longitud 2, dos de ellos corresponden al valor propio 3 y uno al valor propio 2. Es decir es del tipo

$$\mathcal{B} = \{(T-3\mathrm{Id})(u),\ u,\ (T-3\mathrm{Id})(v),\ v,\ (T-2\mathrm{Id})(w),\ w\}$$

siendo $u, v \in \operatorname{Ker}(T-3\mathrm{Id})^2 \setminus \operatorname{Ker}(T-3\mathrm{Id})$ y $w \in \operatorname{Ker}(T-2\mathrm{Id})^2 \setminus \operatorname{Ker}(T-2\mathrm{Id})$. Realizando los cálculos obtenemos

$$\operatorname{Ker}(T-3\mathrm{Id}) = [e_1, e_2],\ \operatorname{Ker}(T-3\mathrm{Id})^2 = [e_1, e_2, e_3, e_4],\ \operatorname{Ker}(T-2\mathrm{Id}) = [3e_5 + 2e_6],\ \operatorname{Ker}(T-2\mathrm{Id})^2 = [e_5, e_6],$$

siendo $\{e_1, \ldots, e_6\}$ la base canónica de $\mathbb{R}^6$. Si elegimos $u = e_3$, $v = e_4$ y $w = e_5$, entonces es $(T-3\mathrm{Id})(e_3) = e_1$, $(T-3\mathrm{Id})(e_4) = e_2$ y $(T-2\mathrm{Id})(e_5) = 6e_5 + 4e_6$, luego una base de Jordan es

$$\mathcal{B} = \{e_1,\ e_3,\ e_2,\ e_4,\ 6e_5 + 4e_6,\ e_5\}.$$

Capítulo 8

Apéndice

8.1. Suma directa

Sea V un espacio vectorial y W_1, W_2 dos subespacios de V. Definimos su *intersección* $W_1 \cap W_2$ y su *suma* $W_1 + W_2$, mediante

$$W_1 \cap W_2 := \{v \in V : \ v \in W_1 \text{ y } v \in W_2\}, \qquad W_1 + W_2 := \{w_1 + w_2 : \ w_1 \in W_1, \ w_2 \in W_2\}.$$

Proposición 8.1.1. *Si W_1, W_2 son subespacios de V, entonces $W_1 \cap W_2$ y $W_1 + W_2$ son también subespacios.*

Dem. Ejercicio. □

Observación 8.1.2. Dados dos subespacios W_1, W_2, su intersección $W_1 \cap W_2$ es el mayor subespacio contenido en W_1 y W_2, mientras que $W_1 + W_2$ es el menor subespacio que contiene a W_1 y W_2.

Ejemplos 8.1.3. Consideremos el espacio $\mathbb{R}^3$.

1. Si $W_1 = \{(x, y, z) : \ z = 0\}$ (plano Oxy) y $W_2 = \{(x, y, z) : \ y = 0\}$ (plano Oxz), entonces

$$W_1 \cap W_2 = \{(x, y, z) : \ y = z = 0\} \text{ (eje Ox)}, \qquad W_1 + W_2 = \mathbb{R}^3.$$

2. Si $W_1 = \{(x, y, z) : \ y = z = 0\}$ (eje Ox) y $W_2 = \{(x, y, z) : \ x = z = 0\}$ (eje Oy), entonces

$$W_1 \cap W_2 = \{(0, 0, 0)\}, \qquad W_1 + W_2 = \{(x, y, z) : \ z = 0\} \text{ (plano Oxy)}.$$

La siguiente proposición relaciona las dimensiones de $W_1 + W_2$ y $W_1 \cap W_2$, con las de W_1 y W_2.

Proposición 8.1.4. *Si W_1, W_2 son subespacios de dimensión finita de un espacio V, entonces vale*

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2).$$

Dem. Sea $\mathcal{B} = \{u_1, \ldots, u_n\}$ una base de $W_1 \cap W_2$. Como $\mathcal{B}$ es un subconjunto LI de W_1 y de W_2, entonces existen $v_1, \ldots, v_p \in W_1$ y $w_1, \ldots, w_q \in W_2$, tales que si

$$\mathcal{B}_1 = \{u_1, \ldots, u_n, v_1, \ldots, v_p\} \quad \text{y} \quad \mathcal{B}_2 = \{u_1, \ldots, u_n, w_1, \ldots, w_q\},$$

entonces $\mathcal{B}_1$ es base de W_1 y $\mathcal{B}_2$ es base de W_2. Probaremos que $\mathcal{C} = \{u_1, \ldots, u_n, v_1, \ldots, v_p, w_1, \ldots, w_q\}$ es base de $W_1 + W_2$. La prueba de que $\mathcal{C}$ es un conjunto generador de $W_1 + W_2$ es fácil y queda como ejercicio. Veamos que $\mathcal{C}$ es LI. Sean escalares a_i, b_j, c_k tales que

$$a_1 u_1 + \cdots + a_n u_n + b_1 v_1 + \cdots + b_p v_p + c_1 w_1 \cdots + c_q w_q = 0. \tag{8.1}$$

Luego

$$c_1w_1 \cdots + c_qw_q = -(a_1u_1 + \cdots + a_nu_n + b_1v_1 + \cdots + b_pv_p). \tag{8.2}$$

Como el lado izquierdo de (8.2) está en W_2 y el derecho en W_1, deducimos $c_1w_1 \cdots + c_qw_q \in W_1 \cap W_2$. Luego existen escalares d_i tales que

$$c_1w_1 \cdots + c_qw_q = d_1u_1 + \cdots + d_nu_n \quad \Rightarrow \quad c_1w_1 \cdots + c_qw_q + (-d_1)u_1 + \cdots + (-d_n)u_n = 0.$$

Dado que $\mathcal{B}_2$ es LI, deducimos $c_1 = \cdots = c_q = d_1 = \cdots = d_n = 0$. Sustituyendo estos valores en (8.1) obtenemos

$$a_1u_1 + \cdots + a_nu_n + b_1v_1 + \cdots + b_pv_p = 0.$$

Ahora usando que $\mathcal{B}_1$ es LI deducimos $a_1 = \cdots = a_n = b_1 = \cdots = b_p = 0$. Esto prueba que $\mathcal{C}$ es LI y por lo tanto es base de $W_1 + W_2$. Luego

$$\dim(W_1 + W_2) = n + p + q = (n + p) + (n + q) - n = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2). \quad \square$$

La proposición anterior se suele usar para calcular la dimensión de la suma de dos subespacios, ya que en general es más fácil determinar la intersección que la suma.

Ejemplo 8.1.5. Sea $V = \mathbb{R}^4$ y consideramos los subespacios

$$W_1 = \left\{(x, y, z, t) \in \mathbb{R}^4 : \ x = 0, \ y + z + t = 0\right\} \quad \text{y} \quad W_2 = \left\{(x, y, z, t) \in \mathbb{R}^4 : \ t = 0\right\}.$$

Es fácil de probar que es $\dim W_1 = 2$ y $\dim W_2 = 3$. Además

$$W_1 \cap W_2 = \left\{(x, y, z, t) \in \mathbb{R}^4 : \ x = 0, \ y + z + t = 0, \ t = 0\right\} = \left\{(x, y, z, t) \in \mathbb{R}^4 : \ x = t = 0, \ y + z = 0\right\},$$

luego $\dim(W_1 \cap W_2) = 1$ y por lo tanto

$$\dim(W_1 + W_2) = \dim W_1 + \dim W_2 - \dim(W_1 \cap W_2) = 2 + 3 - 1 = 4 = \dim \mathbb{R}^4.$$

Luego $W_1 + W_2 = \mathbb{R}^4$.

Definición 8.1.6. Sean W_1, W_2 dos subespacios de un espacio V. Decimos que un subespacio U de V es *suma directa* de W_1 y W_2 y escribimos $U = W_1 \oplus W_2$ si se cumple $U = W_1 + W_2$ y $W_1 \cap W_2 = \{0\}$.

Ejemplo 8.1.7. Si consideramos los ejemplos 8.1.3, deducimos del primero que el espacio $\mathbb{R}^3$ es suma del plano Oxy con el plano Oxz, pero esta suma no es directa, y del segundo que el plano Oxy es suma directa de la recta Ox y la recta Oy. En el ejercicio 8.1.5 vemos también que $\mathbb{R}^4$ es suma de W_1 y W_2, pero esta suma no es directa.

Proposición 8.1.8. *Si W_1, W_2 son dos subespacios de V, entonces $V = W_1 \oplus W_2$ si y solo si todo vector de V se escribe en forma única como suma de un vector de W_1 con uno de W_2.*

Dem. Ver la prueba de la proposición 8.1.14. $\square$

De la proposición 8.1.4 y de su demostración, se deduce inmediatamente el siguiente resultado.

Proposición 8.1.9. *Sea V un espacio de dimensión finita y W_1, W_2 dos subespacios de V, tales que $V = W_1 \oplus W_2$. Valen las siguientes afirmaciones.*

1. *Si $\mathcal{B}_1$ y $\mathcal{B}_2$ son bases respectivas de W_1 y W_2, entonces su unión $\mathcal{B}_1 \cup \mathcal{B}_2$ es una base de V.*

2. *$dimV = \dim W_1 + \dim W_2$.* $\square$

Corolario 8.1.10. *Sean W_1, W_2 dos subespacios de un espacio de dimensión finita V tales que $W_1 \cap W_2 = \{0\}$. Si $\dim V = \dim W_1 + \dim W_2$, entonces $V = W_1 \oplus W_2$.*

Dem. La proposición 8.1.9 implica $\dim(W_1 \oplus W_2) = \dim W_1 + \dim W_2 = \dim V$, luego $W_1 \oplus W_2 = V$. $\square$

Este último corolario nos da un método fácil para probar que un espacio es suma directa de dos subespacios dados.

Ejemplo 8.1.11. Consideremos el espacio $\mathbb{R}^3$ y los subespacios

$$W_1 = \{(x, y, z) : \ 2x + y + z = 0\} \qquad \text{y} \qquad W_2 = \{(x, y, z) : \ y = z = 0\}.$$

Si $(x, y, z) \in W_1 \cap W_2$, entonces vale

$$\begin{cases} 2x + y + z = 0 \\ \quad\;\; y = z = 0 \end{cases} \quad \Rightarrow \quad x = y = z = 0 \quad \Rightarrow \quad (x, y, z) = (0, 0, 0).$$

Luego $W_1 \cap W_2 = \{(0, 0, 0)\}$. Por otro lado W_1 es un plano y W_2 una recta, así que es $\dim W_1 = 2$ y $\dim W_2 = 1$ y por lo tanto $\dim W_1 + \dim W_2 = 2 + 1 = 3 = \dim \mathbb{R}^3$; luego $\mathbb{R}^3 = W_1 \oplus W_2$.

Observación 8.1.12. Dados dos subespacios W_1, W_2 de V, no alcanza con que valga $\dim V = \dim W_1 + \dim W_2$ para que la suma sea directa, hay que probar que también vale $W_1 \cap W_2 = \{0\}$. Por ejemplo, si el espacio es $\mathbb{R}^3$ y consideramos los subespacios

$$W_1 = \{(x, y, z) : \ z = 0\} \qquad \text{y} \qquad W_2 = \{(x, y, z) : \ y = z = 0\},$$

entonces W_1 es un plano y W_2 una recta, y por lo tanto $\dim W_1 + \dim W_2 = 2 + 1 = 3 = \dim \mathbb{R}^3$. Pero W_2 está contenido en W_1, luego $W_1 + W_2 = W_1 \subsetneq \mathbb{R}^3$ y por lo tanto nunca puede ser $W_1 \oplus W_2 = \mathbb{R}^3$.

Generalización. A continuación veremos que todos los resultados anteriores para la suma e intersección de dos subespacios se pueden generalizar a una cantidad finita arbitraria de subespacios.

Sea V un espacio vectorial y $W_1, \ldots, W_m$ una cantidad finita de subespacios de V. Definimos su *intersección* $\bigcap_{i=1}^m W_i$ y su *suma* $\sum_{i=1}^m W_i$, mediante

$$\bigcap_{i=1}^m W_i := \{v \in V : \ v \in W_i, \ \forall i = 1, \ldots, n\}, \qquad \sum_{i=1}^m W_i := \left\{ \sum_{i=1}^m w_i : \ w_i \in W_i, \ \forall i = 1, \ldots, n \right\}.$$

La prueba de la siguiente proposición es simple y queda como ejercicio.

Proposición 8.1.13. *Si $W_1, \ldots, W_m$ son subespacios de V, entonces $\bigcap_{i=1}^m W_i$ y $\sum_{i=1}^m W_i$ son también subespacios.* $\square$

La definición general de suma directa es un poco más delicada que para el caso de dos subespacios. Para definirla introducimos previamente el siguiente concepto.

Decimos que una familia[1] $\{W_1, \ldots, W_m\}$ de subespacios de V es *independiente* si verifica la siguiente condición.

$$\text{Si } w_i \in W_i, \ \forall i = 1, \ldots, m \text{ y } \sum_{i=1}^m w_i = 0, \text{ entonces } w_1 = \cdots = w_m = 0.$$

En lo que sigue escribimos

$$\sum_{j \neq i} W_j := W_1 + \cdots + W_{i-1} + W_{i+1} + \cdots + W_m, \quad \forall i = 1, \ldots, m.$$

[1] En este texto llamaremos *familia* a un conjunto de conjuntos.

Proposición 8.1.14. *Sea $\{W_1, \dots, W_m\}$ una familia de subespacios de V. Entonces las siguientes condiciones son equivalentes:*

1. *La familia $\{W_1, \dots, W_m\}$ es independiente.*
2. *Para cada $i = 1, \dots, m$, vale $W_i \cap \sum_{j \neq i} W_j = \{0\}$.*
3. *Si $W = \sum_{i=1}^m W_i$, entonces para todo $w \in W$, existen únicos $w_i \in W_i$, $i = 1, \dots, m$, tales que $w = \sum_{i=1}^m w_i$.*

Dem. $(1 \Rightarrow 2)$. Supongamos que la familia es independiente. Sea $i \in \{1, \dots, m\}$. Si $w \in W_i \cap \sum_{j \neq i} W_j$, entonces $w \in W_i$ y existen $w_j \in W_j$, para todo $j \neq i$ tales que

$$w = w_1 + \cdots + w_{i-1} + w_{i+1} + \cdots + w_m \quad \Rightarrow \quad w_1 + \cdots + w_{i-1} + (-w) + w_{i+1} + \cdots + w_m = 0.$$

Como la familia es independiente, entonces todos los sumandos de la suma anterior son nulos, luego $w = 0$.

$(2 \Rightarrow 1)$. Supongamos ahora que vale la segunda condición. Sean $w_i \in W_i$, $i = 1, \dots, m$, tales que $\sum_{i=1}^m w_i = 0$. Para cada $i \in \{1, \dots, m\}$, es

$$-w_i = w_1 + \cdots + w_{i-1} + w_{i+1} + \cdots + w_m.$$

Como $-w_i \in W_i$ y $w_1 + \cdots + w_{i-1} + w_{i+1} + \cdots + w_m \in \sum_{j \neq i} W_j$, deducimos que es $-w_i \in W_i \cap \sum_{j \neq i} W_j = \{0\}$. Luego $w_i = 0$, para todo $i = 1, \dots, m$.

$(3 \Rightarrow 1)$. Supongamos vale la tercer afirmación. Sean $w_i \in W_i$, $i = 1, \dots, m$, tales que $\sum_{i=1}^m w_i = 0$. Como $0 \in W_i$ para todo $i = 1, \dots, m$ y $\sum_{i=1}^m 0 = 0$, entonces la unicidad implica que es $w_i = 0$, para todo $i = 1, \dots, m$. Luego la familia $\{W_1, \dots, W_m\}$ es independiente.

$(1 \Rightarrow 3)$. Supongamos ahora que la familia $\{W_1, \dots, W_m\}$ es independiente. Si un elemento $w \in W$ se escribe de dos formas $w = \sum_{i=1}^m w_i$ y $w = \sum_{i=1}^m w_i'$, con $w_i, w_i' \in W_i$, para todo $i = 1, \dots, m$, entonces

$$\sum_{i=1}^m (w_i - w_i') = 0, \text{ siendo } w_i - w_i' \in W_i, \ \forall i = 1, \dots, m.$$

Luego $w_i - w_i' = 0$ y por lo tanto $w_i = w_i'$, para todo $i = 1, \dots, m$. □

Observación 8.1.15. Una familia con solo dos subespacios W_1 y W_2 es independiente si y solo si $W_1 \cap W_2 = \{0\}$, que es la condición que pedimos para que la suma $W_1 + W_2$ sea directa. Una familia de tres subespacios $\{W_1, W_2, W_3\}$ es independiente si y solo si W_1, W_2, W_3 verifican

$$W_1 \cap (W_2 + W_3) = \{0\}, \quad W_2 \cap (W_1 + W_3) = \{0\}, \quad W_3 \cap (W_1 + W_2) = \{0\}.$$

Notar que las condiciones $W_1 \cap W_2 = W_1 \cap W_3 = W_2 \cap W_3 = \{0\}$ no implican las condiciones anteriores. Por ejemplo, si consideramos $V = \Bbbk^3$ y

$$W_1 = \{(x, 0, 0) : \ x \in \Bbbk\}, \quad W_2 = \{(0, y, 0) : \ y \in \Bbbk\}, \quad W_3 = \{(x, x, 0) : \ x \in \Bbbk\},$$

entonces es fácil de probar que las intersecciones dos a dos son triviales, pero $W_1 + W_2 = \{(x, y, 0); \ x, y \in \Bbbk\}$, luego $W_3 \subset W_1 + W_2$ y por lo tanto $W_3 \cap (W_1 + W_2) = W_3 \neq \{0\}$.

Definición 8.1.16. Decimos que un subespacio W de V es *suma directa* de una familia de subespacios $\{W_1, \dots, W_m\}$ y escribimos $W = \bigoplus_{i=1}^m W_i$, si $W = \sum_{i=1}^m W_i$ y la familia $\{W_1, \dots, W_m\}$ es independiente.

Observación 8.1.17. Lo que distingue la suma directa de la suma, y la hace más interesante, es la unicidad de la tercer afirmación de la proposición 8.1.14. Es similar a la diferencia entre una base y un conjunto generador.

Corolario 8.1.18. *Sea $\{W_1, \ldots, W_m\}$ una familia de subespacios de un espacio V. Entonces $V = \bigoplus_{i=1}^m W_i$ si y solo si para todo $v \in V$, existen únicos $w_i \in W_i$, $i = 1, \ldots, m$, tales que $v = \sum_{i=1}^m w_i$.* □

Ejemplo 8.1.19. Es fácil probar usando el corolario 8.1.18 que vale $M_n = W_1 \oplus W_2 \oplus W_3$, siendo W_1 es el espacio de las matrices estrictamente triangulares inferiores, W_2 el de las matrices estrictamente triangulares superiores y W_3 el de las matrices diagonales, es decir

$$W_1 = \{(a_{ij}) : \ a_{ij} = 0 \text{ si } i \leq j\}, \quad W_2 = \{(a_{ij}) : \ a_{ij} = 0 \text{ si } i \geq j\}, \quad W_1 = \{(a_{ij}) : \ a_{ij} = 0 \text{ si } i \neq j\}.$$

Diagramáticamente, la descomposición $M_n = W_1 \oplus W_2 \oplus W_3$ consiste en escribir $\left(\begin{smallmatrix} * & * \\ * & * \end{smallmatrix}\right) = \left(\begin{smallmatrix} 0 & 0 \\ * & 0 \end{smallmatrix}\right) + \left(\begin{smallmatrix} 0 & * \\ 0 & 0 \end{smallmatrix}\right) + \left(\begin{smallmatrix} * & 0 \\ 0 & * \end{smallmatrix}\right)$.

El siguiente resultado es fácil de probar.

Proposición 8.1.20. *Sean $W, W_1, \ldots, W_m$ subespacios de V tales que $W = \sum_{i=1}^m W_i$. Si $\mathcal{G}_i$ es un conjunto generador de W_i, para todo $i = 1, \ldots, m$, entonces su unión $\bigcup_{i=1}^m \mathcal{G}_i$ es un conjunto generador de W.* □

A continuación veremos cómo se relaciona la suma directa con la dimensión.

Proposición 8.1.21. *Sean $W, W_1, \ldots, W_m$ subespacios de dimensión finita de un espacio V, tales que $W = \bigoplus_{i=1}^m W_i$. Valen las siguientes afirmaciones.*

1. *Si $\mathcal{B}_i$ es una base de W_i, para todo $i = 1, \ldots, m$, entonces $\mathcal{B} = \bigcup_{i=1}^m \mathcal{B}_i$ es una base de W.*
2. $\dim W = \sum_{i=1}^m \dim W_i$.

Dem. Por la proposición anterior sabemos que $\mathcal{B}$ es un conjunto generador de W, lo que resta probar es que es LI. Para cada $i = 1, \ldots, m$, sea $\mathcal{B}_i = \{w_1^i, \ldots, w_{l_i}^i\}$. Supongamos que tenemos escalares a_j^i tales que

$$a_1^1 w_1^1 + \cdots + a_{l_1}^1 w_{l_1}^1 + \cdots + a_1^m w_1^m + \cdots + a_{l_m}^m w_{l_m}^m = 0.$$

Como $a_1^i w_1^i + \cdots + a_{l_i}^i w_{l_i}^i \in W_i$ (para todo i) y la familia $\{W_1, \ldots, W_m\}$ es independiente, deducimos que es

$$a_1^i w_1^i + \cdots + a_{l_i}^i w_{l_i}^i = 0, \quad \forall i = 1, \ldots, m.$$

Luego como cada $\mathcal{B}_i$ es LI, deducimos que es $a_j^i = 0$, para todo i, j y por lo tanto $\mathcal{B}$ es LI. La segunda afirmación se deduce inmediatamente de la primera. □

Corolario 8.1.22. *Sean $W_1, \ldots, W_m$ subespacios de un espacio de dimensión finita V. Si $\{W_1, \ldots, W_m\}$ es una familia independiente y vale $\dim V = \sum_{i=1}^m \dim W_i$, entonces $V = \bigoplus_{i=1}^m W_i$.* □

A continuación definiremos las proyecciones y estudiaremos su relación con las sumas directas. Las proyecciones son los elementos básicos para definir la descomposición espectral, que veremos más adelante.

Sean U y W subespacios de V tales que $V = U \oplus W$. Definimos una función $T : V \to V$ mediante $T(v) = u$, si $v = u + w$, con $u \in U$, $w \in W$, para todo $v \in V$. Es un ejercicio el probar que T es una transformación lineal. La función T se llama la *proyección sobre U en la dirección de W*.

Observar que si $v = u + w$, con $u \in U$, $w \in W$, entonces $T(v) = u = u + 0$, con $u \in U$, $0 \in W$; luego $T(T(v)) = u = T(v)$, para todo $v \in V$, es decir $T^2 = T$, siendo $T^2 := T \circ T$.

Una *proyección* es un operador $T \in \mathcal{L}(V)$ que verifica $T^2 = T$. El comentario anterior muestra que para toda descomposición $V = U \oplus W$, la proyección sobre U en la dirección de W es una proyección. La proposición siguiente muestra que también vale el recíproco.

Proposición 8.1.23. *Sea $T \in \mathcal{L}(V)$ tal que $T^2 = T$, entonces $V = \operatorname{Im}(T) \oplus \operatorname{Ker}(T)$ y T es la proyección sobre $\operatorname{Im}(T)$ en la dirección de $\operatorname{Ker}(T)$.*

Dem. Sea $v \in V$. Observar que la condición $T^2 = T$ implica $v - T(v) \in \operatorname{Ker}(T)$. Luego escribiendo $v = T(v) + v - T(v)$, deducimos $V = \operatorname{Im}(T) + \operatorname{Ker}(T)$. Por otro lado, si $v \in \operatorname{Ker}(T) \cap \operatorname{Im}(T)$, entonces $T(v) = 0$ y existe $u \in V$ tal que $v = T(u)$. Luego, $0 = T(v) = T\big(T(u)\big) = T^2(u) = T(u) = v$. Esto prueba que vale $\operatorname{Ker}(T) \cap \operatorname{Im}(T) = \{0\}$ y por lo tanto $V = \operatorname{Im}(T) \oplus \operatorname{Ker}(T)$.
Sea $v \in V = \operatorname{Im}(T) \oplus \operatorname{Ker}(T)$. Por lo que vimos anteriormente es $v = T(v) + v - T(v)$, con $T(v) \in \operatorname{Im}(T)$ y $v - T(v) \in \operatorname{Ker}(T)$. Luego la proyección sobre $\operatorname{Im}(T)$ en la dirección de $\operatorname{Ker}(T)$ de v es $T(v)$. $\square$

El siguiente resultado generaliza lo anterior a una cantidad finita arbitraria de subespacios. El mismo no se necesita para estas notas, pero lo incluimos por completitud.

Proposición 8.1.24. *Consideremos los siguientes datos.*

(A) *$W_1, \dots, W_k$ son subespacios de V tales que $V = \bigoplus_{i=1}^k W_i$.*

(B) *$P_1, \dots, P_k \in \mathcal{L}(V)$ son operadores que verifican las siguientes condiciones.*

1. *$P_i^2 = P_i$, para todo $i = 1, \dots, k$.*
2. *$P_i \circ P_j = 0$, para todo $i \neq j$, con $i, j = 1, \dots, k$.*
3. *$\operatorname{Id} = P_1 + \dots + P_k$.*

Entonces **(A)** *y* **(B)** *son equivalentes.*

Dem. Sean $W_1, \dots, W_k$ como en **(A)** y definimos $P_1, \dots, P_k : V \to V$ por $P_i(v) = w_i$, si $v = \sum_{j=1}^k w_j$, con $w_j \in W_j$, para todo $j = 1, \dots, k$ y todo $v \in V$. Notar que para cada $i \in \{1, \dots, n\}$ es

$$V = W_i \oplus \left(\bigoplus_{j\neq i} W_j\right) \quad \text{y} \quad v = \sum_{j=1}^k w_j = w_i + \sum_{j\neq i} w_j, \quad \text{con } w_l \in W_l,\ \forall l = 1, \dots, k.$$

Luego $P_i : V \to V$ es la proyección sobre W_i en la dirección de $\bigoplus_{j\neq i} W_j$; esto prueba que P_i es una proyección ($P_i \in \mathcal{L}(V)$ y $P_i^2 = P_i$), para todo $i = 1, \dots, k$. Observar que por como definimos los operadores P_i, es

$$v = \sum_{i=1}^k P_i(v),\ \forall v \in V \quad \Rightarrow \quad \operatorname{Id} = P_1 + \dots + P_k.$$

Finalmente, si $v \in V$, entonces $P_i(v) \in W_i$, por lo tanto su descomposición correspondiente a $V = \bigoplus_{i=1}^k W_i$ es $P_i(v) = 0 + \dots + P_i(v) + \dots + 0$, es decir, la componente de $P_i(v)$ en el subespacio W_j es 0 si $j \neq i$. Esto implica que vale $P_j\big(P_i(v)\big) = 0$, para todo $j \neq i$ y $v \in V$; luego $P_j \circ P_i = 0$, para todo $j \neq i$.

Supongamos ahora que $P_1, \dots, P_k \in \mathcal{L}(V)$ verifican las condiciones de **(B)** y definimos $W_i = \operatorname{Im}(P_i)$, para todo i. De $\operatorname{Id} = \sum_{i=1}^k P_i$ deducimos $v = \sum_{i=1}^k P_i(v) \in \sum_{i=1}^k W_i$, para todo $v \in V$. Luego $V = \sum_{i=1}^k W_i$. Solo resta probar que $W_1, \dots, W_k$ son independientes. Sean $w_i \in W_i$, $i = 1, \dots, k$, tales que $w_1 + \dots + w_k = 0$. Al ser $w_i \in W_i = \operatorname{Im}(P_i)$, entonces existe $v_i \in V$ tal que $w_i = P_i(v_i)$, para todo i. Luego para cada i, es

$$0 = P_i\left(\sum_{j=1}^k w_j\right) = P_i\left(\sum_{j=1}^k P_j(v_j)\right) = \sum_{j=1}^k P_i(P_j(v_j)) = \sum_{j=1}^k (P_i \circ P_j)(v_j) = P_i(v_i) = w_i.$$

Por lo tanto es $w_1 = \dots = w_k = 0$. Esto muestra que $W_1, \dots, W_k$ son independientes y concluye la prueba de que es $V = \bigoplus_{i=1}^k W_i$. $\square$

Observación 8.1.25. En las hipótesis de la observación anterior. Notar que si partimos de $W_1, \dots, W_k$ y les asociamos las correspondientes proyecciones $P_1, \dots, P_k$, entonces $\operatorname{Im}(P_i) = W_i$, para todo $i = 1, \dots, k$. Recíprocamente, si partimos de proyecciones $P_1, \dots, P_k$ y definimos $W_i = \operatorname{Im}(P_i)$, para todo $i = 1, \dots, k$, entonces las proyecciones asociadas a la descomposición $V = \bigoplus_{i=1}^{k} W_i$, son exactamente $P_1, \dots, P_k$. Es decir, las correspondencias que vimos entre proyecciones y sumas directas, son inversas una de la otra.

8.2. Matrices en bloques

En esta sección estudiaremos las propiedades de matrices cuyos elementos están agrupados en submatrices (bloques) de la matriz original. Esto tiene varias aplicaciones, en particular se usa para el estudio del polinomio minimal y de la forma de Jordan. En lo que sigue $\Bbbk$ es un cuerpo arbitrario.

Empezamos considerando un caso particular. Dadas cuatro matrices

$$A = (a_{ij}) \in M_{p\times p}(\Bbbk), \quad B = (b_{ij}) \in M_{p\times q}(\Bbbk), \quad C = (c_{ij}) \in M_{q\times p}(\Bbbk), \quad D = (d_{ij}) \in M_{q\times q}(\Bbbk),$$

las podemos combinar para obtener una matriz $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right) \in M_{n\times n}(\Bbbk)$, con $n = p + q$, definida por

$$\begin{pmatrix} A & B \\ C & D \end{pmatrix} := \begin{pmatrix} a_{11} & \cdots & a_{1p} & b_{11} & \cdots & b_{1q} \\ \vdots & & \vdots & \vdots & & \vdots \\ a_{p1} & \cdots & a_{pp} & b_{p1} & \cdots & b_{pq} \\ c_{11} & \cdots & c_{1p} & d_{11} & \cdots & d_{1q} \\ \vdots & & \vdots & \vdots & & \vdots \\ c_{q1} & \cdots & c_{qp} & d_{q1} & \cdots & d_{qq} \end{pmatrix}.$$

Notar que lo que hicimos para cuatro matrices, se puede aplicar a una cantidad arbitraria de matrices que tengan tamaños adecuados para poder combinarse correctamente. Una matriz expresada de esa forma diremos que es una *matriz en bloques*. Es claro que toda matriz se puede expresar como matriz en bloques (en general de muchas formas). Lo interesante de esto es cuando esos bloques tienen alguna forma especial. Para evitar confusiones, cuando escribamos una matriz en bloques, a las submatrices que la componen las escribiremos siempre con mayúsculas. Así $\left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ representa una matriz en bloques y no a una matriz 2×2 (a menos que todos los bloques sean de tamaño 1×1).

Las matrices que más nos interesan son las matrices diagonales en bloques, es decir, las de la forma

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_h \end{pmatrix},$$

en que cada A_i es una matriz de tamaño $n_i \times n_i$, $i = 1, \dots, h$ y las otras entradas son nulas. Es un ejercicio el probar que valen las fórmulas siguientes

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_h \end{pmatrix} + c \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_h \end{pmatrix} = \begin{pmatrix} A_1 + cB_1 & 0 & \cdots & 0 \\ 0 & A_2 + cB_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_h + cB_h \end{pmatrix},$$

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_h \end{pmatrix} \begin{pmatrix} B_1 & 0 & \cdots & 0 \\ 0 & B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & B_h \end{pmatrix} = \begin{pmatrix} A_1B_1 & 0 & \cdots & 0 \\ 0 & A_2B_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_hB_h \end{pmatrix}$$

En particular de la fórmula para el producto se obtiene la fórmula para las potencias

$$\begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_h \end{pmatrix}^k = \begin{pmatrix} A_1^k & 0 & \cdots & 0 \\ 0 & A_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_h^k \end{pmatrix}, \quad \forall k \in \mathbb{N}.$$

Notar la analogía de estas fórmulas con las correspondientes en matrices diagonales. A continuación veremos que el cálculo de determinantes se simplifica para ciertos tipos de matrices en bloques.

Proposición 8.2.1. *Si* $X = \left(\begin{smallmatrix} A & B \\ 0 & D \end{smallmatrix}\right)$*, siendo* A *y* D *matrices cuadradas, entonces* $\det X = \det A \det D$.

Dem. Supongamos $A = (a_{ij}) \in M_{p\times p}(\Bbbk)$. Si $p = 1$, entonces $A = (a_{11})$. Luego desarrollando $\det X$ por la primer comlumna obtenemos

$$\det X = a_{11} \det D = \det A \det D.$$

El caso general se deduce por inducción en p, desarrollando $\det X$ por su primer columna. Como este cálculo es muy engorroso, lo mostraremos solo para $p = 3$, que es ilustrativo del paso inductivo, dejando la prueba general como ejercicio. Supongamos entonces que tenemos probado el resultado para $p = 2$ y consideramos el caso $p = 3$. Desarrollando $\det X$ por la primer columna obtenemos

$$\begin{aligned}
&\begin{vmatrix} a_{11} & a_{12} & a_{13} & b_{11} & \cdots & b_{1q} \\ a_{21} & a_{22} & a_{23} & b_{21} & \cdots & b_{2q} \\ a_{31} & a_{32} & a_{33} & b_{31} & \cdots & b_{3q} \\ 0 & 0 & 0 & d_{11} & \cdots & d_{1q} \\ \vdots & \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & d_{q1} & \cdots & d_{qq} \end{vmatrix} = \\
&= a_{11} \begin{vmatrix} a_{22} & a_{23} & b_{21} & \cdots & b_{2q} \\ a_{32} & a_{33} & b_{31} & \cdots & b_{3q} \\ 0 & 0 & d_{11} & \cdots & d_{1q} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & d_{q1} & \cdots & d_{qq} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} & b_{11} & \cdots & b_{1q} \\ a_{32} & a_{33} & b_{31} & \cdots & b_{3q} \\ 0 & 0 & d_{11} & \cdots & d_{1q} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & d_{q1} & \cdots & d_{qq} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} & b_{11} & \cdots & b_{1q} \\ a_{22} & a_{23} & b_{21} & \cdots & b_{2q} \\ 0 & 0 & d_{11} & \cdots & d_{1q} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & d_{q1} & \cdots & d_{qq} \end{vmatrix} \\
&= a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} |D| - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} |D| + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} |D| \quad \text{(usando la hipótesis inductiva)} \\
&= \left(a_{11} \begin{vmatrix} a_{22} & a_{23} \\ a_{32} & a_{33} \end{vmatrix} - a_{21} \begin{vmatrix} a_{12} & a_{13} \\ a_{32} & a_{33} \end{vmatrix} + a_{31} \begin{vmatrix} a_{12} & a_{13} \\ a_{22} & a_{23} \end{vmatrix} \right) |D| = |A||D|. \quad \square
\end{aligned}$$

Proposición 8.2.2. *Si* X *es una matriz triangular superior en bloques del tipo*

$$X = \begin{pmatrix} A_{11} & A_{12} & \cdots & A_{1h} \\ 0 & A_{22} & \cdots & A_{2h} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_{hh} \end{pmatrix},$$

en que $A_{11}, \ldots, A_{hh}$ *son matrices cuadradas, entonces* $\det X = \prod_{i=1}^h \det A_{ii}$.

Dem. El caso $h = 2$ es la proposición anterior. El caso general se prueba por inducción en h, observando que podemos escribir $X = \left(\begin{smallmatrix} A_{11} & B \\ 0 & T \end{smallmatrix}\right)$, siendo $T = \begin{pmatrix} A_{22} & \cdots & A_{2h} \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_{hh} \end{pmatrix}$ y $B = \begin{pmatrix} A_{12} & \cdots & A_{1h} \end{pmatrix}$. $\square$

Corolario 8.2.3. *Si X es una matriz diagonal en bloques del tipo* $X = \begin{pmatrix} A_1 & 0 & \cdots & 0 \\ 0 & A_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & A_h \end{pmatrix}$, *en que* $A_1, \ldots, A_h$ *son matrices cuadradas, entonces* $\det X = \prod_{i=1}^{h} \det A_i$. □

8.3. Matrices elementales

En esta sección $\Bbbk$ es un cuerpo arbitrario.

Definición 8.3.1. Sea $A \in M_{m\times n}$. Se llaman *operaciones elementales* en A a las siguientes:

1. *Operación de tipo I.* Intercambiar dos filas o columnas de A.
2. *Operación de tipo II.* Multiplicar una fila o columna de A por una constante no nula.
3. *Operación de tipo III.* Sumarle a una fila o columna un múltiplo de otra fila o columna, respectivamente.

Definición 8.3.2. Se llaman *matrices elementales* de tipo I, II o III a las matrices que se obtienen aplicando las operaciones elementales correspondientes a la matriz identidad.

Es decir que las matrices elementales son las siguientes:

$$\text{Tipo I: } \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}; \quad \text{Tipo II: } \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & p & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \quad p \neq 0;$$

$$\text{Tipo III: } \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & q & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix} \quad \text{o} \quad \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & \vdots & \ddots & & & \\ & & q & \cdots & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \quad q \in \Bbbk.$$

Proposición 8.3.3. *Realizar una operación elemental de tipo I, II o III en las filas (columnas) de una matriz A, equivale a multiplicar a A por la izquierda (derecha) con una matriz elemental del mismo tipo.* □

Ejemplo 8.3.4. Sea $A = \begin{pmatrix} 1 & 2 & 3 \\ 5 & 6 & 7 \\ 9 & 1 & 2 \\ 4 & 5 & 6 \end{pmatrix}$. Consideremos $E_1 = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$ matriz elemental de tipo I.

Multiplicando obtenemos $E_1 \cdot A = \begin{pmatrix} 1 & 2 & 3 \\ 9 & 1 & 2 \\ 5 & 6 & 7 \\ 4 & 5 & 6 \end{pmatrix}$. Luego multiplicar la matriz A por la izquierda con la matriz

E_1 equivale a intercambiar las filas 2 y 3 de A. Por otro lado considerando $E_2 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 0 & 1 \end{pmatrix}$ matriz elemental de tipo III, obtenemos $A \cdot E_2 = \begin{pmatrix} 1 & 2 & 1 \\ 5 & 6 & 1 \\ 9 & 1 & 1 \\ 4 & 5 & 1 \end{pmatrix}$. Luego multiplicar la matriz A por la derecha con la matriz E_2 equivale a restarle la columna 2 a la columna 3.

Corolario 8.3.5. *Las matrices elementales de tipo I, II y III son invertibles y sus inversas son también matrices elementales del mismo tipo. Explícitamente*

Matrices de tipo I:

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 0 & \cdots & 1 & & \\ & & \vdots & \ddots & \vdots & & \\ & & 1 & \cdots & 0 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}.$$

Matrices de tipo II:

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & p & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & & & & \\ & & & p^{-1} & & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \quad p \neq 0.$$

Matrices de tipo III:

$$\begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & q & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}^{-1} = \begin{pmatrix} 1 & & & & & & \\ & \ddots & & & & & \\ & & 1 & \cdots & -q & & \\ & & & \ddots & \vdots & & \\ & & & & 1 & & \\ & & & & & \ddots & \\ & & & & & & 1 \end{pmatrix}, \quad q \in \Bbbk.$$

La fórmula para la inversa de la otra matriz de tipo III es análoga. □

Observación 8.3.6.

1. La traspuesta de una matriz elemental, es una matriz elemental del mismo tipo.
2. Si realizar una operación elemental en las columnas de A corresponde a multiplicar por la derecha a A con una cierta matriz elemental E, entonces realizar la misma operación elemental en las filas de A corresponde a multiplicar por la izquierda a A con E^t.