Versión en revisión

Apuntes de Modelización Discreta

(IMA311)

Versión del 24 de julio de 2024

Dr. Víctor Andrés Osores Escalona
Departamento de Matemática, Física y Estadística
Universidad Católica del Maule, Chile

Índice general

Prefacio

Este apunte es el resultado de una recopilación de recursos disponibles en la red. A lo largo de su proceso de escritura, he consultado una variedad de fuentes en línea, desde documentos académicos hasta tutoriales y ejemplos prácticos. Quiero enfatizar que la base de conocimiento en la que se sustenta este libro es el resultado de la generosidad de la comunidad en línea y de aquellos que han compartido su experiencia y conocimientos libremente.

Agradezco profundamente a todas las personas que han contribuido a la creación de este apunte a través de su valiosa información disponible en la red. Sin su generosidad y el espíritu colaborativo de la era digital, este documento no habría sido posible.

Es importante destacar que, a lo largo del libro, se citarán y referenciarán las fuentes relevantes de manera apropiada. Mi objetivo es proporcionar a los lectores una compilación coherente y útil de información sobre ecuaciones diferenciales ordinarias, al tiempo que se reconoce y respeta el trabajo de quienes la han compartido en línea.

CAPÍTULO 1

Preliminares

1.1 Sistemas de ecuaciones lineales

Definición 1.1 **(Sistema lineal de ecuaciones)** *Sea $\mathbb{K}$ el cuerpo $\mathbb{R}$ o $\mathbb{C}$. Un sistema lineal de m ecuaciones y n incógnitas en $\mathbb{K}$ es un conjunto de m ecuaciones lineales en que cada una tiene a lo más n incógnitas, esto es,*

$$\left.\begin{array}{rcl} a_{11}x_1 + a_{12}x_2 + \cdots + a_{1n}x_n & = & b_1 \\ a_{21}x_1 + a_{22}x_2 + \cdots + a_{2n}x_n & = & b_2 \\ \vdots \qquad \vdots \qquad \vdots & & \vdots \\ a_{m1}x_1 + a_{m2}x_2 + \cdots + a_{mn}x_n & = & b_m \end{array}\right\} \tag{1.1.1}$$

donde, para $i \in \{1, \ldots, m\}$ y $j \in \{1, \ldots, n\}$, $a_{ij} \in \mathbb{K}$ son los coeficientes del sistema, $b_i \in \mathbb{K}$ son los términos independientes del sistema y $x_1, \ldots, x_n$ son las incógnitas del sistema.
Si $b_1 = b_2 = \cdots = b_m = 0$, entonces el sistema se dice homogéneo, en caso contrario se dice no homogéneo.

El sistema (1.1.1) se puede escribir de la siguiente forma,

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

lo que da por resultado la ecuación matricial,

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

Definición 1.2 *Decimos que la n-upla $(y_1, \ldots, y_n)^t \in \mathbb{K}^n (= M_{n\times 1}(\mathbb{K}))$ es una solución del sistema* (1.1.1)*, si al reemplazar ordenadamente cada x_i por y_i con $i \in \{1, \ldots, n\}$, se satisfacen simultáneamente las m igualdades del sistema* (1.1.1)*. Llamaremos conjunto solución del sistema* (1.1.1)*, al* *conjunto formado por todas las soluciones del sistema.*

Definición 1.3 *El sistema* (1.1.1) *se dice:*

i) **Incompatible**, *si no tiene solución.*

ii) **Compatible determinado**, *si tiene única solución.*

iii) **Compatible indeterminado**, *si tiene más de una solución.*

Definición 1.4 **(Matriz ampliada del sistema)** *Dado el sistema* (1.1.1), $A\boldsymbol{x} = \boldsymbol{b}$, *llamaremos matriz ampliada del sistema a la matriz* $(A|\boldsymbol{b})$ *de orden* $m \times (n+1)$

Teorema 1.1 **(Existencia de soluciones)** *El sistema* (1.1.1) *es compatible si y sólo si* $r(A) = r(A|\boldsymbol{b})$.

Teorema 1.2 **(Unicidad de soluciones)** *Supongamos que el sistema* (1.1.1) *de* m *ecuaciones y* n *incógnitas es compatible y que* $r(A) = n$. *Entonces la solución del sistema es única.*

Teorema 1.3 **(Multiplicidad de soluciones)** *Si el sistema* (1.1.1) *es compatible y* $r =: r(A) < n$, *entonces a lo más* r *incógnitas se expresan en términos de las* $n - r$ *restantes.*

Observaciones:

i) Consideremos el sistema (1.1.1), $A\boldsymbol{x} = \boldsymbol{b}$. Si F representa a cualquiera de las tres operaciones elementales por filas, entonces $F(A)\boldsymbol{x} = F(\boldsymbol{b})$.

ii) Si $(A|\boldsymbol{b})$ es equivalente por filas a la matriz $(A_1|\boldsymbol{b}_1)$, entonces el sistema $A_1\boldsymbol{x} = \boldsymbol{b}_1$, es compatible si y sólo si el sistema (1.1.1) es compatible. En este caso, el conjunto solución de ambos sistemas es el mismo.

iii) El método en donde se obtiene $(C|\boldsymbol{c})$ (equivalente por filas con $(A|\boldsymbol{b})$) escalonada por filas se denomina eliminación Gaussiana.

iv) Un caso particular de lo anterior es la sucesión de operaciones elementales que transforman la matriz A en la matriz identidad. Aplicando las mismas operaciones a la matriz ampliada se obtiene el **Método de eliminación de Gauss-Jordan**.

v) Notar que el sistema homogéneo $A\boldsymbol{x} = \theta$ siempre tiene solución. Además el sistema homogéneo $A\boldsymbol{x} = \theta$ tiene solución no nula si y sólo si $r(A) < n$ (n es el número de incógnitas del sistema).

Ejercicio 1.1 *Muestre que el siguiente sistema de ecuaciones tiene única solución. Encuentrela,*

$$\left.\begin{array}{rcr} x_1 - 2x_2 + 4x_3 & = & 3 \\ -x1 + 3x_2 + 2x_3 & = & 4 \\ 5x_1 - 6x_3 & = & -1 \end{array}\right\}$$

Definición 1.5 (Sistemas de Cramer) *Si $A \in \mathcal{M}_n(\mathbb{K})$, con $|A| \neq 0$, entonces la matriz A es invertible y el sistema $A\boldsymbol{x} = \boldsymbol{b}$, de n ecuaciones y n incógnitas, tiene solución única*

$$\boldsymbol{x} = A^{-1}\boldsymbol{b}.$$

Recordando que $A^{-1} = \dfrac{1}{|A|}\text{adj}(A) = \dfrac{1}{|A|}(c_{ij})^t_{1\le i,j\le n}$, obtenemos la regla de Cramer.

Definición 1.6 (Regla de Cramer) *Si $A \in \mathcal{M}_n(\mathbb{K})$ con $|A| \neq 0$, entonces la única solución del sistema $A\boldsymbol{x} = \boldsymbol{b}$ es*

$$\boldsymbol{x} = (x_1, \ldots, x_n)^t, \quad \textit{con} \quad x_i = \frac{1}{|A|}\sum_{k=1}^{n} c_{ki}b_k, \quad i \in \{1, \ldots, n\}.$$

Observación: Notar que para $i \in \{1, \ldots, n\}$, se tiene que:

$$x_i = \frac{|A_i|}{|A|},$$

donde A_i es la matriz de orden n, obtenida de la matriz A en que la columna i-ésima de A es reemplazada por los elementos de B.

Ejercicio 1.2 *Resuelva usando la regla de Cramer el siguiente sistema de ecuaciones lineales,*

$$\left\{\begin{array}{rcl} x + y - z & = & 0 \\ 3x + y - z & = & 2 \\ 4x - 2y + z & = & 2 \end{array}\right.$$

Ejercicio 1.3 (Interpolación polinomial) *Determine el polinomio cuadrático que interpola los puntos en el plano* $(1, 3)$*,* $(2, 4)$ *y* $(3, 7)$*.*

Ejercicio 1.4 *Considere el siguiente sistema de ecuaciones lineales,*

$$\left\{\begin{array}{rcl} x + \lambda y + z & = & 1 \\ -2x - y + z & = & 3 \\ -x + y + \lambda z & = & \lambda \end{array}\right.$$

con $\lambda \in \mathbb{R}$. Detrermine todos los valores de $\lambda \in \mathbb{R}$ para los cuales el sistema tiene:

a) Única solución.

b) *Infinitas soluciones.*

c) *no tiene solución.*

Ejercicio 1.5 (Problema de valores y vectores propios de una matriz) *Considere la matriz A dada por,*

$$A = \begin{pmatrix} 1 & 0 & 5 \\ 1 & 1 & 1 \\ 0 & 1 & -4 \end{pmatrix}$$

a) *Encuentre los valores de $\lambda \in \mathbb{R}$ tal que el sistema $A\boldsymbol{x} = \lambda\boldsymbol{x}$ tenga solución no trivial (valores propios de A).*

b) *Resuelva el sistema $(A + 4I)\boldsymbol{x} = \boldsymbol{0}$ (Espacio propio asociado al valor propio -4).*

1.1.1 Sistemas rectangulares: Solución en el sentido de mínimos cuadrados

Considere un sistema rectangular de ecuaciones

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

donde $A \in \mathbb{R}^{m\times n}$, con $n < m$, es una matriz rectangular de m filas y n columnas y $\boldsymbol{b} \in \mathbb{R}^m$. Este problema, en general, no tiene solución: **sistema sobredeterminado**.

Una alternativa es buscar una solución en el siguiente sentido generalizado:

Hallar $\boldsymbol{x} \in \mathbb{R}^n$ tal que $\|\boldsymbol{b} - A\boldsymbol{x}\|_2$ sea **mínima**.

Definición 1.7 *El vector $\boldsymbol{x}$ que minimiza $\|\boldsymbol{b} - A\boldsymbol{x}\|_2$ es la* **solución en el sentido de mínimos cuadrados** *del sistema rectangular.*

En general:

$$A\boldsymbol{x} \neq \boldsymbol{b}.$$

1.1.2 Aplicación

Dado un conjunto de puntos

$$(x_1, y_1), \ldots, (x_m, y_m),$$

nos proponemos encontrar el polinomio

$$p(x) = c_0 + c_1 x + \ldots + c_{n-1}x^{n-1},$$

con $n < m$ que esté **más cerca** de estos puntos en el sentido que

$$\sum_{i=1}^{m} |p(x_i) - y_i|^2,$$

sea mínima.

Esta suma de cuadrados es el cuadrado de la norma del residuo del sistema rectangular:

$$\begin{pmatrix} 1 & x_1 & \dots & x_1^{n-1} \\ 1 & x_2 & \dots & x_2^{n-1} \\ 1 & x_3 & \dots & x_3^{n-1} \\ \vdots & \vdots & & \vdots \\ 1 & x_m & \dots & x_m^{n-1} \end{pmatrix} \begin{pmatrix} c_0 \\ c_1 \\ \vdots \\ c_{n-1} \end{pmatrix} = \begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_m \end{pmatrix}.$$

Teorema 1.4 *Sean $A \in \mathbb{R}^{m\times n}(m \geq n)$ y $\boldsymbol{b} \in \mathbb{R}^m$. Un vector $\boldsymbol{x} \in \mathbb{R}^n$ minimiza la norma del residuo $\|\boldsymbol{r}\|_2 = \|\boldsymbol{b} - A\boldsymbol{x}\|_2$ si y sólo si el residuo $\boldsymbol{r}$ es ortogonal a la imagen de A; esto es si*

$$A^t\boldsymbol{r} = \boldsymbol{0},$$

donde A^t es la matriz transpuesta de A.

En consecuencia, $\boldsymbol{x}$ debe satisfacer

$$A^t\boldsymbol{r} = \boldsymbol{0} \quad \Longleftrightarrow \quad A^t(\boldsymbol{b} - A\boldsymbol{x}) = \boldsymbol{0} \quad \Longleftrightarrow \quad A^tA\boldsymbol{x} = A^t\boldsymbol{b}.$$

Estas últimas ecuaciones reciben el nombre de **ecuaciones normales**.

Obs:

1. En el caso en que $m = n$ y que la matriz A sea una matriz no singular, entonces las ecuaciones normales entregan como solución la solución del sistema lineal $A\boldsymbol{x} = \boldsymbol{b}$.

2. Las ecuaciones normales tienen solución única si y sólo si todas las columnas de A son l.i.; es decir, si $rango(A) = n$.

3. En este caso, además, la matriz A^tA es **simétrica y definida positiva**, de donde, las ecuaciones normales tienen solución única.

Para resolver las ecuaciones normales se puede proceder del siguiente modo:

1. Calcular la matriz A^tA y el vector $A^t\boldsymbol{b}$.

2. Obtener la matriz L de la factorización de Cholesky: $A^tA = LL^t$.

3. Resolver el sistema triangular inferior $L\boldsymbol{y} = A^t\boldsymbol{b}$.

4. Resolver el sistema triangular superior $L^t\boldsymbol{x} = \boldsymbol{y}$.

Ejemplo 1.1 *Ajustar los siguientes datos a un polinomio de grado 2 en el sentido de mínimos cuadrados.*

x	*f(x)*
-3	14
-1	4
1	2
3	8
5	22
7	44

1.2 Espacios vectoriales

Definición 1.8 (Cuerpo) *Sea $\mathbb{K}$ un conjunto y sean $+$ y $\cdot$ dos operaciones binarias internas definidas sobre $\mathbb{K}$, llamadas* **suma** *y* **producto** *respectivamente. Diremos que $\mathbb{K}$, con estas operaciones, es un* **cuerpo** *si se satisfacen los siguientes axiomas:*

1. $\forall x, y, z \in \mathbb{K}, \quad x + (y + z) = (x + y) + z.$ **Asociatividad de +**
2. $\forall x, y \in \mathbb{K}, \quad x + y = y + x.$ **Conmutatividad de +**
3. $\exists 0 \in \mathbb{K}, \forall x \in \mathbb{K}, \quad x + 0 = x.$ **Elemento neutro para +**
4. $\forall x \in \mathbb{K}, \exists - x \in \mathbb{K}, \quad x + (-x) = 0.$ **Simétrico**
5. $\forall x, y, z \in \mathbb{K}, \quad x \cdot (y \cdot z) = (x \cdot y) \cdot z.$ **Asociatividad de ·**
6. $\exists 1 \in \mathbb{K}, \forall x \in \mathbb{K}, \quad 1 \cdot x = x \cdot 1 = x.$ **Elemento neutro para ·**
7. $\forall x \in \mathbb{K}, x \neq 0, \exists x^{-1} \in \mathbb{K}, \quad x \cdot x^{-1} = x^{-1} \cdot x = 1.$ **Inverso de x**
8. $\forall x, y, z \in \mathbb{K}, \quad x \cdot (y + z) = x \cdot y + x \cdot z \quad \wedge \quad (x + y) \cdot z = x \cdot z + y \cdot z.$
   Distributividad de · respecto de la suma +
9. *Diremos que $\mathbb{K}$ es un* **cuerpo conmutativo**, *si además se satisface:*

$$\forall x, y \in \mathbb{K}, \quad x \cdot y = y \cdot x.$$

Observaciones:

1. Escribiremos la terna $(\mathbb{K}, +, \cdot)$ para indicar que el conjunto $\mathbb{K}$ con las operaciones $+$ y $\cdot$ es un cuerpo.
2. Los conjuntos $\mathbb{Q}$, $\mathbb{R}$ y $\mathbb{C}$ de los números racionales, reales y complejos respectivamente, constituyen

cuerpos conmutativos con las operaciones de suma y producto usuales.

Definición 1.9 **(Espacio vectorial)** *Sean V un conjunto, $\mathbb{K}$ un cuerpo y*

$$\begin{aligned} + : V \times V &\to V \\ (x, y) &\to x + y, \end{aligned}$$

una operación binaria interna llamada **suma**,

$$\begin{aligned} \cdot : \mathbb{K} \times V &\to V \\ (\alpha, x) &\to \alpha x, \end{aligned}$$

una operación binaria externa llamada **producto por escalar**.
Diremos que V con las operaciones $+$ y $\cdot$ es un **espacio vectorial** *sobre $\mathbb{K}$ o un* **$\mathbb{K}$-espacio vectorial**, *si:*

1. $\forall x, y, z \in V, \quad x + (y + z) = (x + y) + z.$
2. $\forall x, y \in V, \quad x + y = y + x.$
3. $\exists \theta_V \in V$ *(vector nulo) para* $+$, *tal que,* $\quad x + \theta_V = x, \forall x \in V.$
4. $\forall x \in V, \exists - x \in V, \quad x + (-x) = \theta_V.$
5. $\forall \alpha, \beta \in \mathbb{K}, \forall x \in V, \quad \alpha(\beta x) = (\alpha\beta)x.$
6. $\forall \alpha \in \mathbb{K}, \forall x, y \in V, \quad \alpha(x + y) = \alpha x + \alpha y.$
7. $\forall \alpha, \beta \in \mathbb{K}, \forall x \in V, \quad (\alpha + \beta)x = \alpha x + \beta x.$
8. $\forall x \in V,\ 1 \cdot x = x,$ *donde 1 es el elemento unidad de* $\mathbb{K}$.

Observaciones:

1. Los elementos de V se denominan **vectores** y los elementos de $\mathbb{K}$, **escalares**.
2. Del axioma 3 se concluye que $V \neq \phi$.
3. Si $\mathbb{K} = \mathbb{R}$, diremos que V es un espacio vectorial real. Si $\mathbb{K} = \mathbb{C}$, diremos que V es un espacio vectorial complejo.
4. Cualesquiera sean los vectores x e y de V, $x + (-y)$ se escribe como $x - y$ y se llama **diferencia** entre x e y.

1.2.1 Propiedades de los espacios vectoriales

Teorema 1.5 *Sea V un espacio vectorial sobre un cuerpo $\mathbb{K}$. Entonces*

1. *El elemento neutro θ_V para la operación suma, es único.*
2. *Para cada $x \in V$ existe un único simétrico (inverso aditivo) $-x \in V$.*
3. *Ley de cancelación*
$$\forall x, y, z \in V, \;\; x + y = x + z \Rightarrow y = z$$
4. *Para todo $x \in V$, $0 \cdot x = \theta_V$.*
5. *Para todo $\alpha \in \mathbb{K}$, $\alpha \cdot \theta_V = \theta_V$.*
6. *Para todo $\alpha \in \mathbb{K}$, para todo $x \in V$, $(-\alpha) \cdot x = -(\alpha \cdot x)$.*
7. *Para todo $\alpha \in \mathbb{K}$, para todo $x \in V$, $\alpha \cdot x = \theta_V \Leftrightarrow (\alpha = 0 \vee x = \theta_V)$.*

Ejemplo 1.2

1. *Sea $\mathbb{K}$ un cuerpo, entonces $\mathbb{K}$ es un espacio vectorial sobre si mismo, lo cual se sigue trivialmente de las propiedades de cuerpo.*
2. *Consideremos el conjunto $\mathbb{R}^n$ de todas las n-uplas de números reales, es decir:*
$$V = \mathbb{R}^n = \{(x_1, x_2, \ldots, x_n) : x_i \in \mathbb{R}, i = 1, \ldots, n.\},$$
$\mathbb{R}^n$ es un $\mathbb{R}$-espacio vectorial con las siguientes operaciones
$$\begin{aligned} + : \; & \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}^n \\ & (x_1, x_2, \ldots, x_n) + (y_1, y_2, \ldots, y_n) = (x_1 + y_1, x_2 + y_2, \ldots, x_n + y_n) \\ \cdot : \; & \mathbb{R} \times \mathbb{R}^n \to \mathbb{R}^n \\ & \alpha(x_1, x_2, \ldots, x_n) = (\alpha x_1, \alpha x_2, \ldots, \alpha x_n) \end{aligned}$$
3. *Si en $\mathbb{R}^2$ se definen la suma como en el ejemplo anterior y el producto por escalar como sigue:*
$$\begin{aligned} \cdot : \; & \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2 \\ & \alpha(x_1, x_2) = (\alpha x_1, x_2), \end{aligned}$$
Luego $\mathbb{R}^2$ con estas operaciones no es un espacio vectorial.
4. *El símbolo $\mathbb{K}^X$ denota el conjunto de todas las funciones con dominio un conjunto $X \neq \phi$ y codominio en el cuerpo $\mathbb{K}$, es decir*
$$\mathbb{K}^X = \{f | f : X \to \mathbb{K}\}$$

En $\mathbb{K}^X$ definimos la suma de funciones y el producto de un escalar por una función como sigue. Si f y g son dos elementos cualesquiera de $\mathbb{K}^X$, entonces:

$$\begin{aligned} f+g : & X \to \mathbb{K} \\ & (f+g)(x) = f(x)+g(x), \quad \forall x \in X. \end{aligned}$$

Si α es cualquier escalar en $\mathbb{K}$ y f cualquier elemento de $\mathbb{K}$, entonces:

$$\begin{aligned} \alpha f : & X \to \mathbb{K} \\ & (\alpha f)(x) = \alpha f(x), \quad \forall x \in X. \end{aligned}$$

$(\mathbb{K}^X, +, \cdot)$ un $\mathbb{K}$-espacio vectorial.

5. *El conjunto $\mathcal{M}_{n\times m}(\mathbb{K})$ de las matrices de orden $n \times m$ con elementos en el cuerpo $\mathbb{K}$ y con las operaciones de suma y producto por un escalar usuales, es un $\mathbb{K}$-espacio vectorial. Los vectores de este espacio son matrices.*

1.2.2 Subespacio vectorial

Definición 1.10 *Sea V un $\mathbb{K}$-espacio vectorial y S un subconjunto de V. Diremos que S es un* **subespacio vectorial** *de V, si S es un espacio vectorial sobre $\mathbb{K}$ con las mismas operaciones de suma y producto por escalar definidas en V.*

Observación: Cualquiera sea el espacio vectorial V, tanto $\{\theta_V\}$ como V son subespacios de V llamados subespacios triviales.

Teorema 1.6 *Sea V un espacio vectorial sobre $\mathbb{K}$. Diremos que $S \subset V$ es* **subespacio** *vectorial de V si y sólo si*

1. *$S \neq \emptyset$*
2. *si para cualquier $x, y \in S$, se cumple que $x + y \in S$ y*
3. *si $\lambda \in \mathbb{K}$ y $x \in S$, entonces $\lambda x \in S$.*

Ejemplo 1.3

1. *Sea $S = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{R}) : a = b \wedge c = -d \right\}$. S es un subespacio del espacio vectorial $\mathcal{M}_2(\mathbb{R})$.*
2. *Sea $S = \{(x, y, z) \in \mathbb{R}^3 : x + y + z = 1\}$. S no es subespacio vectorial de $\mathbb{R}^3$.*
3. *El conjunto $V = C_{[0,1]}(\mathbb{R})$ de las funciones reales continuas sobre $[0, 1]$ es un subespacio del espacio vectorial $\mathbb{R}^{[0,1]}$ de las funciones de $[0, 1]$ en $\mathbb{R}$.*

4. *Sea* $A \in M_{m\times n}(\mathbb{K})$*. Si* $x = (x_1, \ldots, x_n) \in \mathbb{K}^n$*, entonces* Ax *denotará la multiplicación de* A *por la matriz columna formada por* $x_1, \ldots, x_n$*, es decir*

$$Ax = A \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}.$$

Sea

$$S = \{x \in \mathbb{K}^n : Ax = 0\}.$$

Es decir, S *es el subconjunto de* $\mathbb{K}^n$ *de las soluciones del sistema* $Ax = 0$*. Entonces,* S *es un subespacio vectorial de* $\mathbb{K}^n$*.*

1.2.3 Intersección de subespacios

Teorema 1.7 *Si* S *y* T *son dos subespacios vectoriales de un mismo* $\mathbb{K}$*-espacio vectorial* V*, entonces la intersección de* S *y* T*,* $S \cap T$*, es un subespacio vectorial de* V*.*

Demostración: Hacer en clase.

1. $S \cap T \neq \phi$.
2. $S \cap T$ es cerrado para la suma.
3. $S \cap T$ es cerrado para multiplicación por escalar.

1.2.4 Unión de subespacios

En general la unión de subespacios no es un subespacio vectorial. Veamos el siguiente ejemplo.

Ejemplo 1.4 1. *Sean* $S = \{(x, y) \in \mathbb{R}^2 : y = 2x\}$ *y* $T = \{(x, y) \in \mathbb{R}^2 : y = \frac{x}{2}\}$ *subespacios vectoriales. Luego* $u = (1, 2) \in S$*, por lo tanto* $u \in S \cup T$*. Análogamente,* $v = (2, 1) \in T$*, por lo tanto* $v \in S \cup T$*. Pero,* $u + v = (3, 3) \notin S$ *y* $(3, 3) \notin T$*. Luego,* $u + v \notin S \cup T$*. Por lo tanto* $S \cup T$ *no es un subespacio vectorial.*

Teorema 1.8 *Si* S *y* T *son subespacios de un* $\mathbb{K}$*-espacio vectorial* V*, entonces* $S \cup T$ *es subespacio si y sólo si* $S \subseteq T \vee T \subseteq S$*.*

1.2.5 Suma de espacios vectoriales

Sean S y T dos subespacios del mismo $\mathbb{K}$-espacio vectorial V, se llama **suma** de S y T al conjunto

$$S + T = \{v \in V : v = s + t, s \in S \wedge t \in T\}.$$

Teorema 1.9 *$S + T$ es un subespacio vectorial de V.*

Demostración: Hacer en clase.

Observación: Si $S \cap T = \{\theta_V\}$, la suma $S + T$ se llama **suma directa** y se escribe $S \oplus T$.

Ejemplo 1.5

1. *Sea*

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

S es el subespacio representado por el eje X, y

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

El subespacio T es el plano coordenado YZ. Muestre que $S \oplus T = \mathbb{R}^3$.

2. *En $V = \mathcal{M}_2(\mathbb{R})$, consideremos los subespacios*

$$S = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{R}) : b = c = 0 \right\},$$

y

$$T = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{R}) : b = d = 0 \right\}.$$

Calcule $S + T$ y muestre que $S + T$ no es suma directa.

3. *Sean $S = \{A \in \mathcal{M}_n(\mathbb{R}) : A = A^t\}$ y $T = \{A \in \mathcal{M}_n(\mathbb{R}) : A = -A^t\}$. Muestre que $\mathcal{M}_n(\mathbb{R}) = S \oplus T$.*

1.2.6 Combinaciones lineales. Dependencia e independencia lineal

Definición 1.11 (Combinación lineal) *Sean, V un $\mathbb{K}$-espacio vectorial, $x_1, \ldots, x_n$ vectores de V. El vector x es* **combinación lineal** *(C.L.) de $x_1, \ldots, x_n$ si existen escalares $\alpha_1, \ldots, \alpha_n \in \mathbb{K}$ tales que*

$$x = \alpha_1 x_1 + \cdots + \alpha_n x_n.$$

Observación: El vector nulo es *C.L.* de cualquier conjunto de vectores.

Ejemplo 1.6

1. *Decidir si $p(t) = t^2 - 2t + 3$ es C.L. de $p_1(t) = (t-1)^2$, $p_2(t) = \frac{1}{2}t + 1$, $p_3(t) = 5$.*

2. *Investigar si $A = \begin{pmatrix} 2 & -10 \\ 20 & 4 \end{pmatrix}$ es C.L de las matrices $A_1 = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$, $A_2 = \begin{pmatrix} 0 & -4 \\ 8 & 0 \end{pmatrix}$ y $A_3 = \begin{pmatrix} 0 & 1/4 \\ -1/2 & 0 \end{pmatrix}$ Respuesta: Sí (de infinitas maneras).*

V. Osores

Teorema 1.10 *Sean V un $\mathbb{K}$-espacio vectorial y $A = \{v_1, v_2, \ldots, v_r\} \subseteq V$. El conjunto de todas las combinaciones lineales de vectores de A es un subespacio vectorial de V.*

$$S = \{x \in V : x = \sum_{i=1}^{r} \alpha_i v_i, \alpha_i \in \mathbb{K}, i = 1, \ldots, r\}.$$

El subespacio S se llama **subespacio generado** *por A o* **subespacio generado** *por los vectores $v_1, \ldots, v_r$. Los vectores $v_1, \ldots, v_r$ se llaman* **generadores** *de S.*
Se denota por $S = \langle A \rangle = \langle \{v_1, \ldots, v_r\} \rangle$. También se dice que $v_1, \ldots, v_r$ **generan al subespacio** *o que A es un* **sistema** *de generadores de S.*

Demostración: Hacer en clase.

Ejemplo 1.7

1. *Caracterice el subespacio generado por $v_1 = (0, 1, 2)$, $v_2 = (-1, 3, -1)$ y $v_3 = (2, -11/2, 3)$.*

   Solución: $S = \{(x, y, z) \in \mathbb{R}^3 : z - 2y - 7x = 0\}$.

2. *Consideremos el subespacio de las matrices simétricas*

$$S = \left\{ \begin{pmatrix} a & b \\ b & c \end{pmatrix} \in \mathcal{M}_2(\mathbb{R}) : a, b, c \in \mathbb{R} \right\}.$$

   Encuentre el conjunto A tal que $S = \langle A \rangle$.

   Respuesta: $A = \left\{ \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix} \right\}$.

Definición 1.12 *Sean V un $\mathbb{K}$-espacio vectorial y $A = \{v_1, \ldots, v_r\} \subseteq V$; A es un conjunto* **linealmente dependiente** *(L.D.) si existen escalares no todos nulos $\alpha_1, \ldots, \alpha_r \in \mathbb{K}$ tales que*

$$\alpha_1 v_1 + \cdots + \alpha_r v_r = \theta_V,$$

Si A no es un conjunto L.D. se dice que es **linealmente independiente** *(L.I.).*

Definición 1.13 *S es un conjunto* **linealmente independiente** *si todo subconjunto finito de S es L.I.*

Observaciones:

1. Todo conjunto que consta de un único vector distinto del nulo es $L.I.$
2. Todo conjunto que contiene al vector nulo es $L.D.$

Ejemplo 1.8

1. *Muestre que $A = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\} \subseteq \mathbb{R}^3$, es un conjunto $L.I.$*

2. *Sea V el espacio de los polinomios de grado menor o igual que 2 con coeficientes reales, donde θ_V es el polinomio nulo. Sea $A = \{3x^2 - 2x, x^2 + 1, -3x + 2, x^2 - 1\} \subseteq \mathcal{P}_2(x)$. Muestre que A es $L.D.$*

Observaciones:

1. Si $v_1, \ldots, v_k$ es un conjunto de vectores $L.D.$, entonces uno de los vectores es $C.L.$ de los restantes.
2. Recíprocamente, si un vector v es $C.L.$ de $v_1, \ldots, v_k$, entonces $\{v, v_1, \ldots, v_k\}$ es $L.D.$
3. Si el conjunto $A = \{v_1, \ldots, v_k\}$ es un sistema de generadores $L.D.$ del espacio vectorial V, entonces existe $v_j \in A$ tal que $A - \{v_j\}$ es un sistema de generadores de V.

Definición 1.14 (*L.I.* **maximal**) *Un subconjunto $A = \{v_1, \ldots, v_n\}$ del $\mathbb{K}$-espacio vectorial V es $L.I.$* **maximal**, *si A es $L.I.$ y si $A \cup \{w\}$ es $L.D.$, cualquiera sea $w \in V$, $w \neq v_i$, $i = 1, \ldots, n$.*

Ejemplo 1.9 *$A = \{(1, 0), (0, 1)\} \subseteq \mathbb{R}^2$ es un conjunto $L.I.$ maximal puesto que es $L.I.$ y cualquier vector de $\mathbb{R}^2$ se puede escribir como combinación lineal de los vectores de A. Luego $A \cup \{w\}$ es $L.D.$, para todo $w \in \mathbb{R}^2$.*

Definición 1.15 **(Base de V)** *Sea V un $\mathbb{K}$-espacio vectorial. $A = \{v_1, \ldots, v_n\} \subseteq V$ es una* **base** *de V si:*

1. *A es $L.I.$*
2. *A es un sistema de generadores de V.*

Teorema 1.11 *Todo espacio vectorial posee base.*

Demostración: La demostración del Teorema 1.11 no está al alcance de este curso.

Ejemplo 1.10
1. *Si $V = \mathbb{R}^3$, se puede demostrar fácilmente que el conjunto $A = \{(1, 0, 0), (0, 1, 0), (0, 0, 1)\}$ es $L.I.$ y un sistema de generadores de $\mathbb{R}^3$. Por lo tanto, A constituye una base de $\mathbb{R}^3$, llamada* **base canonica** *de $\mathbb{R}^3$.*
2. *El conjunto $B = \{3, x - 1, x^2 + x\}$ es base del espacio vectorial $\mathcal{P}_2(x)$ (polinomios de grado menor o igual a 2, con coeficientes reales).*

Teorema 1.12 *Sea V un $\mathbb{K}$-espacio vectorial generado por un conjunto finito de vectores $v_1, \ldots, v_n$. Entonces todo conjunto $L.I.$ de vectores de V es finito y contiene a lo más n vectores.*

Definición 1.16 **(Dimension de V)** *Se llama* **dimensión** *de un* $\mathbb{K}$*-espacio vectorial* V *al número de elementos de una base cualquiera de* V. *Se denota* $\dim(V)$. *Si* V *consiste únicamente en él vector nulo, diremos que su dimensión es* 0.

1.3 Espacios vectoriales con producto interior

Definición 1.17 **(Productor interior)** *Sea* V *un espacio vectorial sobre* $\mathbb{K}$. *Una aplicación* $\langle\ ,\ \rangle : V \times V \to \mathbb{K}$ *se dice un* **producto interior** *sobre* V *si verifica:*

1. $\langle v, v\rangle \geq 0 \quad \forall v \in V, \qquad \langle v, v\rangle = 0 \Leftrightarrow v = \theta.$
2. $\forall v_1, v_2$ *y* $w \in V$, $\forall \alpha, \beta \in \mathbb{K}$,
$$\langle \alpha v_1 + \beta v_2, w\rangle = \alpha\langle v_1, w\rangle + \beta\langle v_2, w\rangle.$$
3. $\langle v, w\rangle = \overline{\langle w, v\rangle} \quad \forall v, w \in V.$

Propiedades 1.1 1. *De (3) se tiene que*
$$\langle v, v\rangle = \overline{\langle v, v\rangle} \quad \forall v \in V,$$
de donde se puede concluir que $\langle v, v\rangle \in \mathbb{R}$.

2. *Si* $\mathbb{K} = \mathbb{R}$ *y* V *espacio vectorial sobre* $\mathbb{R}$ *entonces (3) se transforma en*
$$\langle v, w\rangle = \langle w, v\rangle \quad \forall v, w \in V.$$

3. $\forall w_1, w_2$ *y* $v \in V$, $\forall \alpha_1, \alpha_2 \in \mathbb{K}$
$$\langle v, \alpha_1 w_1 + \alpha_2 w_2\rangle = \overline{\alpha}_1\langle v, w_1\rangle + \overline{\alpha}_2\langle v, w_2\rangle.$$
En particular, si $\mathbb{K} = \mathbb{R}$ *se tiene*
$$\langle v, \alpha_1 w_1 + \alpha_2 w_2\rangle = \alpha_1\langle v, w_1\rangle + \alpha_2\langle v, w_2\rangle.$$

Ejemplo 1.11 **(Producto interior)** *1.- Sea* $V = \mathbb{C}^n$ *espacio vectorial sobre* $\mathbb{C}$.
$$\langle\ ,\ \rangle : \mathbb{C}^n \times \mathbb{C}^n \to \mathbb{C}$$
$$\langle u, v\rangle = \sum_{i=1}^{n} u_i \overline{v}_i,$$
es un producto interior.

V. Osores

2. *Sea* $\mathcal{M}_{n\times m}(\mathbb{R})$,

$$\begin{aligned} \langle\ ,\ \rangle : \mathcal{M}_{n\times m} \times \mathcal{M}_{n\times m} &\to \mathbb{R} \\ \langle A, B\rangle &= tr(A^t B), \end{aligned}$$

es un producto interior. Donde la **traza** *de una matriz cuadrada* A *es la suma de los elementos de la diagonal principal y se denota por* $tr(A)$.

3. *Sea* $C_{\mathbb{R}}[a,b]$ *el espacio vectorial de las funciones continuas de* $[a,b]$ *en* $\mathbb{R}$.

$$\begin{aligned} \langle\ ,\ \rangle : C_{\mathbb{R}}[a,b] \times C_{\mathbb{R}}[a,b] &\to \mathbb{R} \\ \langle f, g\rangle &= \int_a^b f(x)g(x)dx, \end{aligned}$$

es un producto interior.

Definición 1.18 **(Norma)** *Sea* V *un espacio vectorial con producto interior, se llama norma del vector* v *al número* $\|v\| = \sqrt{\langle v, v\rangle}$.

Propiedades 1.2 $\forall u, v, w \in V$ *y* $\alpha \in \mathbb{K}$,

1. $\|v\| = 0 \Leftrightarrow v = \theta$.
2. $\|\alpha v\| = |\alpha|\|v\|$.
3. $|\langle v, w\rangle| \leq \|v\|\|w\|$. **(Desigualdad de Cauchy-Schwartz)**
4. $\|u + v\| \leq \|u\| + \|v\|$. **(Desigualdad triangular)**

Definición 1.19 *Sea* V *un* $\mathbb{K}$*-espacio vectorial con producto interior*

1. $v, w \in V$, v *es* **ortogonal** *a* w *si* $\langle v, w\rangle = 0$.
2. *un subconjunto* $\{v_1, \ldots, v_n\} \subseteq V$ *es un* **conjunto ortogonal** *si* $\langle v_i, v_j\rangle = 0 \quad \forall i \neq j$.
3. *un subconjunto* $\{v_1, \ldots, v_n\} \subseteq V$ *es un conjunto* **ortonormal** *si es un conjunto ortogonal y* $\|v_i\| = 1 \quad \forall i = \{1, \ldots, n\}$.

Lema 1.1 *Si* $\{v_1, \ldots, v_n\}$ *es un conjunto ortogonal de vectores no nulos, entonces* $\{v_1, \ldots, v_n\}$ *es L.I.*

Demostración: Demostrar en clases.

Corolario 1.1 *Si un vector w es C.L. de un conjunto ortogonal de vectores no nulos $\{x_1,\dots,x_n\}$, entonces w es igual a*

$$w=\sum_{k=1}^{n}\frac{\langle w,x_k\rangle}{\|x_k\|^2}x_k.$$

Demostración: Demostrar en clases.

Proposición 1.1 (**Proceso de ortogonalización de Gram-Schmidt**) *Sea V espacio vectorial de dimensión finita con producto interior $\langle\,,\rangle$ y $\{v_1,\dots,v_n\}$ una base de V. Entonces existe una base ortogonal $\{w_1,\dots,w_n\}$ tal que el subespacio generado por los vectores $\{w_1,\dots,w_m\}$ es el mismo que el subespacio generado por $\{v_1,\dots,v_m\}$ $(1\le m\le n)$. Explícitamente, la base es*

$$\begin{aligned}
w_1 &= v_1, && (1)\\
w_2 &= v_2-\frac{\langle v_2,w_1\rangle}{\|w_1\|^2}w_1, && (2)\\
w_3 &= v_3-\frac{\langle v_3,w_1\rangle}{\|w_1\|^2}w_1-\frac{\langle v_3,w_2\rangle}{\|w_2\|^2}w_2, && (3)\\
\vdots\ & \ \vdots\\
w_n &= v_n-\frac{\langle v_n,w_1\rangle}{\|w_1\|^2}w_1-\frac{\langle v_n,w_2\rangle}{\|w_2\|^2}w_2-\cdots-\frac{\langle v_n,w_{n-1}\rangle}{\|w_{n-1}\|^2}w_{n-1}. && (n)
\end{aligned}$$

Ejemplo 1.12 1. *Sea $B=\{(3,0,4),(-1,0,7),(2,9,11)\}$ una base de $\mathbb{R}^3$. Encuentre una base ortogonal para $\mathbb{R}$.*

Respuesta: $B'=\{(3,0,4),(-4,0,3),(0,9,0)\}$

Observación: Sea V espacio vectorial de dimensión finita con producto interior $\langle\,,\,\rangle$ y $\{v_1,\dots,v_n\}$ una base de V. Entonces, por Gram-Schmidt existe una base ortogonal $\{w_1,\dots,w_n\}$ de V. Sea $u_i=w_i/\|w_i\|$, entonces $\{u_1,\dots,u_n\}$ es una base de V donde los vectores son ortogonales y de norma 1.

Definición 1.20 *Sea V espacio vectorial con producto interior $\langle\,,\,\rangle$ y sean U,W subespacios de V. Diremos que U es* **ortogonal** *a W y denotaremos $U\perp W$ si para todo $u\in U$ y para todo $w\in W$ tenemos que $\langle u,w\rangle=0$.*
Si X es subconjunto de V, definimos

$$X^{\perp}:=\{u\in V:\langle u,x\rangle=0,\forall x\in X\}. \qquad (\textbf{Complemento ortogonal de X})$$

Proposición 1.2 *Sea V espacio vectorial con producto interior $\langle\,,\,\rangle$ y $X\subseteq V$. Entonces $X^{\perp}$ es un subespacio de V.*

Demostración: Demostrar en clases.

Observación: Si W es subespacio y x es ortogonal a todo vector de una base de W entonces $x \in W^{\perp}$.

1. Decida si los sistemas que siguen son incompatibles o compatibles. En el último caso, decida si son determinados o indeterminados y encuentre la solución.

$$a)\ \left.\begin{array}{rcrcr} x & - & y & = & 2 \\ 2x & - & 2y & = & 8 \\ 3x & - & y & = & 4 \end{array}\right\}, \qquad b)\ \left.\begin{array}{rcrcr} x & - & y & = & 2 \\ x & + & 3y & = & 8 \\ 2x & + & 2y & = & 10 \end{array}\right\},$$

$$c)\ \left.\begin{array}{rcrcrcr} x & - & 2y & - & z & = & 2 \\ x & - & 2y & + & 2z & = & 1 \\ 2x & - & 3y & + & z & = & 2 \end{array}\right\}$$

2. Muestre que el siguiente sistema es compatible determinado y encuentre su solución por los método de Cramer y usando operaciones elementales (escalonando):

$$\left.\begin{array}{rcrcrcrcr} 2x_1 & - & 4x_2 & & & & & = & -10 \\ x_1 & - & 3x2 & & & + & x_4 & = & -4 \\ x_1 & & & - & x_3 & + & 2x_4 & = & 9 \\ 3x_1 & - & 4x_2 & + & 3x_3 & - & x_4 & = & -15 \end{array}\right\}.$$

¿Qué método le tomó más tiempo?

3. Determine el o los valores de p y q tales que el sistema: i) No tenga solución. ii) Tenga una única solución. iii) Tenga infinitas soluciones.

$$a)\ \left.\begin{array}{rcrcrcr} x & - & y & + & 2z & = & 2 \\ 3x & + & y & + & z & = & 1 \\ 2x & & & + & pz & = & q \end{array}\right\}, \qquad b)\ \left.\begin{array}{rcrcrcr} x & + & 2y & + & z & = & q \\ 2x & - & y & + & z & = & p \\ 3x & + & y & + & 2z & = & 1 \end{array}\right\}.$$

4. Determine los valores de $\alpha \in \mathbb{R}$ para que el sistema siguiente posea solución no trivial:

$$\left.\begin{array}{rcrcrcr} & & \alpha x & + & z & = & 0 \\ 2x & + & y & - & z & = & 0 \\ & & y & + & z & = & 0 \end{array}\right\}$$

5. Determine los valores que debe tomar el parámetro $\lambda \in \mathbb{R}$ para que el sistema siguiente sea compatible determinado:

$$\left.\begin{array}{ccccccc} (1-\lambda)x & + & y & + & z & = & a \\ x & + & (1-\lambda)y & + & z & = & b \\ x & + & y & + & (1-\lambda)z & = & c \end{array}\right\}$$

En el caso en que el sistema no sea determinado, determine las condiciones que debe satisfacer $(a,b,c) \in \mathbb{R}^3$ para que el sistema sea compatible indeterminado.

6. Sean U, V, W, Z los siguientes subespacios de $\mathbb{R}^3$.
$U = \{(x,y,z) \in \mathbb{R}^3 : x + y + z = 0\}$,
$V = \{(x,y,z) \in \mathbb{R}^3 : x = 0\}$,
$W = \{(0,0,z) \in \mathbb{R}^3 : z \in \mathbb{R}\}$,
$Z = \{(x,y,z) \in \mathbb{R}^3 : x = 3y = 2z\}$.
Caracterice los elementos de cada uno de los siguientes espacios:

a) $U+V$
b) $U+W$
c) $V+W$
d) $W+Z$
e) $U\cap W$
f) $V\cap W$
g) $U\cap Z$

7. Determine si los siguientes conjuntos son linealmente independientes.

   a) $\{(3,6,1),(2,1,1),(-1,0,-1)\}$ en $\mathbb{R}^3$
   b) $\left\{\begin{pmatrix}1&0\\0&1\end{pmatrix},\begin{pmatrix}1&1\\0&0\end{pmatrix},\begin{pmatrix}1&1\\1&1\end{pmatrix}\right\}$ en $\mathcal{M}_2(\mathbb{R})$
   c) $\{t^3-t^2+4t+1,\quad 2t^3-2t^2+9t-1,\quad t^3+6t-5,\quad 2t^3-5t^2+7t+5\}$ en $\mathcal{P}_3(\mathbb{R})$

8. Demuestre que los polinomios $\{(1-t)^3,\,(1-t)^2,\,(1-t),\,1\}$, generan el espacio de los polinomios de grado menor o igual que tres.

9. Sean $S_1=\{\operatorname{sen}^2(x),\cos^2(x),\operatorname{sen}(x)\cos(x)\}$ y $S_2=\{1,\operatorname{sen}(2x),\cos(2x)\}$. Muestre que los vectores de cada conjunto son $L.I.$

10. Encuentre una base y determine la dimensión de los siguientes subespacios:

    a) $Z=\{(x,y,z)\in\mathbb{R}^3 : x-2y-3z=0\}$,
    b) $Y=\{(x,y,x)\in\mathbb{R}^3 : x=3y\}$,
    c) $X=\{(x,y,z)\in\mathbb{R}^3 : 6x=3y=z\}$,
    d) $W=\{(a,b,c,d)\in\mathbb{R}^4 : b-2c+d=0\}$,
    e) $V=\{(a,b,c,d)\in\mathbb{R}^4 : a=d, b=2c\}$
    f) $U=\left\{\begin{pmatrix}a&b&b\\0&a&c\\0&d&a\end{pmatrix}\in\mathcal{M}_3(\mathbb{R})\right\}$,
    g) $T=\langle\{7-x^2,x^2+1,x^2-1\}\rangle$,
    h) $S=\langle\{\cos^2(x),\sin^2(x),\cos(2x)\}\rangle$,
    i) $R=\{p\in\mathcal{P}_3(\mathbb{R}) : p(0)=0\}$,
    j) $Q=\{p\in\mathcal{P}_3(\mathbb{R}) : p(1)=0\}$,
    k) $P=\{p\in\mathcal{P}_3(\mathbb{R}) : p(0)=p'(0)=0\}$.

11. Considere el conjunto $\mathcal{P}_2(\mathbb{R})$ con la suma usual de polinomios y la multiplicación por escalar definida por
$$\alpha p(x)=\alpha p'(x),\forall\alpha\in\mathbb{R},\forall p(x)\in\mathcal{P}_2(\mathbb{R}).$$
¿Es $\mathcal{P}_2(\mathbb{R})$ un espacio vectorial con estas operaciones?.

12. Considere la ecuación $x-2y+3z=0$.

    a) Muestre que el conjunto solución S de esta ecuación es un subespacio de $\mathbb{R}^3$.
    b) Encuentre una base para S y su dimensión.

13. Sea V un $\mathbb{K}$-espacio vectorial y $S_1=\{u,v,w\}$ un subconjunto $L.I$ de V. Demuestre que: $S_2=\{u+v,u-v,u-2v+w\}$ es también $L.I.$

14. Considere los siguientes subespacios de $\mathbb{R}^3$.
$U=\{(x,y,z)\in\mathbb{R}^3 : x=2y\}$
$V=\langle\{-1,2,1),(0,0,1)\}\rangle$
Caracterice los subespacios $U+V$ y $U\cap V$.

15. Encuentre la dimensión del subespacio $V = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{R}) : a = b \wedge c = d \right\}$.

16. Encuentre la dimensión del espacio $U = \{ax^2 + bx + c \in \mathcal{P}_2(\mathbb{R}) : 2b - c = 0\}$.

17. Dados los subespacios $U = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{R}) : a + c = 0 \right\}$ y

    $V = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathcal{M}_2(\mathbb{R}) : 2b - d = 0 \right\}$.

    a) Caracterice el subespacio $U \cap V$.
    b) ¿Es $U + V$ suma directa?.

18. Considere el conjunto $S = \{(x, y, z) \in \mathbb{R}^3 : 4x + 3y - z = 0\}$ y el subespacio T de $\mathbb{R}^3$ generado por $(3, -1, 1)$.

    a) Demuestre que S es un subespacio de $\mathbb{R}^3$.
    b) Determine una base para $S + T$ y decida si ésta es una suma directa.

19. *a*) Considere $\mathbb{R}^2$ con el producto interior usual. Si $x = (1, 2)$ y $y = (-1, 1)$, encuentre $v \in \mathbb{R}^2$ tal que
    $$\langle v, x\rangle = -2 \;\wedge\; \langle v, y\rangle = 3.$$
    b) Demuestre que para cada vector $u \in \mathbb{R}^2$, se tiene
    $$u = \langle u, e_1\rangle e_1 + \langle u, e_2\rangle e_2,$$
    donde $\{e_1, e_2\}$ es la base canónica de $\mathbb{R}$.

20. Encuentre una base ortonormal para $\mathbb{R}^3$ a partir de $\{(1, 1, 0), (-1, 1, 0), (-1, 1, 1)\}$.

21. Dado el vector $(2, 1, -1) \in \mathbb{R}^3$, construya a partir de él una base ortonormal de $\mathbb{R}^3$.

22. Considere el espacio vectorial $\mathbb{R}^3$ con el *p.i.* usual. Sea $S = \langle\{(1, 1, 1), (-1, 1, 0)\}\rangle$.

    a) Caracterice $S^\perp$ y determine su dimensión.
    b) Encontrar una base B ortonormal de $\mathbb{R}^3$ tal que uno de sus vectores sea elemento de $S^\perp$.

23. Considere el espacio vectorial real $\mathcal{P}_2(\mathbb{R})$ con el *p.i.*
    $$\langle p, q\rangle = 2\int_0^2 p(x)q(x)dx$$
    Pruebe que el conjunto $\{1, x - 2, x^2 - 2\}$ es *l.i.* y ortonormalice respecto del *p.i.* dado.

24. En $\mathbb{C}^2$ se define el producto interior
    $$\langle x, y\rangle = \sum_{i=1}^{n} x_i \overline{y}_i.$$
    Pruebe que los vectores $x = (3, -i)$, $y = (2, 6i)$ son ortogonales y normalícelos.

25. Pruebe que en el espacio $C_{\mathbb{R}}[0,1]$, con el producto interior
$$\langle f,g\rangle = \int_0^1 f(t)g(t)dt,$$
el conjunto $\{1, \sqrt{3}(2t-1), \sqrt{5}(6t^2-6t+1)\}$ es una base ortonormal.

26. En el espacio $C_{\mathbb{R}}[0,2\pi]$, con el producto interior
$$\langle f,g\rangle = \int_0^{2\pi} f(x)g(x)dx,$$
el conjunto $\{\sin(x), \cos(x)\}$ es ortogonal.

27. Pruebe que $\{\sin(nx), \cos(nx), 1\}$ es un conjunto ortogonal con el producto interior
$$\langle f,g\rangle = \int_{-1}^1 f(x)g(x)dx.$$

28. En el espacio de los polinomios reales de grado menor o igual que 2 con el producto
$$\langle p,q\rangle = \int_{-1}^1 p(x)q(x)dx.$$
construya a partir de la base $\{1, x, x^2\}$ una base ortonormal.

29. Sean x e y vectores de un espacio vectorial con $p.i.$ tales que $x+y$ es ortogonal a $x-y$. Demuestre que $\|x\| = \|y\|$.

30. Sea V un $\mathbb{K}$-espacio vectorial con $p.i.$. Demuestre que: $\forall x, y \in V$,
$$\|x+y\|^2 + \|x-y\|^2 = 2\|x\|^2 + 2\|y\|^2.$$

31. Sea $W = \{(x,y,z) \in \mathbb{R}^3 : x = 2y = z\}$. Halle $W^{\perp}$. ¿Qué representan geométricamente W y $W^{\perp}$?

CAPÍTULO 2

Diagonalización

2.1 Conceptos básicos

Designaremos por $\mathbb{C}^{n\times n}$ el espacio de las matrices cuadradas de orden n en el cuerpo $\mathbb{C}$, mientras que cuando los coeficientes pertenezcan al cuerpo $\mathbb{R}$, usaremos la notación $\mathbb{R}^{n\times n}$.

Definición 2.1 *Para* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *se definen la matriz transpuesta como la matriz* $\boldsymbol{B}$ *de elementos* $b_{ij} := a_{ji}$ *Notación:* $\boldsymbol{A}^T := \boldsymbol{B}$.

Definición 2.2 (Transpuesta Conjugada en $\mathbb{C}^{n\times n}$) *La transpuesta conjugada* $\boldsymbol{A}^*$ *de una matriz compleja* $\boldsymbol{A}$ *se define por*

$$\boldsymbol{A}^* = \bar{\boldsymbol{A}}^T = \overline{(\boldsymbol{A}^T)}$$

Observe que $\boldsymbol{A}^* = \boldsymbol{A}^T$ cuando $\boldsymbol{A}$ es real.

Ejemplo 2.1

$$\begin{bmatrix} 3 & 1-i & 2+i \\ 2i & 5+2i & -i \end{bmatrix}^* = \begin{bmatrix} 3 & -2i \\ 1+i & 5-2i \\ 2-i & i \end{bmatrix}$$

Las siguientes propiedades de $\boldsymbol{A}^*$ se deducen fácilmente de las reglas para la transposición de matrices reales y extienden estas reglas a matrices complejas. Note el conjugado en la propiedad (3).

Teorema 2.1 *Sean* $\boldsymbol{A}$ *y* $\boldsymbol{B}$ *matrices complejas, y sea* λ *un número complejo.*

1. $(\boldsymbol{A}^*)^* = \boldsymbol{A}$.
2. $(\boldsymbol{A}+\boldsymbol{B})^* = \boldsymbol{A}^* + \boldsymbol{B}^*$.
3. $(\lambda\boldsymbol{A})^* = \bar{\lambda}\boldsymbol{A}^*$.

4. $(\boldsymbol{AB})^* = \boldsymbol{B}^*\boldsymbol{A}^*$.

Definición 2.3 *Una matriz* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *se dice simétrica si* $\boldsymbol{A} = \boldsymbol{A}^T$*, hermitiana si* $\boldsymbol{A} = \boldsymbol{A}^*$*, ortogonal si* $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}$ *y unitaria si* $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A} = \boldsymbol{I}$.

Una manera de caracterizar la ortogonalidad, respectivamente la unitariedad, de una matriz $\boldsymbol{A}$ es a través de las igualdades $\boldsymbol{A}^{-1} = \boldsymbol{A}^T$ y $\boldsymbol{A}^{-1} = \boldsymbol{A}^*$, respectivamente.

Definición 2.4 (Producto interno estándar en $\mathbb{C}^n$) *Dado* $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)$ *y* $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$ *en* $\mathbb{C}^n$*, definimos su producto interno estándar* $\langle \boldsymbol{z}, \boldsymbol{w}\rangle$ *por*

$$\langle \boldsymbol{z}, \boldsymbol{w}\rangle = z_1\bar{w}_1 + z_2\bar{w}_2 + \cdots + z_n\bar{w}_n = \boldsymbol{z}\cdot\bar{\boldsymbol{w}}$$

donde $\bar{\boldsymbol{w}}$ *es el conjugado del vector complejo* $\boldsymbol{w}$.

Claramente, si $\boldsymbol{z}$ y $\boldsymbol{w}$ están en $\mathbb{R}^n$, entonces $\langle \boldsymbol{z}, \boldsymbol{w}\rangle = \boldsymbol{z}\cdot\boldsymbol{w}$ es el producto punto usual.

Ejemplo 2.2 *Si* $\boldsymbol{z} = (2, 1-i, 2i, 3-i)$ *y* $\boldsymbol{w} = (1-i, -1, -i, 3+2i)$*, entonces*

$$\langle \boldsymbol{z}, \boldsymbol{w}\rangle = 2(1+i) + (1-i)(-1) + (2i)(i) + (3-i)(3-2i) = 6-6i$$
$$\langle \boldsymbol{z}, \boldsymbol{z}\rangle = 2\cdot 2 + (1-i)(1+i) + (2i)(-2i) + (3-i)(3+i) = 20$$

Note que $\langle \boldsymbol{z}, \boldsymbol{w}\rangle$ es un número complejo en general. Sin embargo, si $\boldsymbol{w} = \boldsymbol{z} = (z_1, z_2, \ldots, z_n)$, la definición da $\langle \boldsymbol{z}, \boldsymbol{z}\rangle = |z_1|^2 + \cdots + |z_n|^2$, que es un número real no negativo, igual a 0 si y sólo si $\boldsymbol{z} = \boldsymbol{0}$. Esto explica la conjugación en la definición de $\langle \boldsymbol{z}, \boldsymbol{w}\rangle$, y da (4) del siguiente teorema.

Teorema 2.2 *Sean* $\boldsymbol{z}, \boldsymbol{z}_1, \boldsymbol{w}$ *y* $\boldsymbol{w}_1$ *vectores en* $\mathbb{C}^n$*, y sea* λ *un número complejo.*

1. $\langle \boldsymbol{z}+\boldsymbol{z}_1, \boldsymbol{w}\rangle = \langle \boldsymbol{z}, \boldsymbol{w}\rangle + \langle \boldsymbol{z}_1, \boldsymbol{w}\rangle$ *y* $\langle \boldsymbol{z}, \boldsymbol{w}+\boldsymbol{w}_1\rangle = \langle \boldsymbol{z}, \boldsymbol{w}\rangle + \langle \boldsymbol{z}, \boldsymbol{w}_1\rangle$.

2. $\langle \lambda\boldsymbol{z}, \boldsymbol{w}\rangle = \lambda\langle \boldsymbol{z}, \boldsymbol{w}\rangle$ *y* $\langle \boldsymbol{z}, \lambda\boldsymbol{w}\rangle = \bar{\lambda}\langle \boldsymbol{z}, \boldsymbol{w}\rangle$.

3. $\langle \boldsymbol{z}, \boldsymbol{w}\rangle = \overline{\langle \boldsymbol{w}, \boldsymbol{z}\rangle}$.

4. $\langle \boldsymbol{z}, \boldsymbol{z}\rangle \geq 0$*, y* $\langle \boldsymbol{z}, \boldsymbol{z}\rangle = 0$ *si y sólo si* $\boldsymbol{z} = \boldsymbol{0}$.

Demostración: *Se deja (1) y (2) al lector, y (4) ya ha sido demostrado. Para demostrar (3), escribimos* $\boldsymbol{z} = (z_1, z_2, \ldots, z_n)$ *y* $\boldsymbol{w} = (w_1, w_2, \ldots, w_n)$*. Entonces*

$$\begin{aligned}\overline{\langle \boldsymbol{w}, \boldsymbol{z}\rangle} = (\overline{w_1\bar{z}_1 + \cdots + w_n\bar{z}_n}) &= \bar{w}_1\bar{\bar{z}}_1 + \cdots + \bar{w}_n\bar{\bar{z}}_n \\ &= z_1\bar{w}_1 + \cdots + z_n\bar{w}_n = \langle \boldsymbol{z}, \boldsymbol{w}\rangle\end{aligned}$$

Teorema 2.3 *Si* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *es una matriz simétrica, entonces*

$$(\boldsymbol{A}\boldsymbol{x})\cdot\boldsymbol{y} = \boldsymbol{x}\cdot(\boldsymbol{A}\boldsymbol{y}).$$

para todas las columnas $\boldsymbol{x}$ y $\boldsymbol{y}$ en $\mathbb{R}^n$.[a]

[a] El reciproco también se tiene.

Demostración: *Recuerde que $\boldsymbol{x} \cdot \boldsymbol{y} = \boldsymbol{x}^T \boldsymbol{y}$ para todas las columnas $\boldsymbol{x}$ y $\boldsymbol{y}$. Como $\boldsymbol{A}^T = \boldsymbol{A}$, obtenemos*

$$(\boldsymbol{A}\boldsymbol{x}) \cdot \boldsymbol{y} = (\boldsymbol{A}\boldsymbol{x})^T \boldsymbol{y} = \boldsymbol{x}^T \boldsymbol{A}^T \boldsymbol{y} = \boldsymbol{x}^T \boldsymbol{A}\boldsymbol{y} = \boldsymbol{x} \cdot (\boldsymbol{A}\boldsymbol{y})$$

Definición 2.5 *Una matriz $\boldsymbol{A} \in \mathbb{K}^{n \times n}$ se llama estrictamente diagonal dominante si*

$$\forall i = 1, \dots, n : \quad |\alpha_{ii}| > \sum_{\substack{j=1 \\ j \neq i}}^{n} |\alpha_{ij}|$$

Una matriz $\boldsymbol{A} = \boldsymbol{A}^*$ se llama definida positiva si

$$\forall \boldsymbol{x} \in \mathbb{C}^n, \boldsymbol{x} \neq 0 : \quad \boldsymbol{x}^* \boldsymbol{A}\boldsymbol{x} > 0$$

Teorema 2.4 *Todos los subdeterminantes principales de una matriz $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ estrictamente diagonal dominante son diferentes de cero.*

Demostración: *Para demostrar el teorema, es suficiente demostrar que una matriz $\boldsymbol{A} = (a_{ij}) \in \mathbb{C}^{n \times n}$ estrictamente diagonal dominante es no singular, dado que cada submatriz principal de una matriz estrictamente diagonal dominante es estrictamente diagonal dominante. Para tal efecto, supongamos que $\boldsymbol{A}$ es una matriz estrictamente diagonal dominante del tamaño $n \times n$, pero que existe un vector $\boldsymbol{x} = (x_1, \dots, x_n)^T \neq 0$ tal que*

$$\boldsymbol{A}\boldsymbol{x} = 0 \tag{2.1.1}$$

Dado que $\boldsymbol{x} \neq 0$, existe un índice $m \in \{1, \dots, n\}$ tal que

$$|x_m| = \max\{|x_1|, \dots, |x_n|\} > 0 \tag{2.1.2}$$

Evaluando la m-ésima componente de (2.1.1), tenemos

$$a_{mm}x_m + \sum_{\substack{j \neq m \\ j=1}}^{n} a_{mj}x_j = 0$$

lo que podemos reescribir como

$$a_{mm}x_m = -\sum_{\substack{j \neq m \\ j=1}}^{n} a_{mj}x_j$$

Tomando valores absolutos, tenemos

$$|a_{mm}| \, |x_m| \leqslant \sum_{\substack{j \neq m \\ j=1}}^{n} |a_{mj}| \, |x_j|$$

y usando (2.1.2)*, llegamos a*

$$|a_{mm}|\,|x_m| \leqslant |x_m| \sum_{\substack{j\neq m\\ j=1}}^{n} |a_{mj}|$$

Dividiendo por $|x_m|$*, obtenemos*

$$|a_{mm}| \leqslant \sum_{\substack{j\neq m\\ j=1}}^{n} |a_{mj}|$$

una contradicción a la diagonaldominancia estricta de $\boldsymbol{A}$*.* ■

Definición 2.6 (Espacio de columnas y filas de una matriz) *El espacio de columnas,* $\operatorname{col} A$*, de* A *es el subespacio de* $\mathbb{R}^m$ *generado por las columnas de* A*. El espacio de filas, row* A*, de* A *es el subespacio de* $\mathbb{R}^n$ *generado por las filas de* A*.*

Lema 2.1 *Sean* A, U *y* V *matrices de tamaños* $m \times n, p \times m$ *y* $n \times q$ *respectivamente.*

1. $\operatorname{col}(AV) \subseteq \operatorname{col} A$*, con igualdad si* $VV' = I_n$ *para algún* V'*.*
2. $\operatorname{row}(UA) \subseteq \operatorname{row} A$*, con igualdad si* $U'U = I_m$ *para algún* U'*.*

Demostración: *Para (1), escriba* $V = [\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_q]$ *donde* $\boldsymbol{v}_j$ *es la columna* j *de* V*. Entonces tenemos* $AV = [A\boldsymbol{v}_1, A\boldsymbol{v}_2, \ldots, A\boldsymbol{v}_q]$*, y cada* $A\boldsymbol{v}_j$ *está en* $\operatorname{col} A$*. De ello se sigue que* $\operatorname{col}(AV) \subseteq \operatorname{col} A$*. Si* $VV' = I_n$*, obtenemos* $\operatorname{col} A = \operatorname{col}[(AV)V'] \subseteq \operatorname{col}(AV)$ *de la misma manera. Esto prueba (1).*

En cuanto a (2), tenemos $\operatorname{col}\left[(UA)^T\right] = \operatorname{col}\left(A^T U^T\right) \subseteq \operatorname{col}\left(A^T\right)$ *por (1), de lo cual se sigue que* $\operatorname{row}(UA) \subseteq \operatorname{row} A$*. Si* $U'U = I_m$*, esto es igualdad como en la demostración de (1).*

Teorema 2.5 (Teorema Fundamental) *Sea* U *un subespacio de* $\mathbb{R}^n$*. Si* U *está generado por* m *vectores, y si* U *contiene* k *vectores linealmente independientes, entonces* $k \leq m$*.*

Teorema 2.6 *Sea* U *un subespacio de* $\mathbb{R}^n$ *donde* $\dim U = m$ *y sea* $B = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_m\}$ *un conjunto de* m *vectores en* U*. Entonces,* B *es linealmente independiente si y sólo si* B *genera* U*.*

Teorema 2.7 *Sea* V *un espacio vectorial de dimensión finita generado por* m *vectores.*

1. V *tiene una base finita, y* $\dim V \leq m$*.*
2. *Todo conjunto de vectores linealmente independientes en* V *puede ampliarse a una base de* V *añadiendo vectores de cualquier base fija de* V*.*
3. *Si* U *es un subespacio de* V*, entonces*
   a) U *es de dimensión finita y* $\dim U \leq \dim V$*.*
   b) *Si* $\dim U = \dim V$*, entonces* $U = V$*.*

2.2 Valores y vectores propios

Definición 2.7 *Un escalar $\lambda \in \mathbb{C}$ se dice un valor propio de una matriz $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ si existe $\boldsymbol{x} \in \mathbb{C}^n, \boldsymbol{x} \neq 0$ tal que $\boldsymbol{Ax} = \lambda \boldsymbol{x}$. En tal caso, el vector $\boldsymbol{x}$ se llama vector propio de $\boldsymbol{A}$ asociado a λ.*

Definición 2.8 *Sea $\lambda \in \mathbb{C}$ un valor propio de $\boldsymbol{A}$. Se llama espacio propio asociado a λ al conjunto $L(\lambda) := \{\boldsymbol{x} \in \mathbb{C}^n \mid \boldsymbol{Ax} = \lambda \boldsymbol{x}\}$.*

Note que $L(\lambda)$ contiene, además del vector nulo, a todos los vectores propios asociados a λ. Se puede demostrar que $L(\lambda)$ es un subespacio vectorial de $\mathbb{C}^n$ con dimensión $\varrho(\lambda)$ dada por $\varrho(\lambda) = n - \operatorname{rango}(\boldsymbol{A} - \lambda \boldsymbol{I})$. El número $\varrho(\lambda)$ se llama también multiplicidad geométrica de λ.

Lema 2.2 *Sea $\boldsymbol{A} \in \mathbb{C}^{n\times n}$. Un escalar $\lambda \in \mathbb{C}$ es un valor propio de $\boldsymbol{A}$ si y sólo si*

$$\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0.$$

Demostración: *De acuerdo a la Definición 2.7, un escalar $\lambda \in \mathbb{C}$ es un valor propio de $\boldsymbol{A}$ si y sólo si existe $\boldsymbol{x} \in \mathbb{C}^n, \boldsymbol{x} \neq 0$ tal que $\boldsymbol{Ax} = \lambda \boldsymbol{x}$, equivalentemente, si y sólo si $(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{x} = 0$ con $\boldsymbol{x} \neq 0$. Esta última relación es un sistema lineal de ecuaciones lineales homogéneo de n ecuaciones y n incógnitas. Para no obtener únicamente la solución trivial $\boldsymbol{x} = 0$, que no nos interesa, imponemos la condición necesaria y suficiente* $\det(\boldsymbol{A} - \lambda \boldsymbol{I}) = 0$. ∎

La expresión $f_{\boldsymbol{A}}(\lambda) := \det(\boldsymbol{A} - \lambda \boldsymbol{I})$ es un polinomio de grado n que se llama polinomio característico de $\boldsymbol{A}$, y tiene la forma $f_{\boldsymbol{A}}(\lambda) = (-1)^n \left(\lambda^n + \alpha_{n-1}\lambda^{n-1} + \cdots + \alpha_1 \lambda + \alpha_0\right)$. Si $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los ceros del polinomio característico, entonces $f_{\boldsymbol{A}}(\lambda)$ puede factorizarse como

$$f_{\boldsymbol{A}}(\lambda) = (-1)^n (\lambda - \lambda_1)^{\beta_1} (\lambda - \lambda_2)^{\beta_2} \cdot \cdots \cdot (\lambda - \lambda_k)^{\beta_k}$$

donde $\beta_1, \ldots, \beta_k$ son números naturales tales que $\beta_1 + \cdots + \beta_k = n$. El número $\beta_i, i = 1, \ldots, k$ de veces que se repite el factor $(\lambda - \lambda_i)$ se llama multiplicidad algebraica de λ_i. Al valor propio λ_i pueden corresponder a lo más β_i vectores propios linealmente independientes. El número de vectores propios de $\boldsymbol{A}$ asociados al valor propio λ_i, y que son linealmente independientes, es igual a $\varrho(\lambda_i)$. En otras palabras, se tiene que $\varrho(\lambda_i) \leqslant \beta_i$ para $i = 1, \ldots, k$.

Ejemplo 2.3 *La matriz diagonal de orden n, $\boldsymbol{D} := \mu \boldsymbol{I}$ con $\mu \in \mathbb{C}$, tiene el polinomio característico $f_{\boldsymbol{D}}(\lambda) = (\mu - \lambda)^n$. Luego $\lambda = \mu$, único valor propio de $\boldsymbol{D}$, tiene multiplicidad algebraica n y multiplicidad geométrica n. Esto indica que $L(\mu) = \mathbb{C}^n$, es decir, todo vector $\boldsymbol{x} \in \mathbb{C}^n, \boldsymbol{x} \neq 0$ es vector propio de $\boldsymbol{D}$ asociado a μ.*

Ejemplo 2.4 *Consideremos la matriz* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *dada por*

$$\boldsymbol{A} = \begin{bmatrix} \mu & 1 & 0 & \dots & 0 \\ 0 & \mu & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \mu & 1 \\ 0 & \dots & \dots & 0 & \mu \end{bmatrix}$$

con $\mu \in \mathbb{C}$, *que tiene el mismo polinmomio característico que la matriz del ejemplo anterior, en efecto,*

$$f_{\boldsymbol{A}}(\lambda) = \begin{vmatrix} \mu-\lambda & 1 & 0 & \cdots & 0 \\ 0 & \mu-\lambda & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ \vdots & & \ddots & \mu-\lambda & 1 \\ 0 & \cdots & \cdots & 0 & \mu-\lambda \end{vmatrix} = (\mu-\lambda)^n$$

En este caso, el único valor propio de $\boldsymbol{A}, \lambda = \mu$, *tiene multiplicidad algebraica n, mientras que su multiplicidad geométrica es* $\varrho(\mu) = 1$. *En efecto, de la Definición 2.8 tenemos*

$$\begin{aligned} L(\mu) &= \{\boldsymbol{x} \in \mathbb{C}^n \mid (\boldsymbol{A} - \mu \boldsymbol{I})\boldsymbol{x} = 0\} \\ &= \left\{ \boldsymbol{x} \in \mathbb{C}^n \;\middle|\; \begin{bmatrix} 0 & 1 & \dots & & 0 \\ & 0 & 1 & & \\ & & \ddots & \ddots & \\ & & & 0 & 1 \\ 0 & & & & 0 \end{bmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_{n-1} \\ x_n \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ \vdots \\ 0 \\ 0 \end{pmatrix} \right\}, \end{aligned}$$

esto es,

$$L(\mu) = \{\boldsymbol{x} \in \mathbb{C}^n \mid x_2 = 0, \dots, x_n = 0\} = \left\{\boldsymbol{x} = (x_1, 0, \dots, 0)^{\boldsymbol{T}} \mid x_1 \in \mathbb{C}\right\}$$

Lo anterior muestra que $\varrho(\mu) = \dim L(\mu) = 1$.

Ejercicio 2.1 *Si* $\boldsymbol{A}$ *es una matriz cuadrada, muestre que* $\boldsymbol{A}$ *y* $\boldsymbol{A}^T$ *tienen el mismo polinomio característico, y por lo tanto, los mismos valores propios.*

Definición 2.9 *Sean* $\boldsymbol{A}$ *y* $\boldsymbol{B} \in \mathbb{C}^{n\times n}$. *Las matrices* $\boldsymbol{A}$ *y* $\boldsymbol{B}$ *se dicen similares si existe una matriz* $\boldsymbol{P} \in \mathbb{C}^{n\times n}$ *invertible tal que*

$$\boldsymbol{B} = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}. \tag{2.2.1}$$

Lema 2.3 *Sean* $\boldsymbol{A}$ *y* $\boldsymbol{B} \in \mathbb{C}^{n\times n}$. *Si* $\boldsymbol{A}$ *y* $\boldsymbol{B}$ *son similares, entonces ellas tienen los mismos n valores propios, contando su multiplicidad algebraica. Además, si* $\boldsymbol{x}$ *es un vector propio de* $\boldsymbol{A}$, *entonces*

$P^{-1}x$ es vector propio de B, con P que satisface (2.2.1).

Demostración: *Como A y B son similares, existe P invertible tal que $B = P^{-1}AP$. De lo anterior se deduce que $A = PBP^{-1}$ y luego*

$$f_A(\lambda) = \det(A - \lambda I) = \det\left(PBP^{-1} - \lambda PP^{-1}\right) = \det\left(P(B - \lambda I)P^{-1}\right).$$

Puesto que

$$\forall A, B \in \mathbb{C}^{n\times n}: \quad \det(AB) = \det(A)\det(B)$$

entonces $\det(P)\det\left(P^{-1}\right) = \det\left(PP^{-1}\right) = \det(I) = 1$ *y en consecuencia,*

$$f_A(\lambda) = \det(P)\det(B - \lambda I)\det\left(P^{-1}\right) = \det(B - \lambda I) = f_B(\lambda)$$

Eso muestra que A y B tienen el mismo polinomio característico y por lo tanto los mismos n valores propios, contando su multiplicidad algebraica.

Consideremos ahora un valor propio λ de A y un vector propio x asociado a λ. Multiplicando a la izquierda por P^{-1} la ecuación $Ax = \lambda x$ obtenemos

$$P^{-1}Ax = \lambda\left(P^{-1}x\right). \tag{2.2.2}$$

Por otra parte, $P^{-1}Ax = P^{-1}A\left(PP^{-1}\right)x = \left(P^{-1}AP\right)\left(P^{-1}x\right)$, lo cual, en virtud de la igualdad $B = P^{-1}AP$, conduce a

$$P^{-1}Ax = B\left(P^{-1}x\right). \tag{2.2.3}$$

Se sigue de las igualdades (2.2.2) y (2.2.3) que $B\left(P^{-1}x\right) = \lambda\left(P^{-1}x\right)$. Notando que $P^{-1}x \neq 0$, concluimos que $P^{-1}x$ es un vector propio de B asociado al valor propio λ. ■

Definición 2.10 (Vectores ortogonales y ortonormales en $\mathbb{C}^n$) *Al igual que en el caso real, un conjunto de vectores no nulos $\{z_1, z_2, \ldots, z_m\}$ en $\mathbb{C}^n$ se llama ortogonal si $\langle z_i, z_j\rangle = 0$ siempre que $i \neq j$, y es ortonormal si, además, $\|z_i\| = 1$ para cada i.*

Teorema 2.8 *Las siguientes condiciones son equivalentes para una matriz $P \in \mathbb{R}^{n\times n}$.*

1. *P es invertible y $P^{-1} = P^T$.*
2. *Las filas de P son ortonormales.*
3. *Las columnas de P son ortonormales.*

Demostración: *Primero recordemos que la condición (1) es equivalente a $PP^T = I$. Sea $x_1, x_2, \ldots, x_n$ las filas de P. Entonces x_j^T es la columna j de P^T, por lo que la entrada (i,j) de PP^T es $x_i \cdot x_j$. Por lo tanto, $PP^T = I$ significa que $x_i \cdot x_j = 0$ si $i \neq j$ y $x_i \cdot x_j = 1$ si $i = j$. Por lo tanto, la condición (1) es equivalente a (2). La demostración de la equivalencia entre (1) y (3) es similar.*

Teorema 2.9 *Las siguientes afirmaciones son equivalentes para una matriz $A \in \mathbb{C}^{n\times n}$.*

1. *A es invertible y $A^{-1} = A^*$.*

2. Las filas de $\boldsymbol{A}$ son un conjunto ortonormal en $\mathbb{C}^n$.

3. Las columnas de $\boldsymbol{A}$ son un conjunto ortonormal en $\mathbb{C}^n$.

Demostración: *Si* $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \cdots & \boldsymbol{c}_n \end{bmatrix}$ *es una matriz compleja con la columna j-ésima $\boldsymbol{c}_j$, entonces* $\boldsymbol{A}^T \bar{\boldsymbol{A}} = [\langle \boldsymbol{c}_i, \boldsymbol{c}_j \rangle]$*, como en el Teorema 2.8. Ahora* $(1) \Leftrightarrow (2)$ *se tiene, y (1) ⇔ (3) se demuestra de la misma manera.*

2.3 Diagonalización de Matrices

El mundo está lleno de ejemplos de sistemas que evolucionan con el tiempo: el clima en una región, la economía de una nación, la diversidad de un ecosistema, etc. Describir estos sistemas es difícil en general, y se han desarrollado varios métodos en casos especiales. En esta sección describiremos uno de estos métodos, llamado diagonalización, que es una de las técnicas más importantes en álgebra lineal. Un ejemplo de este procedimiento es en la modelización del crecimiento de la población de una especie animal. Esto ha atraído más atención en los últimos años con la creciente conciencia de que muchas especies están en peligro de extinción. Para motivar la técnica, comenzamos estableciendo un modelo simple de una población de aves en el que hacemos suposiciones sobre las tasas de supervivencia y reproducción.

Ejemplo 2.5 *Considera la evolución de la población de una especie de aves. Debido a que el número de machos y hembras es casi igual, contamos solo a las hembras. Suponemos que cada hembra permanece como juvenil durante un año y luego se convierte en adulta, y que solo las adultas tienen descendencia. Hacemos tres suposiciones sobre las tasas de reproducción y supervivencia:*

1. *El número de hembras juveniles que nacen en cualquier año es el doble del número de hembras adultas vivas el año anterior (decimos que la tasa de reproducción es 2).*

2. *La mitad de las hembras adultas en cualquier año sobreviven al siguiente año (la tasa de supervivencia de adultos es $\frac{1}{2}$).*

3. *Un cuarto de las hembras juveniles en cualquier año sobreviven hasta la edad adulta (la tasa de supervivencia de juveniles es $\frac{1}{4}$).*

Si inicialmente había 100 hembras adultas y 40 hembras juveniles vivas, calcula la población de hembras k años después.

Sol: *Sea a_k y j_k denotan, respectivamente, el número de hembras adultas y juveniles después de k años, de modo que la población total de hembras es la suma $a_k + j_k$. La suposición 1 muestra que $j_{k+1} = 2a_k$, mientras que las suposiciones 2 y 3 muestran que $a_{k+1} = \frac{1}{2}a_k + \frac{1}{4}j_k$. Por lo tanto, los números a_k y j_k en años sucesivos están relacionados por las siguientes ecuaciones:*

$$\begin{aligned} a_{k+1} &= \frac{1}{2}a_k + \frac{1}{4}j_k \\ j_{k+1} &= 2a_k \end{aligned}$$

Si escribimos $\boldsymbol{v}_k = \begin{bmatrix} a_k \\ j_k \end{bmatrix}$ *y* $A = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 2 & 0 \end{bmatrix}$ *estas ecuaciones toman la forma matricial*

$$\boldsymbol{v}_{k+1} = A\boldsymbol{v}_k, \text{ para cada } k = 0, 1, 2, \ldots$$

Tomando $k = 0$ *obtenemos* $\boldsymbol{v}_1 = A\boldsymbol{v}_0$*, luego tomando* $k = 1$ *obtenemos* $\boldsymbol{v}_2 = A\boldsymbol{v}_1 = A^2\boldsymbol{v}_0$*, y tomando* $k = 2$ *obtenemos* $\boldsymbol{v}_3 = A\boldsymbol{v}_2 = A^3\boldsymbol{v}_0$*. Continuando de esta manera, obtenemos*

$$\boldsymbol{v}_k = A^k\boldsymbol{v}_0, \text{ para cada } k = 0, 1, 2, \ldots$$

Dado que $\boldsymbol{v}_0 = \begin{bmatrix} a_0 \\ j_0 \end{bmatrix} = \begin{bmatrix} 100 \\ 40 \end{bmatrix}$ *es conocido, encontrar el perfil de población* $\boldsymbol{v}_k$ *implica calcular* A^k *para todo* $k \geq 0$*. Completaremos este cálculo después de desarrollar algunas nuevas técnicas.*

Sea A una matriz fija $n \times n$. Una secuencia $\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ de vectores columna en $\mathbb{R}^n$ se llama un sistema dinámico lineal[1] si $\boldsymbol{v}_0$ es conocido y los otros $\boldsymbol{v}_k$ están determinados por las condiciones

$$\boldsymbol{v}_{k+1} = A\boldsymbol{v}_k \text{ para cada } k = 0, 1, 2, \ldots$$

Estas condiciones se llaman una recurrencia matricial para los vectores $\boldsymbol{v}_k$. Como en el Ejemplo 2.5, implican que

$$\boldsymbol{v}_k = A^k\boldsymbol{v}_0 \text{ para todo } k \geq 0$$

por lo que encontrar las columnas $\boldsymbol{v}_k$ implica calcular A^k para $k \geq 0$.

El cálculo directo de las potencias A^k de una matriz cuadrada A puede ser laborioso, por lo que adoptamos un método indirecto que se usa comúnmente. La idea es primero diagonalizar la matriz A, es decir, encontrar una matriz invertible P tal que

$$P^{-1}AP = D \text{ es una matriz diagonal} \tag{2.3.1}$$

Esto funciona porque las potencias D^k de la matriz diagonal D son fáciles de calcular, y la ecuación (2.3.1) nos permite calcular las potencias A^k de la matriz A en términos de las potencias D^k de D. De hecho, podemos resolver la cuación (2.3.1) para A obteniendo $A = PDP^{-1}$. Elevando al cuadrado obtenemos

$$A^2 = (PDP^{-1})(PDP^{-1}) = PD^2P^{-1}$$

Usando esto podemos calcular A^3 de la siguiente manera:

$$A^3 = AA^2 = (PDP^{-1})(PD^2P^{-1}) = PD^3P^{-1}.$$

Continuando de esta manera, obtenemos el Teorema 2.10 (incluso si $\boldsymbol{D}$ no es diagonal).

Teorema 2.10 *Si* $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$*, entonces* $\boldsymbol{A}^k = \boldsymbol{P}\boldsymbol{D}^k\boldsymbol{P}^{-1}$ *para cada* $k = 1, 2, \ldots$*.*

Por lo tanto, calcular $\boldsymbol{A}^k$ se reduce a encontrar una matriz invertible $\boldsymbol{P}$ como en la ecuación (2.3.1). Para hacer esto, es necesario primero calcular ciertos números (llamados valores propios) asociados con la matriz $\boldsymbol{A}$.

[1] Más precisamente, este es un sistema dinámico lineal discreto. Muchos modelos consideran $\boldsymbol{v}_t$ como una función continua del tiempo t, y reemplazan nuestra condición entre $\boldsymbol{v}_{k+1}$ y $A\boldsymbol{v}_k$ con una relación diferencial vista como función del tiempo.

Definición 2.11 (Matrices Diagonalizables) *Una matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ se llama diagonalizable si*

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} \text{ es diagonal para alguna matriz invertible } \boldsymbol{P} \text{ de tamaño } n \times n$$

Aquí, la matriz invertible $\boldsymbol{P}$ se llama una matriz diagonalizadora para A.

Para descubrir cuándo existe tal matriz $\boldsymbol{P}$, denotamos las columnas de $\boldsymbol{P}$ como $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ y buscamos formas de determinar cuándo existen tales $\boldsymbol{x}_i$ y cómo calcularlos. Con este fin, escribimos $\boldsymbol{P}$ en términos de sus columnas de la siguiente manera:

$$\boldsymbol{P} = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]$$

Observamos que $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$ para alguna matriz diagonal $\boldsymbol{D}$ se cumple si y sólo si

$$\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{D}.$$

Si escribimos $\boldsymbol{D} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$, donde los λ_i son números a determinar, la ecuación $\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}\boldsymbol{D}$ se convierte en

$$\boldsymbol{A}[\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n] = [\boldsymbol{x}_1, \boldsymbol{x}_2, \cdots, \boldsymbol{x}_n]\begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}.$$

Por la definición de multiplicación de matrices, cada lado se simplifica de la siguiente manera

$$\begin{bmatrix} \boldsymbol{A}\boldsymbol{x}_1 & \boldsymbol{A}\boldsymbol{x}_2 & \cdots & \boldsymbol{A}\boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \lambda_1\boldsymbol{x}_1 & \lambda_2\boldsymbol{x}_2 & \cdots & \lambda_n\boldsymbol{x}_n \end{bmatrix}$$

Comparando columnas, se muestra que $\boldsymbol{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i$ para cada i, entonces

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D} \quad \text{si y sólo si } A\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i \text{ para cada } i$$

En otras palabras, $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$ se cumple si y sólo si las entradas diagonales de $\boldsymbol{D}$ son los valores propios de A y las columnas de $\boldsymbol{P}$ son los vectores propios correspondientes. Esto demuestra el siguiente resultado fundamental.

Teorema 2.11 *Sea $\boldsymbol{A} \in \mathbb{R}^{n\times n}$.*

1. *$\boldsymbol{A}$ es diagonalizable si y sólo si tiene vectores propios $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ tales que la matriz $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \end{bmatrix}$ es invertible.*

2. *Cuando esto ocurre, $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_n)$ donde, para cada i, λ_i es el valor propio de $\boldsymbol{A}$ correspondiente a $\boldsymbol{x}_i$.*

Ejemplo 2.6 *Diagonalice la matriz* $A = \begin{bmatrix} 2 & 0 & 0 \\ 1 & 2 & -1 \\ 1 & 3 & -2 \end{bmatrix}$.

Sol: *Los valores propios de A son $\lambda_1 = 2$, $\lambda_2 = 1$, y $\lambda_3 = -1$, con los vectores propios básicos correspondientes* $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$, $\boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 1 \\ 1 \end{bmatrix}$, *y* $\boldsymbol{x}_3 = \begin{bmatrix} 0 \\ 1 \\ 3 \end{bmatrix}$ *respectivamente (verificar!). Dado que la matriz*

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 1 & 1 \\ 1 & 1 & 3 \end{bmatrix}$$

es invertible, el Teorema 2.11 garantiza que

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{bmatrix} = \begin{bmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} = \boldsymbol{D}$$

El lector puede verificar esto directamente, es más fácil comprobar $\boldsymbol{AP} = \boldsymbol{PD}$.

En el ejemplo anterior, cada valor propio ha tenido solo un vector propio básico. El siguiente ejemplo muestra una matriz diagonalizable donde esto no es el caso.

Ejemplo 2.7 *Diagonalice la matriz* $\boldsymbol{A} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}$.

Sol: *Para calcular el polinomio característico de $\boldsymbol{A}$, primero suma las filas 2 y 3 de $x\boldsymbol{I} - \boldsymbol{A}$ (o $\boldsymbol{A} - x\boldsymbol{I}$) a la fila 1:*

$$f_{\boldsymbol{A}}(x) = \det \begin{bmatrix} x & -1 & -1 \\ -1 & x & -1 \\ -1 & -1 & x \end{bmatrix} = \det \begin{bmatrix} x-2 & x-2 & x-2 \\ -1 & x & -1 \\ -1 & -1 & x \end{bmatrix}$$

$$= \det \begin{bmatrix} x-2 & 0 & 0 \\ -1 & x+1 & 0 \\ -1 & 0 & x+1 \end{bmatrix} = (x-2)(x+1)^2$$

Por lo tanto, los valores propios son $\lambda_1 = 2$ y $\lambda_2 = -1$, con λ_2 repetido dos veces (decimos que λ_2 tiene multiplicidad dos). Sin embargo, $\boldsymbol{A}$ es diagonalizable. Para $\lambda_1 = 2$, el sistema de ecuaciones $(\lambda_1\boldsymbol{I} - \boldsymbol{A})\,\boldsymbol{x} = \boldsymbol{0}$ tiene solución general $\boldsymbol{x} = t \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$ *como el lector puede verificar, por lo que un vector propio básico λ_1 es* $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}$.

Pasando al valor propio repetido $\lambda_2 = -1$, debemos resolver $(\lambda_2\boldsymbol{I} - \boldsymbol{A})\,\boldsymbol{x} = \boldsymbol{0}$. Por eliminación gaussiana, la solución general es $\boldsymbol{x} = s \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix} + t \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$ *donde s y t son arbitrarios. Por lo tanto,*

V. Osores

el algoritmo gaussiano produce dos vectores propios básicos λ_2: $\boldsymbol{x}_2 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}$ *y* $\boldsymbol{y}_2 = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix}$. *Si tomamos* $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{y}_2 \end{bmatrix} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}$ *encontramos que* P *es invertible. Por lo tanto,* $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \operatorname{diag}(2, -1, -1)$ *según el Teorema 2.11.*

Definición 2.12 (Multiplicidad de un valor propio) *Un valor propio* λ *de una matriz cuadrada* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *se dice que tiene multiplicidad* m *si ocurre* m *veces como raíz del polinomio característico* $f_{\boldsymbol{A}}(x)$.

Teorema 2.12 *Una matriz* $A \in \mathbb{R}^{n\times n}$ *es diagonalizable si y sólo si cada valor propio* λ *de multiplicidad* m *produce exactamente* m *vectores propios básicos; es decir, si y sólo si la solución general del sistema* $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ *tiene exactamente* m *parámetros.*

Teorema 2.13 *Una matriz* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *con* n *valores propios distintos es diagonalizable.*

Las demostraciones del Teorema 2.12 y el Teorema 2.13 requieren técnicas más avanzadas.

Definición 2.13 (Algoritmo de Diagonalización) *Para diagonalizar una matriz* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$:

Paso 1. *Encuentra los valores propios* λ *distintos de* A.

Paso 2. *Calcula un conjunto de vectores propios básicos correspondientes a cada uno de estos valores propios* λ *como soluciones básicas del sistema homogéneo* $(\lambda \boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$.

Paso 3. *La matriz* $\boldsymbol{A}$ *es diagonalizable si y sólo si hay* n *vectores propios básicos en total.*

Paso 4. *Si* $\boldsymbol{A}$ *es diagonalizable, la matriz* $n\times n$ $\boldsymbol{P}$ *con estos vectores propios básicos como sus columnas es una matriz diagonalizante para* $\boldsymbol{A}$, *es decir,* $\boldsymbol{P}$ *es invertible y* $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ *es diagonal.*

El algoritmo de diagonalización es válido incluso si los valores propios son números complejos no reales. En este caso, los vectores propios también tendrán entradas complejas, pero no perseguiremos esto aquí.

Ejemplo 2.8 *Muestre que* $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$ *no es diagonalizable.*

Sol: 1. *El polinomio característico es* $f_{\boldsymbol{A}}(x) = (x-1)^2$, *por lo que* $\boldsymbol{A}$ *tiene solo un valor propio* $\lambda_1 = 1$ *de multiplicidad 2. Pero el sistema de ecuaciones* $(\lambda_1 \boldsymbol{I} - \boldsymbol{A})\,\boldsymbol{x} = \boldsymbol{0}$ *tiene una solución general* $t\begin{bmatrix} 1 \\ 0 \end{bmatrix}$, *por lo que hay solo un parámetro, y por lo tanto solo un vector propio básico* $\begin{bmatrix} 1 \\ 2 \end{bmatrix}$. *Por lo tanto,* $\boldsymbol{A}$ *no es diagonalizable.*

Sol: 2. *Tenemos* $f_A(x) = (x-1)^2$ *así que el único valor propio de* $\boldsymbol{A}$ *es* $\lambda = 1$. *Por lo tanto, si* $\boldsymbol{A}$ *fuera diagonalizable, el Teorema 2.11 daría* $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = \boldsymbol{I}$ *para alguna matriz invertible* $\boldsymbol{P}$. *Pero entonces* $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{I}\boldsymbol{P}^{-1} = \boldsymbol{I}$, *lo cual no es el caso. Entonces,* $\boldsymbol{A}$ *no puede ser diagonalizable.*

Las matrices diagonalizables comparten muchas propiedades de sus valores propios. El siguiente ejemplo ilustra por qué.

Ejemplo 2.9 *Si* $\lambda^3 = 5\lambda$ *para cada valor propio de la matriz diagonalizable* $\boldsymbol{A}$, *muestre que* $\boldsymbol{A}^3 = 5\boldsymbol{A}$.

Sol: *Sea* $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D} = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. *Debido a que* $\lambda_i^3 = 5\lambda_i$ *para cada* i, *obtenemos*

$$\boldsymbol{D}^3 = \operatorname{diag}\left(\lambda_1^3, \ldots, \lambda_n^3\right) = \operatorname{diag}(5\lambda_1, \ldots, 5\lambda_n) = 5\boldsymbol{D}.$$

Por lo tanto, $\boldsymbol{A}^3 = \left(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}\right)^3 = \boldsymbol{P}\boldsymbol{D}^3\boldsymbol{P}^{-1} = \boldsymbol{P}(5\boldsymbol{D})\boldsymbol{P}^{-1} = 5\left(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}\right) = 5\boldsymbol{A}$ *usando el Teorema 2.10. Esto es lo que queríamos.*

Si $p(x)$ es cualquier polinomio y $p(\lambda) = 0$ para cada valor propio de la matriz diagonalizable $\boldsymbol{A}$, un argumento similar al del ejemplo 2.9 muestra que $p(\boldsymbol{A}) = 0$. Así que ejemplo 2.9 trata el caso $p(x) = x^3 - 5x$. En general, $p(\boldsymbol{A})$ se llama la evaluación del polinomio $p(x)$ en la matriz A. Por ejemplo, si $p(x) = 2x^3 - 3x + 5$, entonces $p(\boldsymbol{A}) = 2A^3 - 3A + 5I$, note el uso de la matriz identidad.

En particular, si $f_{\boldsymbol{A}}(x)$ denota el polinomio característico de $\boldsymbol{A}$, ciertamente tenemos $f_{\boldsymbol{A}}(\lambda) = 0$ para cada valor propio λ de $\boldsymbol{A}$. Por lo tanto, $f_{\boldsymbol{A}}(\boldsymbol{A}) = 0$ para cada matriz diagonalizable $\boldsymbol{A}$. Esto es, de hecho, cierto para cualquier matriz cuadrada, diagonalizable o no, y el resultado general se llama el teorema de Cayley-Hamilton. Se introduce mas tarde.

Ejercicio 2.2 *Muestre que los valores propios de una matriz triangular son las entradas de la diagonal principal.*

Comenzamos la sección con un ejemplo de ecología que modela la evolución de la población de una especie de aves a medida que pasa el tiempo. Como prometimos, ahora completamos el ejemplo.

La población de aves se describió calculando el perfil de población de hembras $\boldsymbol{v}_k = \begin{bmatrix} a_k \\ j_k \end{bmatrix}$ de la especie, donde a_k y j_k representan el número de hembras adultas y juveniles presentes k años después de observar los valores iniciales a_0 y j_0. El modelo asume que estos números están relacionados por las siguientes ecuaciones:

$$a_{k+1} = \frac{1}{2}a_k + \frac{1}{4}j_k$$
$$j_{k+1} = 2a_k$$

V. Osores

Si escribimos $\boldsymbol{A} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 2 & 0 \end{bmatrix}$, las columnas $\boldsymbol{v}_k$ satisfacen $\boldsymbol{v}_{k+1} = \boldsymbol{A}\boldsymbol{v}_k$ para cada $k = 0, 1, 2, \ldots$. Por lo tanto, $\boldsymbol{v}_k = \boldsymbol{A}^k\boldsymbol{v}_0$ para cada $k = 1, 2, \ldots$ Ahora podemos usar nuestras técnicas de diagonalización para determinar el perfil de población $\boldsymbol{v}_k$ para todos los valores de k en términos de los valores iniciales.

Ejemplo 2.10 *Suponiendo que los valores iniciales fueron $a_0 = 100$ hembras adultas y $j_0 = 40$ hembras juveniles, calcula a_k y j_k para $k = 1, 2, \ldots$.*

Sol: *El polinomio característico de la matriz $\boldsymbol{A} = \begin{bmatrix} \frac{1}{2} & \frac{1}{4} \\ 2 & 0 \end{bmatrix}$ es $f_A(x) = x^2 - \frac{1}{2}x - \frac{1}{2} = (x-1)\left(x+\frac{1}{2}\right)$, por lo que los eigenvalores son $\lambda_1 = 1$ y $\lambda_2 = -\frac{1}{2}$ y la eliminación gaussiana da como resultado los eigenvectores básicos correspondientes $\begin{bmatrix} \frac{1}{2} \\ 1 \end{bmatrix}$ y $\begin{bmatrix} -\frac{1}{4} \\ 1 \end{bmatrix}$. Para mayor comodidad, podemos usar los múltiplos $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$ y $\boldsymbol{x}_2 = \begin{bmatrix} -1 \\ 4 \end{bmatrix}$ respectivamente. Por lo tanto, una matriz diagonalizadora es $\boldsymbol{P} = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}$ y obtenemos*

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D} \text{ donde } \boldsymbol{D} = \begin{bmatrix} 1 & 0 \\ 0 & -\frac{1}{2} \end{bmatrix}$$

Esto da $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$, así que, para cada $k \geq 0$, podemos calcular $\boldsymbol{A}^k$ explícitamente:

$$\begin{aligned} \boldsymbol{A}^k = \boldsymbol{P}\boldsymbol{D}^k\boldsymbol{P}^{-1} &= \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \left(-\frac{1}{2}\right)^k \end{bmatrix} \frac{1}{6} \begin{bmatrix} 4 & 1 \\ -2 & 1 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 4 + 2\left(-\frac{1}{2}\right)^k & 1 - \left(-\frac{1}{2}\right)^k \\ 8 - 8\left(-\frac{1}{2}\right)^k & 2 + 4\left(-\frac{1}{2}\right)^k \end{bmatrix} \end{aligned}$$

Por lo tanto, obtenemos

$$\begin{aligned} \begin{bmatrix} a_k \\ j_k \end{bmatrix} = \boldsymbol{v}_k = \boldsymbol{A}^k\boldsymbol{v}_0 &= \frac{1}{6} \begin{bmatrix} 4 + 2\left(-\frac{1}{2}\right)^k & 1 - \left(-\frac{1}{2}\right)^k \\ 8 - 8\left(-\frac{1}{2}\right)^k & 2 + 4\left(-\frac{1}{2}\right)^k \end{bmatrix} \begin{bmatrix} 100 \\ 40 \end{bmatrix} \\ &= \frac{1}{6} \begin{bmatrix} 440 + 160\left(-\frac{1}{2}\right)^k \\ 880 - 640\left(-\frac{1}{2}\right)^k \end{bmatrix} \end{aligned}$$

Igualando las entradas superiores e inferiores, obtenemos fórmulas exactas para a_k y j_k:

$$a_k = \frac{220}{3} + \frac{80}{3}\left(-\frac{1}{2}\right)^k \text{ y } j_k = \frac{440}{3} + \frac{320}{3}\left(-\frac{1}{2}\right)^k \text{ para } k = 1, 2, \cdots.$$

En la práctica, generalmente no se requieren los valores exactos de a_k y j_k. Lo que se necesita es una medida de cómo se comportan estos números para valores grandes de k. Esto es fácil de obtener aquí. Dado que $\left(-\frac{1}{2}\right)^k$ es casi cero para k grande, obtenemos los siguientes valores aproximados

$$a_k \approx \frac{220}{3} \text{ y } j_k \approx \frac{440}{3} \text{ si } k \text{ es grande}$$

Por lo tanto, a largo plazo, la población de hembras se estabiliza con aproximadamente el doble de juveniles que de adultas.

2.3.1 Aplicación a sistemas lineales dinámicos

Definición 2.14 (Sistema lineal dinámico) *Si $\boldsymbol{A}$ es una matriz $n \times n$, una secuencia $\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ de columnas en $\mathbb{R}^n$ se llama un sistema dinámico lineal si $\boldsymbol{v}_0$ está dado (conocido) y $\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ están dados por la recurrencia matricial $\boldsymbol{v}_{k+1} = \boldsymbol{A}\boldsymbol{v}_k$ para cada $k \geq 0$. Llamamos a $\boldsymbol{A}$ la matriz de migración del sistema.*

Como antes, tenemos $\boldsymbol{v}_1 = \boldsymbol{A}\boldsymbol{v}_0$, luego $\boldsymbol{v}_2 = \boldsymbol{A}\boldsymbol{v}_1 = \boldsymbol{A}^2\boldsymbol{v}_0$, y continuando encontramos

$$\boldsymbol{v}_k = \boldsymbol{A}^k \boldsymbol{v}_0 \text{ para cada } k = 1, 2, \cdots . \tag{2.3.2}$$

Por lo tanto, las columnas $\boldsymbol{v}_k$ están determinadas por las potencias $\boldsymbol{A}^k$ de la matriz $\boldsymbol{A}$ y, como hemos visto, estas potencias pueden ser calculadas eficientemente si $\boldsymbol{A}$ es diagonalizable. De hecho, la ecuación (2.3.2) puede usarse para dar una "fórmula" elegante para las columnas $\boldsymbol{v}_k$ en este caso.

Supongamos que $\boldsymbol{A}$ es diagonalizable con valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$ y vectores propios básicos correspondientes $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$. Si $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \end{bmatrix}$ es una matriz diagonalizante con los $\boldsymbol{x}_i$ como columnas, entonces P es invertible y

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D} = \operatorname{diag}(\lambda_1, \lambda_2, \cdots, \lambda_n)$$

por el Teorema 2.11. Por lo tanto, $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}$ así que la ecuación (2.3.2) y el Teorema 2.10 dan

$$\boldsymbol{v}_k = \boldsymbol{A}^k \boldsymbol{v}_0 = \left(\boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^{-1}\right)^k \boldsymbol{v}_0 = \left(\boldsymbol{P}\boldsymbol{D}^k\boldsymbol{P}^{-1}\right) \boldsymbol{v}_0 = \boldsymbol{P}\boldsymbol{D}^k \left(\boldsymbol{P}^{-1}\boldsymbol{v}_0\right)$$

para cada $k = 1, 2, \ldots$. Para mayor comodidad, denotamos la columna $\boldsymbol{P}^{-1}\boldsymbol{v}_0$ que surge aquí de la siguiente manera:

$$\boldsymbol{b} = \boldsymbol{P}^{-1}\boldsymbol{v}_0 = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

V. Osores

Entonces la multiplicación matricial da

$$\begin{aligned}
\boldsymbol{v}_k &= \boldsymbol{P}\boldsymbol{D}^k\left(\boldsymbol{P}^{-1}\boldsymbol{v}_0\right) \\
&= \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} \lambda_1^k & 0 & \cdots & 0 \\ 0 & \lambda_2^k & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n^k \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} b_1\lambda_1^k \\ b_2\lambda_2^k \\ \vdots \\ b_3\lambda_n^k \end{bmatrix} \\
&= b_1\lambda_1^k\boldsymbol{x}_1 + b_2\lambda_2^k\boldsymbol{x}_2 + \cdots + b_n\lambda_n^k\boldsymbol{x}_n
\end{aligned} \tag{2.3.3}$$

para cada $k \geq 0$. Esta es una fórmula exacta útil para las columnas $\boldsymbol{v}_k$. Nótese que, en particular,

$$\boldsymbol{v}_0 = b_1\boldsymbol{x}_1 + b_2\boldsymbol{x}_2 + \cdots + b_n\boldsymbol{x}_n$$

Sin embargo, a menudo no se requiere una fórmula exacta para $\boldsymbol{v}_k$ en la práctica; lo único que se necesita es estimar $\boldsymbol{v}_k$ para valores grandes de k (como se hizo en el ejemplo 2.10). Esto puede hacerse fácilmente si $\boldsymbol{A}$ tiene un valor propio mayor. Un valor propio λ de una matriz $\boldsymbol{A}$ se llama un valor propio dominante de $\boldsymbol{A}$ si tiene multiplicidad 1 y

$$|\lambda| > |\mu| \text{ para todos los valores propios } \mu \neq \lambda$$

donde $|\lambda|$ denota el valor absoluto del número λ. Por ejemplo, $\lambda_1 = 1$ es dominante en el ejemplo 2.10.

Volviendo a la discusión anterior, supongamos que $\boldsymbol{A}$ tiene un valor propio dominante. Al elegir el orden en el que las columnas $\boldsymbol{x}_i$ se colocan en $\boldsymbol{P}$, podemos asumir que λ_1 es dominante entre los valores propios $\lambda_1, \lambda_2, \ldots, \lambda_n$ de $\boldsymbol{A}$ (ver la discusión que sigue al ejemplo 2.6). Ahora recordemos la expresión exacta para $\boldsymbol{v}_k$ en la ecuación 2.3.3 anterior:

$$\boldsymbol{v}_k = b_1\lambda_1^k\boldsymbol{x}_1 + b_2\lambda_2^k\boldsymbol{x}_2 + \cdots + b_n\lambda_n^k\boldsymbol{x}_n.$$

Toma λ_1^k como factor común en esta ecuación para obtener

$$\boldsymbol{v}_k = \lambda_1^k\left[b_1\boldsymbol{x}_1 + b_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\boldsymbol{x}_2 + \cdots + b_n\left(\frac{\lambda_n}{\lambda_1}\right)^k\boldsymbol{x}_n\right]$$

para cada $k \geq 0$. Dado que λ_1 es dominante, tenemos $|\lambda_i| < |\lambda_1|$ para cada $i \geq 2$, por lo que cada uno de los números $(\lambda_i/\lambda_1)^k$ se vuelven pequeños en valor absoluto a medida que k aumenta. Por lo tanto, $\boldsymbol{v}_k$ es aproximadamente igual al primer término $\lambda_1^k b_1\boldsymbol{x}_1$, y escribimos esto como $\boldsymbol{v}_k \approx \lambda_1^k b_1\boldsymbol{x}_1$. Estas observaciones se resumen en el siguiente teorema (junto con la fórmula exacta anterior para $\boldsymbol{v}_k$).

Teorema 2.14 *Considera el sistema dinámico $\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ con recurrencia matricial*

$$\boldsymbol{v}_{k+1} = \boldsymbol{A}\boldsymbol{v}_k \text{ para } k \geq 0$$

donde $\boldsymbol{A}$ *y* $\boldsymbol{v}_0$ *están dados. Supongamos que* $\boldsymbol{A}$ *es una matriz* $n \times n$ *diagonalizable con valores propios* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *y vectores propios básicos correspondientes* $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$, *y sea* $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \end{bmatrix}$ *la matriz diagonalizante. Entonces una fórmula exacta para* $\boldsymbol{v}_k$ *es*

$$\boldsymbol{v}_k = b_1\lambda_1^k\boldsymbol{x}_1 + b_2\lambda_2^k\boldsymbol{x}_2 + \cdots + b_n\lambda_n^k\boldsymbol{x}_n \text{ para cada } k \geq 0$$

donde los coeficientes b_i *provienen de*

$$\boldsymbol{b} = \boldsymbol{P}^{-1}\boldsymbol{v}_0 = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_n \end{bmatrix}.$$

Además, si $\boldsymbol{A}$ *tiene un valor propio dominante*[a] λ_1, *entonces* $\boldsymbol{v}_k$ *se aproxima por*

$$\boldsymbol{v}_k = b_1\lambda_1^k\boldsymbol{x}_1 \text{ para } k \text{ suficientemente grande.}$$

[a] Resultados similares se pueden encontrar en otras situaciones. Si, por ejemplo, los valores propios λ_1 y λ_2 (posiblemente iguales) satisfacen $|\lambda_1| = |\lambda_2| > |\lambda_i|$ para todo $i > 2$, entonces obtenemos $\boldsymbol{v}_k \approx b_1\lambda_1^k\boldsymbol{x}_1 + b_2\lambda_2^k\boldsymbol{x}_2$ para k grande.

Ejemplo 2.11 *Volviendo al ejemplo 2.10, vemos que* $\lambda_1 = 1$ *es el valor propio dominante, con el vector propio* $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}$. *Aquí* $\boldsymbol{P} = \begin{bmatrix} 1 & -1 \\ 2 & 4 \end{bmatrix}$ *y* $\boldsymbol{v}_0 = \begin{bmatrix} 100 \\ 40 \end{bmatrix}$, *por lo que* $\boldsymbol{P}^{-1}\boldsymbol{v}_0 = \frac{1}{3}\begin{bmatrix} 220 \\ -80 \end{bmatrix}$. *Por lo tanto,* $b_1 = \frac{220}{3}$ *en la notación del Teorema 2.14, así que*

$$\begin{bmatrix} a_k \\ j_k \end{bmatrix} = \boldsymbol{v}_k \approx b_1\lambda_1^k\boldsymbol{x}_1 = \frac{220}{3}1^k\begin{bmatrix} 1 \\ 2 \end{bmatrix}$$

donde k *es grande. Por lo tanto,* $a_k \approx \frac{220}{3}$ *y* $j_k \approx \frac{440}{3}$ *como en el ejemplo 2.10.*

El siguiente ejemplo utiliza el Teorema 2.14 para resolver una “recurrencia lineal”. Véase también la sección **??**.

Ejemplo 2.12 *Supongamos que una secuencia* $x_0, x_1, x_2, \ldots$ *está determinada por*

$$x_0 = 1, x_1 = -1, \text{ y } x_{k+2} = 2x_k - x_{k+1} \text{ para todo } k \geq 0.$$

Encuentre una fórmula para x_k *en términos de* k.

Sol: *Usando la recurrencia lineal* $x_{k+2} = 2x_k - x_{k+1}$ *repetidamente obtenemos*

$$x_2 = 2x_0 - x_1 = 3, \quad x_3 = 2x_1 - x_2 = -5, \quad x_4 = 11, \quad x_5 = -21, \ldots$$

así que los x_i *están determinados, pero no hay un patrón aparente. La idea es encontrar* $\boldsymbol{v}_k = \begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix}$ *para cada* k, *y luego recuperar* x_k *como la primera componente de* $\boldsymbol{v}_k$. *La razón por la que esto funciona* *es que la recurrencia lineal garantiza que estos* $\boldsymbol{v}_k$ *son un sistema dinámico:*

$$\boldsymbol{v}_{k+1} = \begin{bmatrix} x_{k+1} \\ x_{k+2} \end{bmatrix} = \begin{bmatrix} x_{k+1} \\ 2x_k - x_{k+1} \end{bmatrix} = \boldsymbol{A}\boldsymbol{v}_k \text{ donde } \boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}.$$

Los valores propios de $\boldsymbol{A}$ *son* $\lambda_1 = -2$ *y* $\lambda_2 = 1$ *con vectores propios* $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ -2 \end{bmatrix}$ *y* $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$*, por lo que la matriz diagonalizadora es* $\boldsymbol{P} = \begin{bmatrix} 1 & 1 \\ -2 & 1 \end{bmatrix}$*. Además,* $\boldsymbol{b} = \boldsymbol{P}^{-1}\boldsymbol{v}_0 = \frac{1}{3}\begin{bmatrix} 2 \\ 1 \end{bmatrix}$*, por lo que la fórmula exacta para* $\boldsymbol{v}_k$ *es*

$$\begin{bmatrix} x_k \\ x_{k+1} \end{bmatrix} = \boldsymbol{v}_k = b_1\lambda_1^k\boldsymbol{x}_1 + b_2\lambda_2^k\boldsymbol{x}_2 = \frac{2}{3}(-2)^k \begin{bmatrix} 1 \\ -2 \end{bmatrix} + \frac{1}{3}1^k \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

Igualando las primeras componentes obtenemos la fórmula deseada para x_k*:*

$$x_k = \frac{1}{3}\left[2(-2)^k + 1\right] \text{ para todos } k = 0, 1, 2, \ldots.$$

El lector debe verificar esto para los primeros valores de k*.*

El método de la potencia

Una de las razones iniciales del análisis del proceso de diagonalización de matrices fue poder calcular las potencias de una matriz cuadrada, y para esto necesitábamos el cálculo de los valores propios de $\boldsymbol{A}$. Aquí, nos interesa calcular eficientemente los valores propios, y no es sorprendente que el primer método que discutimos use las potencias de una matriz.

Recordemos que un valor propio λ de una matriz $\boldsymbol{A}$ de orden n se llama un valor propio dominante si λ tiene multiplicidad 1, y

$$|\lambda| > |\mu| \quad \text{para todos los valores propios } \mu \neq \lambda.$$

Cualquier vector propio correspondiente se llama un vector propio dominante de $\boldsymbol{A}$. Cuando existe tal valor propio, una técnica para encontrarlo es la siguiente: Sea $\boldsymbol{x}_0$ en $\mathbb{R}^n$ una primera aproximación a un vector propio dominante, y luego calcular aproximaciones sucesivas $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ como sigue:

$$\boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{x}_0 \quad \boldsymbol{x}_2 = \boldsymbol{A}\boldsymbol{x}_1 \quad \boldsymbol{x}_3 = \boldsymbol{A}\boldsymbol{x}_2 \quad \cdots$$

En general, definimos

$$\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k \quad \text{para cada } k \geq 0.$$

Si la primera estimación $\boldsymbol{x}_0$ es lo suficientemente buena, estos vectores $\boldsymbol{x}_n$ aproximarán al vector propio dominante. Esta técnica se llama el método de la potencia (porque $\boldsymbol{x}_k = \boldsymbol{A}^k\boldsymbol{x}_0$ para cada $k \geq 1$). Observemos que si $\boldsymbol{z}$ es cualquier vector propio correspondiente a λ, entonces

$$\frac{\boldsymbol{z}\cdot(A\boldsymbol{z})}{\|\boldsymbol{z}\|^2} = \frac{\boldsymbol{z}\cdot(\lambda\boldsymbol{z})}{\|\boldsymbol{z}\|^2} = \lambda.$$

Debido a que los vectores $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n, \ldots$ aproximan a los vectores propios dominantes, esto sugiere que definamos los cocientes de Rayleigh como sigue:

$$r_k = \frac{\boldsymbol{x}_k \cdot \boldsymbol{x}_{k+1}}{\|\boldsymbol{x}_k\|^2} \quad \text{para } k \geq 1.$$

Entonces, los números r_k aproximan al valor propio dominante λ.

Ejemplo 2.13 *Use el método de la potencia para aproximar un vector propio dominante y un valor propio de* $A = \begin{bmatrix} 1 & 1 \\ 2 & 0 \end{bmatrix}$.

Sol: *Los valores propios de* $\boldsymbol{A}$ *son 2 y -1, con vectores propios* $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *y* $\begin{bmatrix} 1 \\ -2 \end{bmatrix}$. *Tomemos* $\boldsymbol{x}_0 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *como la primera aproximación y calculemos* $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots,$ *sucesivamente, a partir de* $\boldsymbol{x}_1 = \boldsymbol{A}\boldsymbol{x}_0, \boldsymbol{x}_2 = \boldsymbol{A}\boldsymbol{x}_1, \ldots$. *El resultado es*

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} 3 \\ 2 \end{bmatrix}, \boldsymbol{x}_3 = \begin{bmatrix} 5 \\ 6 \end{bmatrix}, \boldsymbol{x}_4 = \begin{bmatrix} 11 \\ 10 \end{bmatrix}, \boldsymbol{x}_5 = \begin{bmatrix} 21 \\ 22 \end{bmatrix}, \ldots$$

Estos vectores se están aproximando a múltiplos escalares del vector propio dominante $\begin{bmatrix} 1 \\ 1 \end{bmatrix}$. *Además, los cocientes de Rayleigh son*

$$r_1 = \frac{7}{5}, r_2 = \frac{27}{13}, r_3 = \frac{115}{61}, r_4 = \frac{451}{221}, \ldots$$

y estos se están aproximando al valor propio dominante 2.

Para ver por qué funciona el método de la potencia, sea $\lambda_1, \lambda_2, \ldots, \lambda_m$ los valores propios de $\boldsymbol{A}$ con λ_1 dominante y sean $\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_m$ los vectores propios correspondientes. Lo que se requiere es que la primera aproximación $\boldsymbol{x}_0$ sea una combinación lineal de estos vectores propios:

$$\boldsymbol{x}_0 = a_1\boldsymbol{y}_1 + a_2\boldsymbol{y}_2 + \cdots + a_m\boldsymbol{y}_m \quad \text{con } a_1 \neq 0.$$

Si $k \geq 1$, el hecho de que $\boldsymbol{x}_k = \boldsymbol{A}^k\boldsymbol{x}_0$ y $\boldsymbol{A}^k\boldsymbol{y}_i = \lambda_i^k\boldsymbol{y}_i$ para cada i da

$$\boldsymbol{x}_k = a_1\lambda_1^k\boldsymbol{y}_1 + a_2\lambda_2^k\boldsymbol{y}_2 + \cdots + a_m\lambda_m^k\boldsymbol{y}_m \quad \text{para } k \geq 1$$

Por lo tanto

$$\frac{1}{\lambda_1^k}\boldsymbol{x}_k = a_1\boldsymbol{y}_1 + a_2\left(\frac{\lambda_2}{\lambda_1}\right)^k\boldsymbol{y}_2 + \cdots + a_m\left(\frac{\lambda_m}{\lambda_1}\right)^k\boldsymbol{y}_m.$$

El lado derecho se aproxima a $a_1\boldsymbol{y}_1$ a medida que k aumenta, porque λ_1 es dominante ($\left|\frac{\lambda_i}{\lambda_1}\right| < 1$ para cada $i > 1$). Debido a que $a_1 \neq 0$, esto significa que $\boldsymbol{x}_k$ aproxima al vector propio dominante $a_1\lambda_1^k\boldsymbol{y}_1$.

El método de la potencia requiere que la primera aproximación $\boldsymbol{x}_0$ sea una combinación lineal de los vectores propios. Pero incluso en este caso, el método falla si $a_1 = 0$, donde a_1 es el coeficiente del vector propio dominante. En general, la tasa de convergencia es bastante lenta si alguno de los cocientes $\left|\frac{\lambda_i}{\lambda_1}\right|$ está cerca de 1. Además, debido a que el método requiere multiplicaciones repetidas por $\boldsymbol{A}$, no se recomienda a menos que estas multiplicaciones sean fáciles de realizar (por ejemplo, si la mayoría de las entradas de $\boldsymbol{A}$ son cero).

Un método mucho mejor para aproximar los valores propios de una matriz invertible $\boldsymbol{A}$ depende de la factorización (usando el algoritmo de Gram-Schmidt) de $\boldsymbol{A}$ en la forma

$$\boldsymbol{A} = \boldsymbol{Q}\boldsymbol{R}$$

donde $\boldsymbol{Q}$ es ortogonal y $\boldsymbol{R}$ es invertible y triangular superior (factorización y algoritmo $\boldsymbol{QR}$).

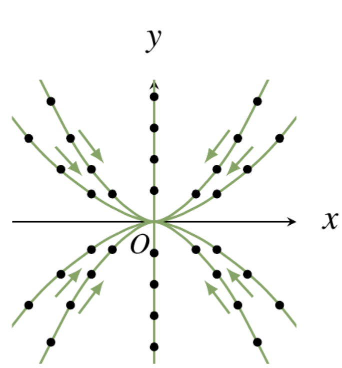
2.3.2 Descripción gráfica de sistemas dinámicos

Si se da un sistema dinámico $\boldsymbol{v}_{k+1} = \boldsymbol{A}\boldsymbol{v}_k$, la secuencia $\boldsymbol{v}_0, \boldsymbol{v}_1, \boldsymbol{v}_2, \ldots$ se llama trayectoria del sistema comenzando en $\boldsymbol{v}_0$. Es instructivo obtener un gráfico del sistema escribiendo $\boldsymbol{v}_k = \begin{bmatrix} x_k \\ y_k \end{bmatrix}$ y trazando los valores sucesivos como puntos en el plano, identificando $\boldsymbol{v}_k$ con el punto (x_k, y_k) en el plano. Damos varios ejemplos que ilustran las propiedades de los sistemas dinámicos. Para facilitar el cálculo, asumimos que la matriz $\boldsymbol{A}$ es simple, generalmente diagonal.

Ejemplo 2.14 *Sea* $\boldsymbol{A} = \begin{bmatrix} \frac{1}{2} & 0 \\ 0 & \frac{1}{3} \end{bmatrix}$. *Entonces los valores propios son* $\frac{1}{2}$ *y* $\frac{1}{3}$, *con vectores propios correspondientes* $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *y* $\boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$. *La fórmula exacta es*

$$\boldsymbol{v}_k = b_1 \left(\frac{1}{2}\right)^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b_2 \left(\frac{1}{3}\right)^k \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

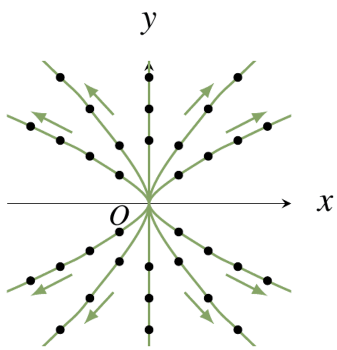
para $k = 0, 1, 2, \ldots$ *por el Teorema 2.14, donde los coeficientes* b_1 *y* b_2 *dependen del punto inicial* $\boldsymbol{v}_0$. *Varias trayectorias se trazan en el diagrama y, para cada elección de* $\boldsymbol{v}_0$, *las trayectorias convergen hacia el origen porque ambos valores propios son menores que 1 en valor absoluto. Por esta razón, el origen se llama un atractor para el sistema.*



Ejemplo 2.15 *Sea* $\boldsymbol{A} = \begin{bmatrix} \frac{3}{2} & 0 \\ 0 & \frac{4}{3} \end{bmatrix}$*. Aquí los valores propios son* $\frac{3}{2}$ *y* $\frac{4}{3}$*, con vectores propios correspondientes* $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$ *y* $\boldsymbol{x}_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$ *como antes. La fórmula exacta es*

$$\boldsymbol{v}_k = b_1 \left(\frac{3}{2}\right)^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + b_2 \left(\frac{4}{3}\right)^k \begin{bmatrix} 0 \\ 1 \end{bmatrix}$$

para $k = 0, 1, 2, \ldots$ *Dado que ambos valores propios son mayores que 1 en valor absoluto, las trayectorias divergen del origen para cada elección del punto inicial* $\boldsymbol{v}_0$*. Por esta razón, el origen se llama un repelente para el sistema.*

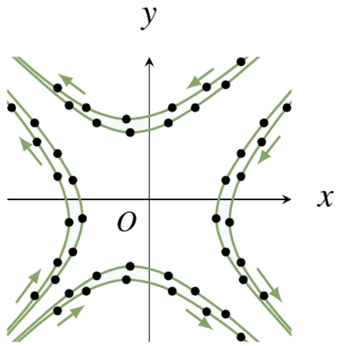


Ejemplo 2.16 *Sea* $\boldsymbol{A} = \begin{bmatrix} 1 & -\frac{1}{2} \\ -\frac{1}{2} & 1 \end{bmatrix}$*. Ahora los valores propios son* $\frac{3}{2}$ *y* $\frac{1}{2}$*, con vectores propios correspondientes* $\boldsymbol{x}_1 = \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *y* $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$*. La fórmula exacta es*

$$\boldsymbol{v}_k = b_1 \left(\frac{3}{2}\right)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix} + b_2 \left(\frac{1}{2}\right)^k \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

para $k = 0, 1, 2, \ldots$ *En este caso* $\frac{3}{2}$ *es el valor propio dominante, así que, si* $b_1 \neq 0$*, tenemos* $\boldsymbol{v}_k \approx b_1 \left(\frac{3}{2}\right)^k \begin{bmatrix} -1 \\ 1 \end{bmatrix}$ *para* k *grande y* $\boldsymbol{v}_k$ *se aproxima a la línea* $y = -x$*.*

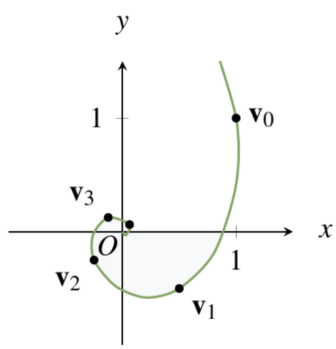
Sin embargo, si $b_1 = 0$*, entonces* $\boldsymbol{v}_k = b_2 \left(\frac{1}{2}\right)^k \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *y así se aproxima al origen a lo largo de la línea* $y = x$*. En general, las trayectorias aparecen como en el diagrama, y el origen se llama un punto de silla para el sistema.*



Ejemplo 2.17 *Sea* $\boldsymbol{A} = \begin{bmatrix} 0 & \frac{1}{2} \\ -\frac{1}{2} & 0 \end{bmatrix}$. *Ahora el polinomio característico es* $f_A(x) = x^2 + \frac{1}{4}$, *por lo que los valores propios son los números complejos* $\frac{i}{2}$ *y* $-\frac{i}{2}$ *donde* $i^2 = -1$. *Por lo tanto,* $\boldsymbol{A}$ *no es diagonalizable como una matriz real. Sin embargo, las trayectorias no son difíciles de describir. Si comenzamos con* $\boldsymbol{v}_0 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$ *entonces la trayectoria comienza como*

$$\boldsymbol{v}_1 = \begin{bmatrix} \frac{1}{2} \\ -\frac{1}{2} \end{bmatrix}, \boldsymbol{v}_2 = \begin{bmatrix} -\frac{1}{4} \\ -\frac{1}{4} \end{bmatrix}, \boldsymbol{v}_3 = \begin{bmatrix} -\frac{1}{8} \\ \frac{1}{8} \end{bmatrix}, \boldsymbol{v}_4 = \begin{bmatrix} \frac{1}{16} \\ \frac{1}{16} \end{bmatrix}, \boldsymbol{v}_5 = \begin{bmatrix} \frac{1}{32} \\ -\frac{1}{32} \end{bmatrix}, \boldsymbol{v}_6 = \begin{bmatrix} -\frac{1}{64} \\ -\frac{1}{64} \end{bmatrix}, \ldots$$

Los primeros cinco de estos puntos se trazan en el diagrama. Aquí cada trayectoria espiral hacia el origen, por lo que el origen es un atractor. Nótese que los dos valores propios (complejos) tienen modulo menor que 1 aquí. Si tuvieran modulo mayor que 1, las trayectorias en espiral se alejarían del origen.



2.3.3 De vuela al proceso de diagonalización

El Teorema 2.11 afirma (en parte) que una matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ es diagonalizable si y sólo si tiene n eigenvectores $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ tales que la matriz $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \cdots & \boldsymbol{x}_n \end{bmatrix}$ con los $\boldsymbol{x}_i$ como columnas es invertible. Esto es equivalente a requerir que $\{\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\}$ sea una base de $\mathbb{R}^n$ que consista en eigenvectores de $\boldsymbol{A}$. Por lo tanto, podemos reformular el Teorema 2.11 de la siguiente manera:

V. Osores

Teorema 2.15 *Sea* $\boldsymbol{A} \in \mathbb{R}^{n \times n}$.

1. $\boldsymbol{A}$ *es diagonalizable si y sólo si* $\mathbb{R}^n$ *tiene una base* $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ *que consista en eigenvectores de* $\boldsymbol{A}$.

2. *Cuando esto es el caso, la matriz* $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix}$ *es invertible y*

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \operatorname{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_n\right),$$

donde, para cada i, λ_i *es el valor propio de* $\boldsymbol{A}$ *correspondiente a* $\boldsymbol{x}_i$.

El siguiente resultado es una herramienta básica para determinar cuándo una matriz es diagonalizable. Revela una conexión importante entre los valores propios y la independencia lineal: Los eigenvectores correspondientes a valores propios distintos son necesariamente linealmente independientes.

Teorema 2.16 *Sean* $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ *eigenvectores correspondientes a valores propios distintos* $\lambda_1, \lambda_2, \ldots, \lambda_k$ *de una matriz* $\boldsymbol{A} \in \mathbb{R}^{n \times n}$. *Entonces* $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ *es un conjunto linealmente independiente.*

Demostración: *Utilizamos la inducción en* k. *Si* $k = 1$, *entonces* $\{\boldsymbol{x}_1\}$ *es independiente porque* $\boldsymbol{x}_1 \neq \mathbf{0}$. *En general, supongamos que el teorema es verdadero para algún* $k \geq 1$. *Dados los eigenvectores* $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_{k+1}\}$, *supongamos que una combinación lineal se anula:*

$$t_1\boldsymbol{x}_1 + t_2\boldsymbol{x}_2 + \cdots + t_{k+1}\boldsymbol{x}_{k+1} = \mathbf{0} \tag{2.3.4}$$

Debemos mostrar que cada $t_i = 0$. *Multiplicamos a la izquierda* (2.3.4) *por* $\boldsymbol{A}$ *y utilizamos el hecho de que* $\boldsymbol{A}\boldsymbol{x}_i = \lambda_i\boldsymbol{x}_i$ *para obtener*

$$t_1\lambda_1\boldsymbol{x}_1 + t_2\lambda_2\boldsymbol{x}_2 + \cdots + t_{k+1}\lambda_{k+1}\boldsymbol{x}_{k+1} = \mathbf{0} \tag{2.3.5}$$

Si multiplicamos (2.3.4) *por* λ_1 *y restamos el resultado de* (2.3.5), *los primeros términos se cancelan y obtenemos*

$$t_2\left(\lambda_2 - \lambda_1\right)\boldsymbol{x}_2 + t_3\left(\lambda_3 - \lambda_1\right)\boldsymbol{x}_3 + \cdots + t_{k+1}\left(\lambda_{k+1} - \lambda_1\right)\boldsymbol{x}_{k+1} = \mathbf{0}$$

Dado que $\boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_{k+1}$ *corresponden a valores propios distintos* $\lambda_2, \lambda_3, \ldots, \lambda_{k+1}$, *el conjunto*

$$\{\boldsymbol{x}_2, \boldsymbol{x}_3, \ldots, \boldsymbol{x}_{k+1}\}$$

es independiente por la hipótesis de inducción. Por lo tanto,

$$t_2\left(\lambda_2 - \lambda_1\right) = 0, \quad t_3\left(\lambda_3 - \lambda_1\right) = 0, \quad \ldots, \quad t_{k+1}\left(\lambda_{k+1} - \lambda_1\right) = 0$$

y así $t_2 = t_3 = \cdots = t_{k+1} = 0$ *porque los* λ_i *son distintos. Por lo tanto,* (2.3.4) *se convierte en* $t_1\boldsymbol{x}_1 = \mathbf{0}$, *lo que implica que* $t_1 = 0$ *porque* $\boldsymbol{x}_1 \neq \mathbf{0}$. *Esto es lo que queríamos.*

El Teorema 2.16 se aplicará varias veces; comenzamos usándolo para dar una condición útil para cuando una matriz es diagonalizable.

Teorema 2.17 *Si $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ es una matriz con n valores propios distintos, entonces $\boldsymbol{A}$ es diagonalizable.*

Demostración: *Elegimos un eigenvector para cada uno de los n valores propios distintos. Luego, estos eigenvectores son independientes por el Teorema 2.16, y por lo tanto son una base de $\mathbb{R}^n$ por el Teorema 2.6. Ahora utilizamos el Teorema 2.15.*

Ejemplo 2.18 *Demuestre que* $\boldsymbol{A} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & 2 & 3 \\ -1 & 1 & 0 \end{bmatrix}$ *es diagonalizable.*

Sol: *Un cálculo rutinario muestra la matriz tiene valores propios distintos 1, 3, y -1. Por lo tanto, se aplica el Teorema 2.17.*

Sin embargo, una matriz puede tener múltiples valores propios. Para tratar esta situación, probamos un lema importante que formaliza una técnica básica para la diagonalización.

Lema 2.4 *Sea $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k\}$ un conjunto linealmente independiente de vectores propios de una matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, extiéndalo a una base $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k, \ldots, \boldsymbol{x}_n\}$ de $\mathbb{R}^n$, y sea*

$$\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix}$$

la matriz $n\times n$ (invertible) con las $\boldsymbol{x}_i$ como sus columnas. Si $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los valores propios (no necesariamente distintos) de $\boldsymbol{A}$ correspondientes a $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_k$ respectivamente, entonces $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ tiene la forma de bloque

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \operatorname{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k) & \boldsymbol{B} \\ 0 & \boldsymbol{A}_1 \end{bmatrix},$$

donde $\boldsymbol{B}$ tiene tamaño $k\times(n-k)$ y $\boldsymbol{A}_1$ tiene tamaño $(n-k)\times(n-k)$.

Demostración: *Si $\{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ es la base estándar de $\mathbb{R}^n$, entonces*

$$\begin{aligned} \begin{bmatrix} \boldsymbol{e}_1 & \boldsymbol{e}_2 & \ldots & \boldsymbol{e}_n \end{bmatrix} = \boldsymbol{I}_n = \boldsymbol{P}^{-1}\boldsymbol{P} &= \boldsymbol{P}^{-1}\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix} \\ &= \begin{bmatrix} \boldsymbol{P}^{-1}\boldsymbol{x}_1 & \boldsymbol{P}^{-1}\boldsymbol{x}_2 & \cdots & \boldsymbol{P}^{-1}\boldsymbol{x}_n \end{bmatrix}. \end{aligned}$$

Comparando columnas, tenemos que $\boldsymbol{P}^{-1}\boldsymbol{x}_i = \boldsymbol{e}_i$ para cada $1 \leq i \leq n$. Por otro lado, observe que

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}^{-1}\boldsymbol{A}\begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix} = \begin{bmatrix} \left(\boldsymbol{P}^{-1}\boldsymbol{A}\right)\boldsymbol{x}_1 & \left(\boldsymbol{P}^{-1}\boldsymbol{A}\right)\boldsymbol{x}_2 & \cdots & \left(\boldsymbol{P}^{-1}\boldsymbol{A}\right)\boldsymbol{x}_n \end{bmatrix}.$$

Por lo tanto, si $1 \leq i \leq k$, la columna i de $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ es

$$\left(\boldsymbol{P}^{-1}\boldsymbol{A}\right)\boldsymbol{x}_i = \boldsymbol{P}^{-1}(\lambda_i\boldsymbol{x}_i) = \lambda_i\left(\boldsymbol{P}^{-1}\boldsymbol{x}_i\right) = \lambda_i\boldsymbol{e}_i$$

Esto describe las primeras k columnas de $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$, y el Lema 2.4 se sigue.

Nótese que el Lema 2.4 (con $k = n$) muestra que una matriz $\boldsymbol{A}$ de tamaño $n\times n$ es diagonalizable si $\mathbb{R}^n$ tiene una base de vectores propios de $\boldsymbol{A}$, como en (1) del Teorema 2.15.

Definición 2.15 **(Espacio propio de una matriz)** *Si λ es un valor propio de una matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$, se define el espacio propio de $\boldsymbol{A}$ correspondiente a λ por*

$$E_\lambda(\boldsymbol{A}) = \{\boldsymbol{x} \in \mathbb{R}^n : \boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}\}.$$

Esto es un subespacio de $\mathbb{R}^n$ y los vectores propios correspondientes a λ son simplemente los vectores no nulos en $E_\lambda(\boldsymbol{A})$. De hecho, $E_\lambda(\boldsymbol{A})$ es el núcleo de la matriz $(\lambda I - A)$:

$$E_\lambda(\boldsymbol{A}) = \{\boldsymbol{x} : (\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}\} = \mathrm{null}(\lambda\boldsymbol{I} - \boldsymbol{A}).$$

Por lo tanto, las soluciones básicas del sistema homogéneo $(\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}$ dadas por el algoritmo de Gauss forman una base para $E_\lambda(\boldsymbol{A})$. En particular

$$\dim E_\lambda(\boldsymbol{A}) \text{ es el número de soluciones básicas } \boldsymbol{x} \text{ de } (\lambda\boldsymbol{I} - \boldsymbol{A})\boldsymbol{x} = \boldsymbol{0}. \tag{2.3.6}$$

Ahora recuerde (Definición 2.12) que la multiplicidad de un valor propio λ de $\boldsymbol{A}$ es el número de veces que λ aparece como raíz del polinomio característico $f_A(x)$ de $\boldsymbol{A}$. En otras palabras, la multiplicidad de λ es el entero más grande $m \geq 1$ tal que

$$f_A(x) = (x-\lambda)^m g(x)$$

para algún polinomio $g(x)$. Debido a (2.3.6), la afirmación (sin prueba) en el Teorema 2.12 se puede expresar de la siguiente manera: Una matriz cuadrada es diagonalizable si y sólo si la multiplicidad de cada valor propio λ es igual a $\dim[E_\lambda(\boldsymbol{A})]$. Vamos a demostrar esto, y la prueba requiere el siguiente resultado que es válido para cualquier matriz cuadrada, diagonalizable o no.

Lema 2.5 *Sea λ un valor propio de multiplicidad m de una matriz cuadrada $\boldsymbol{A} \in \mathbb{R}^{n\times n}$. Entonces $\dim[E_\lambda(\boldsymbol{A})] \leq m$.*

Demostración: *Escriba $\dim[E_\lambda(\boldsymbol{A})] = d$. Es suficiente mostrar que $f_A(x) = (x-\lambda)^d g(x)$ para algún polinomio $g(x)$, porque m es la mayor potencia de $(x-\lambda)$ que divide $f_A(x)$. Con este fin, sea $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d\}$ una base de $E_\lambda(\boldsymbol{A})$. Entonces el Lema 2.4 muestra que existe una matriz invertible $\boldsymbol{P}$ de tamaño $n\times n$ tal que*

$$\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \lambda\boldsymbol{I}_d & \boldsymbol{B} \\ 0 & \boldsymbol{A}_1 \end{bmatrix},$$

en forma de bloque, donde $\boldsymbol{I}_d$ denota la matriz identidad $d\times d$. Ahora escriba $\boldsymbol{A}' = \boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ y observe que $f_{A'}(x) = f_A(x)$.Pero el Teorema de determinantes de matrices por bloques da

$$\begin{aligned} f_A(x) = f_{A'}(x) = \det\left(x\boldsymbol{I}_n - \boldsymbol{A}'\right) &= \det\begin{bmatrix} (x-\lambda)\boldsymbol{I}_d & -\boldsymbol{B} \\ 0 & x\boldsymbol{I}_{n-d} - \boldsymbol{A}_1 \end{bmatrix} \\ &= \det\left[(x-\lambda)\boldsymbol{I}_d\right]\det\left[(x\boldsymbol{I}_{n-d} - \boldsymbol{A}_1)\right] \\ &= (x-\lambda)^d\det(x\boldsymbol{I}_{n-d} - \boldsymbol{A}_1) \\ &= (x-\lambda)^d g(x) \end{aligned}$$

donde $g(x) = \det\left(x\boldsymbol{I}_{n-d} - \boldsymbol{A}_1\right)$ es un polinomio, como se requería.

Es imposible ignorar la cuestión de cuándo se cumple la igualdad en el Lema 2.5 para cada valor propio λ. Resulta que esto caracteriza las matrices diagonalizables $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ para las cuales $f_A(x)$ se factoriza completamente sobre $\mathbb{R}$. Con esto queremos decir que $f_A(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$, donde los λ_i son números reales (no necesariamente distintos); en otras palabras, cada valor propio de $\boldsymbol{A}$ es real. Esto no siempre sucede (considérese $A = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$), y investigamos el caso general a continuación.

Teorema 2.18 *Las siguientes afirmaciones son equivalentes para una matriz cuadrada $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ para la cual $f_A(x)$ se factoriza completamente.*

1. *$\boldsymbol{A}$ es diagonalizable.*
2. *$\dim\left[E_\lambda(\boldsymbol{A})\right]$ es igual a la multiplicidad de λ para cada valor propio λ de la matriz $\boldsymbol{A}$.*[a]

[a]Esto a menudo se llama la multiplicidad algebraica de λ.

Demostración: *Sea $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ una matriz y sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los valores propios distintos de $\boldsymbol{A}$. Para cada i, sea m_i la multiplicidad de λ_i y escriba $d_i = \dim\left[E_{\lambda_i}(\boldsymbol{A})\right]$. Entonces*

$$f_A(x) = (x-\lambda_1)^{m_1}(x-\lambda_2)^{m_2}\ldots(x-\lambda_k)^{m_k},$$

así que $m_1 + \cdots + m_k = n$ porque $f_A(x)$ tiene grado n. Además, $d_i \leq m_i$ para cada i según el Lema 2.5.

(1) $\Rightarrow$ (2). Según (1), $\mathbb{R}^n$ tiene una base de n vectores propios de $\boldsymbol{A}$, así que sea t_i el número de ellos que se encuentran en $E_{\lambda_i}(\boldsymbol{A})$ para cada i. Dado que el subespacio generado por estos t_i vectores propios tiene dimensión t_i, tenemos $t_i \leq d_i$ para cada i según el Teorema 2.5. Por lo tanto,

$$n = t_1 + \cdots + t_k \leq d_1 + \cdots + d_k \leq m_1 + \cdots + m_k = n.$$

Se sigue que $d_1 + \cdots + d_k = m_1 + \cdots + m_k$, por lo que, dado que $d_i \leq m_i$ para cada i, debemos tener $d_i = m_i$. Esto es (2).

(2) $\Rightarrow$ (1). Sea B_i una base de $E_{\lambda_i}(\boldsymbol{A})$ para cada i, y sea $B = B_1 \cup \cdots \cup B_k$. Dado que cada B_i contiene m_i vectores según (2), y dado que los B_i son disjuntos entre sí (los λ_i son distintos), se sigue que B contiene n vectores. Entonces, basta con mostrar que B es linealmente independiente (entonces B es una base de $\mathbb{R}^n$). Supongamos que una combinación lineal de los vectores en B es cero, y sea $\boldsymbol{y}_i$ la suma de todos los términos que provienen de B_i. Entonces $\boldsymbol{y}_i$ está en $E_{\lambda_i}(\boldsymbol{A})$, por lo que los $\boldsymbol{y}_i$ no nulos son independientes según el Teorema 2.16 (ya que los λ_i son distintos). Dado que la suma de los $\boldsymbol{y}_i$ es cero, se sigue que $\boldsymbol{y}_i = \mathbf{0}$ para cada i. Por lo tanto, todos los coeficientes de los términos en $\boldsymbol{y}_i$ son cero (porque B_i es independiente). Dado que esto se cumple para cada i, se demuestra que B es independiente.

Ejemplo 2.19 *Si $\boldsymbol{A} = \begin{bmatrix} 5 & 8 & 16 \\ 4 & 1 & 8 \\ -4 & -4 & -11 \end{bmatrix}$ y $B = \begin{bmatrix} 2 & 1 & 1 \\ 2 & 1 & -2 \\ -1 & 0 & -2 \end{bmatrix}$ muestre que $\boldsymbol{A}$ es diagonalizable pero $\boldsymbol{B}$ no lo es.*

Sol: *Tenemos* $f_A(x) = (x+3)^2(x-1)$ *por lo que los valores propios son* $\lambda_1 = -3$ *y* $\lambda_2 = 1$. *Los correspondientes espacios propios son* $E_{\lambda_1}(\boldsymbol{A}) = \operatorname{span}\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ *y* $E_{\lambda_2}(\boldsymbol{A}) = \operatorname{span}\{\boldsymbol{x}_3\}$ *donde*

$$x_1 = \begin{bmatrix} -1 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} -2 \\ 0 \\ 1 \end{bmatrix}, \boldsymbol{x}_3 = \begin{bmatrix} 2 \\ 1 \\ -1 \end{bmatrix}$$

como se puede verificar. Dado que $\{\boldsymbol{x}_1, \boldsymbol{x}_2\}$ *es independiente, tenemos* $\dim(E_{\lambda_1}(\boldsymbol{A})) = 2$ *que es la multiplicidad de* λ_1. *De manera similar,* $\dim(E_{\lambda_2}(\boldsymbol{A})) = 1$ *iguala la multiplicidad de* λ_2. *Por lo tanto,* $\boldsymbol{A}$ *es diagonalizable según el Teorema 2.18, y una matriz diagonalizadora es* $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 \end{bmatrix}$. *Pasando a* $\boldsymbol{B}$, $f_B(x) = (x+1)^2(x-3)$ *por lo que los valores propios son* $\lambda_1 = -1$ *y* $\lambda_2 = 3$. *Los correspondientes espacios propios son* $E_{\lambda_1}(\boldsymbol{B}) = \operatorname{span}\{\boldsymbol{y}_1\}$ *y* $E_{\lambda_2}(\boldsymbol{B}) = \operatorname{span}\{\boldsymbol{y}_2\}$ *donde*

$$y_1 = \begin{bmatrix} -1 \\ 2 \\ 1 \end{bmatrix}, \boldsymbol{y}_2 = \begin{bmatrix} 5 \\ 6 \\ -1 \end{bmatrix}$$

Aquí $\dim(E_{\lambda_1}(\boldsymbol{B})) = 1$ *es menor que la multiplicidad de* λ_1, *por lo que la matriz* $\boldsymbol{B}$ *no es diagonalizable, nuevamente según el Teorema 2.18. El hecho de que* $\dim(E_{\lambda_1}(\boldsymbol{B})) = 1$ *significa que no hay posibilidad de encontrar tres vectores propios linealmente independientes.*

Teorema 2.19 *Sea* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *una matriz simétrica. Si* λ *es cualquier valor propio complejo de* $\boldsymbol{A}$, *entonces* λ *es real.*

Demostración: *Observe que* $\bar{\boldsymbol{A}} = \boldsymbol{A}$ *porque* $\boldsymbol{A}$ *es real. Si* λ *es un valor propio de* $\boldsymbol{A}$, *mostraremos que* λ *es real al demostrar que* $\bar{\lambda} = \lambda$. *Sea* $\boldsymbol{x}$ *un vector propio (posiblemente complejo) correspondiente a* λ, *de modo que* $\boldsymbol{x} \neq \boldsymbol{0}$ *y* $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$. *Definimos* $c = \boldsymbol{x}^T\bar{\boldsymbol{x}}$. *Si escribimos*

$$\boldsymbol{x} = \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix}$$

donde los z_i *son números complejos, tenemos*

$$c = \boldsymbol{x}^T\bar{\boldsymbol{x}} = z_1\bar{z}_1 + z_2\bar{z}_2 + \cdots + z_n\bar{z}_n = |\bar{z}_1|^2 + |\bar{z}_2|^2 + \cdots + |\bar{z}_n|^2 .$$

Así que c *es un número real, y* $c > 0$ *porque al menos uno de los* $z_i \neq 0$ *(ya que* $\boldsymbol{x} \neq \boldsymbol{0}$*). Mostramos que* $\bar{\lambda} = \lambda$ *al verificar que* $\lambda c = \bar{\lambda} c$. *Tenemos*

$$\lambda c = \lambda\left(\boldsymbol{x}^T\bar{\boldsymbol{x}}\right) = (\lambda\boldsymbol{x})^T\bar{\boldsymbol{x}} = (\boldsymbol{A}x)^T\bar{\boldsymbol{x}} = \boldsymbol{x}^T\boldsymbol{A}^T\bar{\boldsymbol{x}}.$$

En este punto, usamos la hipótesis de que $\boldsymbol{A}$ *es simétrica y real. Esto significa* $\boldsymbol{A}^T = \boldsymbol{A} = \bar{\boldsymbol{A}}$ *por lo que continuamos con el cálculo:*

$$\begin{aligned} \lambda c = \boldsymbol{x}^T\boldsymbol{A}^T\bar{\boldsymbol{x}} = \boldsymbol{x}^T(\bar{\boldsymbol{A}}\bar{\boldsymbol{x}}) = \boldsymbol{x}^T(\overline{\boldsymbol{A}\boldsymbol{x}}) &= \boldsymbol{x}^T(\overline{\lambda\boldsymbol{x}}) \\ &= \boldsymbol{x}^T(\bar{\lambda}\bar{\boldsymbol{x}}) \\ &= \bar{\lambda}\boldsymbol{x}^T\bar{\boldsymbol{x}} \\ &= \bar{\lambda}c \end{aligned}$$

como se requería.

2.3.4 Diagonalización Ortogonal

Definición 2.16 (Matrices ortogonalmente diagonalizables) *Una matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ se dice que es ortogonalmente diagonalizable cuando se puede encontrar una matriz ortogonal $\boldsymbol{P}$ tal que $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}$ es diagonal.*

Esta condición resulta caracterizar las matrices simétricas.

Teorema 2.20 (Teorema de los ejes principales) *Las siguientes condiciones son equivalentes para una matriz $\boldsymbol{A} \in \mathbb{R}^{n\times n}$.*

1. *$\boldsymbol{A}$ tiene un conjunto ortonormal de n eigenvectores.*
2. *$\boldsymbol{A}$ es ortogonalmente diagonalizable.*
3. *$\boldsymbol{A}$ es simétrica.*

Demostración: *(1) $\Leftrightarrow$ (2). Dado (1), sea $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ un conjunto ortonormal de eigenvectores de $\boldsymbol{A}$. Entonces $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \end{bmatrix}$ es ortogonal, y $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ es diagonal por el Teorema 2.11. Esto prueba (2). Por el contrario, dado (2) sea $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}$ diagonal donde $\boldsymbol{P}$ es ortogonal. Si $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n$ son las columnas de $\boldsymbol{P}$ entonces $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ es una base ortonormal de $\mathbb{R}^n$ que consiste en eigenvectores de $\boldsymbol{A}$ por el Teorema 2.11. Esto prueba (1).*

(2) $\Rightarrow$ (3). Si $\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P} = \boldsymbol{D}$ es diagonal, donde $\boldsymbol{P}^{-1} = \boldsymbol{P}^T$, entonces $\boldsymbol{A} = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^T$. Pero $\boldsymbol{D}^T = \boldsymbol{D}$, así que esto da $\boldsymbol{A}^T = \boldsymbol{P}^{TT}\boldsymbol{D}^T\boldsymbol{P}^T = \boldsymbol{P}\boldsymbol{D}\boldsymbol{P}^T = \boldsymbol{A}$.

(3) $\Rightarrow$ (2). Si $\boldsymbol{A}$ es una matriz $n \times n$ simétrica, procedemos por inducción en n. Si $n = 1$, $\boldsymbol{A}$ ya es diagonal. Si $n > 1$, asumimos que (3) $\Rightarrow$ (2) para matrices simétricas de $(n-1)\times(n-1)$. Por el Teorema 2.19 sea λ_1 un eigenvalor (real) de $\boldsymbol{A}$, y sea $\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1$, donde $\|\boldsymbol{x}_1\| = 1$. Utilizamos el algoritmo de Gram-Schmidt para encontrar una base ortonormal $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ para $\mathbb{R}^n$. Sea $\boldsymbol{P}_1 = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \ldots & \boldsymbol{x}_n \end{bmatrix}$, entonces $\boldsymbol{P}_1$ es una matriz ortogonal y $\boldsymbol{P}_1^T\boldsymbol{A}\boldsymbol{P}_1 = \begin{bmatrix} \lambda_1 & \boldsymbol{B} \\ 0 & \boldsymbol{A}_1 \end{bmatrix}$ en forma de bloque por el Lema 2.4. Pero $\boldsymbol{P}_1^T\boldsymbol{A}\boldsymbol{P}_1$ es simétrica ($\boldsymbol{A}$ lo es), por lo que se sigue que $\boldsymbol{B} = 0$ y $\boldsymbol{A}_1$ es simétrica. Luego, por inducción, existe una matriz ortogonal de $(n-1)\times(n-1)$ $\boldsymbol{Q}$ tal que $\boldsymbol{Q}^T\boldsymbol{A}_1\boldsymbol{Q} = \boldsymbol{D}_1$ es diagonal. Observemos que $\boldsymbol{P}_2 = \begin{bmatrix} 1 & 0 \\ 0 & \boldsymbol{Q} \end{bmatrix}$ es ortogonal, y calculemos:

$$
\begin{aligned}
(\boldsymbol{P}_1\boldsymbol{P}_2)^T \boldsymbol{A} (\boldsymbol{P}_1\boldsymbol{P}_2) &= \boldsymbol{P}_2^T \left(\boldsymbol{P}_1^T\boldsymbol{A}\boldsymbol{P}_1\right) \boldsymbol{P}_2 \\
&= \begin{bmatrix} 1 & 0 \\ 0 & \boldsymbol{Q}^T \end{bmatrix} \begin{bmatrix} \lambda_1 & 0 \\ 0 & \boldsymbol{A}_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \boldsymbol{Q} \end{bmatrix} \\
&= \begin{bmatrix} \lambda_1 & 0 \\ 0 & \boldsymbol{D}_1 \end{bmatrix}
\end{aligned}
$$

es diagonal. Debido a que $\boldsymbol{P}_1\boldsymbol{P}_2$ es ortogonal, esto prueba (2).

Un conjunto de eigenvectores ortonormales de una matriz simétrica $\boldsymbol{A}$ se llama un conjunto de ejes principales para $\boldsymbol{A}$. El nombre proviene de la geometría, y esto se discute en la sección **??**. Debido a que los eigenvalores de una matriz simétrica (real) son reales, el Teorema 2.20 también se llama el teorema espectral real, y el conjunto de eigenvalores distintos se llama el espectro de la matriz. En general, el teorema espectral es un resultado similar para matrices con entradas complejas (Teorema 2.27).

Ejemplo 2.20 *Encuentre una matriz ortogonal $\boldsymbol{P}$ tal que $\boldsymbol{P}^{-1}\boldsymbol{AP}$ sea diagonal, donde*

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

Sol:

$$f_{\boldsymbol{A}}(x) = \det \begin{bmatrix} x-1 & 0 & 1 \\ 0 & x-1 & -2 \\ 1 & -2 & x-5 \end{bmatrix} = x(x-1)(x-6)$$

Así, los eigenvalores son $\lambda = 0, 1$ y 6, y los eigenvectores correspondientes son

$$\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ -2 \\ 1 \end{bmatrix}, \boldsymbol{x}_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}, \boldsymbol{x}_3 = \begin{bmatrix} -1 \\ 2 \\ 5 \end{bmatrix}.$$

Además, por lo que parece ser una suerte notable, estos eigenvectores son ortogonales (verifique). Tenemos $\|\boldsymbol{x}_1\|^2 = 6, \|\boldsymbol{x}_2\|^2 = 5$, y $\|\boldsymbol{x}_3\|^2 = 30$, así que

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{\sqrt{6}}\boldsymbol{x}_1 & \frac{1}{\sqrt{5}}\boldsymbol{x}_2 & \frac{1}{\sqrt{30}}\boldsymbol{x}_3 \end{bmatrix} = \frac{1}{\sqrt{30}} \begin{bmatrix} \sqrt{5} & 2\sqrt{6} & -1 \\ -2\sqrt{5} & \sqrt{6} & 2 \\ \sqrt{5} & 0 & 5 \end{bmatrix}$$

es una matriz ortogonal. Así $\boldsymbol{P}^{-1} = \boldsymbol{P}^T$ y

$$\boldsymbol{P}^T\boldsymbol{AP} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 6 \end{bmatrix}$$

según el algoritmo de diagonalización.

De hecho, el hecho de que los eigenvectores en el ejemplo 2.20 sean ortogonales no es una coincidencia. El Teorema 2.16 garantiza que son linealmente independientes (corresponden a eigenvalores distintos); el hecho de que la matriz sea simétrica implica que son ortogonales.

Teorema 2.21 *Si $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ es una matriz simétrica, entonces los vectores propios de $\boldsymbol{A}$ correspondientes a valores propios distintos son ortogonales.*

Demostración: *Sea $\boldsymbol{Ax} = \lambda \boldsymbol{x}$ y $\boldsymbol{Ay} = \mu \boldsymbol{y}$, donde $\lambda \neq \mu$. Utilizando el Teorema 2.3, calculamos*

$$\lambda(\boldsymbol{x}\cdot\boldsymbol{y}) = (\lambda\boldsymbol{x})\cdot\boldsymbol{y} = (\boldsymbol{Ax})\cdot\boldsymbol{y} = \boldsymbol{x}\cdot(\boldsymbol{Ay}) = \boldsymbol{x}\cdot(\mu\boldsymbol{y}) = \mu(\boldsymbol{x}\cdot\boldsymbol{y}).$$

Por lo tanto, $(\lambda-\mu)(\boldsymbol{x}\cdot\boldsymbol{y}) = 0$, y así $\boldsymbol{x}\cdot\boldsymbol{y} = 0$ porque $\lambda \neq \mu$.

Ahora el procedimiento para diagonalizar una matriz simétrica $n \times n$ es claro. Encuentra los valores propios distintos (todos reales por el Teorema 2.19) y encuentra bases ortonormales para cada subespacio propio (puede ser necesario el algoritmo de ortogonalización de Gram-Schmidt). Luego, el conjunto de todos estos vectores base es ortonormal (por el Teorema 2.21) y contiene n vectores. Aquí hay un ejemplo.

Ejemplo 2.21 *Diagonalice ortogonalmente la matriz simétrica* $\boldsymbol{A} = \begin{bmatrix} 8 & -2 & 2 \\ -2 & 5 & 4 \\ 2 & 4 & 5 \end{bmatrix}$.

Sol: *El polinomio característico viene dado por*

$$f_{\boldsymbol{A}}(x) = \det\begin{bmatrix} x-8 & 2 & -2 \\ 2 & x-5 & -4 \\ -2 & -4 & x-5 \end{bmatrix} = x(x-9)^2.$$

Por lo tanto, los valores propios distintos son 0 y 9 con multiplicidades 1 y 2, respectivamente, así que $\dim(E_0) = 1$ y $\dim(E_9) = 2$ por el Teorema 2.18 ($\boldsymbol{A}$ es diagonalizable, siendo simétrica). La eliminación gaussiana da

$$E_0(\boldsymbol{A}) = \operatorname{span}\{\boldsymbol{x}_1\}, \quad \boldsymbol{x}_1 = \begin{bmatrix} 1 \\ 2 \\ -2 \end{bmatrix}, \quad y \quad E_9(\boldsymbol{A}) = \operatorname{span}\left\{\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}, \begin{bmatrix} 2 \\ 0 \\ 1 \end{bmatrix}\right\}$$

Los vectores propios en E_9 son ortogonales a $\boldsymbol{x}_1$ como garantiza el Teorema 2.21, pero no entre sí. Sin embargo, el proceso de Gram-Schmidt produce una base ortogonal

$$\{\boldsymbol{x}_2, \boldsymbol{x}_3\} \text{ de } E_9(\boldsymbol{A}) \quad \text{donde} \quad \boldsymbol{x}_2 = \begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix} \ y\ \boldsymbol{x}_3 = \begin{bmatrix} 2 \\ 4 \\ 5 \end{bmatrix}$$

Normalizando se obtienen vectores ortonormales $\left\{\frac{1}{3}\boldsymbol{x}_1, \frac{1}{\sqrt{5}}\boldsymbol{x}_2, \frac{1}{3\sqrt{5}}\boldsymbol{x}_3\right\}$, así que

$$\boldsymbol{P} = \begin{bmatrix} \frac{1}{3}\boldsymbol{x}_1 & \frac{1}{\sqrt{5}}\boldsymbol{x}_2 & \frac{1}{3\sqrt{5}}\boldsymbol{x}_3 \end{bmatrix} = \frac{1}{3\sqrt{5}}\begin{bmatrix} \sqrt{5} & -6 & 2 \\ 2\sqrt{5} & 3 & 4 \\ -2\sqrt{5} & 0 & 5 \end{bmatrix}$$

es una matriz ortogonal tal que $\boldsymbol{P}^{-1}\boldsymbol{AP}$ es diagonal.

Vale la pena notar que existen otras matrices de diagonalización $\boldsymbol{P}$ *más convenientes. Por ejemplo,* $\boldsymbol{y}_2 = \begin{bmatrix} 2 \\ 1 \\ 2 \end{bmatrix}$ *y* $\boldsymbol{y}_3 = \begin{bmatrix} -2 \\ 2 \\ 1 \end{bmatrix}$ *se encuentran en* $E_9(\boldsymbol{A})$ *y son ortogonales. Además, ambas tienen norma 3 (como también la tiene* $\boldsymbol{x}_1$*), así que*

$$\boldsymbol{Q} = \begin{bmatrix} \frac{1}{3}\boldsymbol{x}_1 & \frac{1}{3}\boldsymbol{y}_2 & \frac{1}{3}\boldsymbol{y}_3 \end{bmatrix} = \frac{1}{3}\begin{bmatrix} 1 & 2 & -2 \\ 2 & 1 & 2 \\ -2 & 2 & 1 \end{bmatrix}$$

es una matriz ortogonal más agradable con la propiedad de que $\boldsymbol{Q}^{-1}\boldsymbol{A}\boldsymbol{Q}$ *es diagonal.*

Si estamos dispuestos a reemplazar "diagonal" por "triangular superior" en el teorema de los ejes principales, podemos debilitar el requisito de que $\boldsymbol{A}$ sea simétrica, insistiendo solo en que $\boldsymbol{A}$ tenga valores propios reales.

Teorema 2.22 (Teorema de triangulación) *Si* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *con* n *valores propios reales, existe una matriz ortogonal* $\boldsymbol{P}$ *tal que* $\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}$ *es triangular superior.*

Demostración: *Modificamos la demostración del Teorema 2.20. Si* $\boldsymbol{A}\boldsymbol{x}_1 = \lambda_1\boldsymbol{x}_1$ *donde* $\|\boldsymbol{x}_1\| = 1$*, sea* $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ *una base ortonormal de* $\mathbb{R}^n$*, y sea* $\boldsymbol{P}_1 = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \cdots & \boldsymbol{x}_n \end{bmatrix}$*. Entonces* $\boldsymbol{P}_1$ *es ortogonal y* $\boldsymbol{P}_1^T\boldsymbol{A}\boldsymbol{P}_1 = \begin{bmatrix} \lambda_1 & \boldsymbol{B} \\ 0 & \boldsymbol{A}_1 \end{bmatrix}$ *en forma de bloque. Por inducción, sea* $\boldsymbol{Q}^T\boldsymbol{A}_1\boldsymbol{Q} = \boldsymbol{T}_1$ *triangular superior donde* $\boldsymbol{Q}$ *es de tamaño* $(n-1)\times(n-1)$ *y ortogonal. Entonces* $\boldsymbol{P}_2 = \begin{bmatrix} 1 & 0 \\ 0 & \boldsymbol{Q} \end{bmatrix}$ *es ortogonal, así que* $\boldsymbol{P} = \boldsymbol{P}_1\boldsymbol{P}_2$ *también es ortogonal y* $\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \lambda_1 & \boldsymbol{B}\boldsymbol{Q} \\ 0 & \boldsymbol{T}_1 \end{bmatrix}$ *es triangular superior.*

La demostración del Teorema 2.22 no proporciona una manera de construir la matriz $\boldsymbol{P}$. Sin embargo, se dará un algoritmo en la sección 3.2.1 donde se presenta una versión mejorada del Teorema 2.22. En una dirección diferente, una versión del Teorema 2.22 es válida para una matriz arbitraria con entradas complejas (teorema de Schur en la sección 2.4).

Como para una matriz diagonal, los valores propios de una matriz triangular superior se muestran a lo largo de la diagonal principal. Debido a que $\boldsymbol{A}$ y $\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}$ tienen el mismo determinante y traza siempre que $\boldsymbol{P}$ sea ortogonal, el Teorema 2.22 proporciona:

Corolario 2.1 *Si* $\boldsymbol{A} \in \mathbb{R}^{n\times n}$ *es una matriz con valores propios reales* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *(posiblemente no todos distintos), entonces* $\det \boldsymbol{A} = \lambda_1\lambda_2\ldots\lambda_n$ *y* $\operatorname{tr} \boldsymbol{A} = \lambda_1 + \lambda_2 + \cdots + \lambda_n$*.*

Este corolario sigue siendo cierto incluso si los valores propios no son reales (usando el teorema de Schur).

2.3.5 Valores Propios Complejos

Todas las matrices que hemos considerado en la sección han tenido valores propios reales. Pero esto no tiene por qué ser así: La matriz $\boldsymbol{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$ tiene polinomio característico $f_A(x) = x^2 + 1$ que no tiene raíces reales. No obstante, esta matriz es diagonalizable; la única diferencia es que debemos usar un conjunto más grande de escalares, los números complejos.

De hecho, casi todo lo que hemos hecho para matrices reales se puede hacer para matrices complejas. Los métodos son los mismos; la única diferencia es que la aritmética se realiza con números complejos en lugar de reales. Por ejemplo, el algoritmo de Gauss funciona de la misma manera para resolver sistemas de ecuaciones lineales con coeficientes complejos, la multiplicación de matrices se define de la misma manera, y el algoritmo de inversión de matrices funciona de la misma manera.

Pero los números complejos son mejores que los números reales en un aspecto: Mientras que hay polinomios como $x^2 + 1$ con coeficientes reales que no tienen raíz real, este problema no surge con los números complejos: Cada polinomio no constante con coeficientes complejos tiene una raíz compleja y, por lo tanto, se factoriza completamente como un producto de factores lineales. Este hecho se conoce como el teorema fundamental del álgebra.[2]

Ejemplo 2.22 *Diagonalice la matriz* $\boldsymbol{A} = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}$.

Sol: *El polinomio característico de* $\boldsymbol{A}$ *es*

$$f_A(x) = \det(x\boldsymbol{I} - \boldsymbol{A}) = x^2 + 1 = (x - i)(x + i)$$

donde $i^2 = -1$. *Por lo tanto, los valores propios son* $\lambda_1 = i$ *y* $\lambda_2 = -i$, *con los correspondientes vectores propios* $\boldsymbol{x}_1 = \begin{bmatrix} 1 \\ -i \end{bmatrix}$ *y* $\boldsymbol{x}_2 = \begin{bmatrix} 1 \\ i \end{bmatrix}$. *Por lo tanto,* $\boldsymbol{A}$ *es diagonalizable según la versión compleja del Teorema 2.17, y la versión compleja del Teorema 2.15 muestra que* $\boldsymbol{P} = \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ -i & i \end{bmatrix}$ *es invertible y* $\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$. *Por supuesto, esto se puede verificar directamente.*

2.4 Matrices complejas

2.4.1 Matrices Hermitianas y Unitarias

Si $\boldsymbol{A}$ es una matriz simétrica real, está claro que $\boldsymbol{A}^* = \boldsymbol{A}$. Las matrices complejas que satisfacen esta condición resultan ser la generalización más natural de las matrices simétricas reales:

[2]Este fue un famoso problema abierto en 1799 cuando Gauss lo resolvió a los 22 años en su disertación doctoral.

Definición 2.17 **(Matrices hermitianas)** *Una matriz compleja cuadrada $\boldsymbol{A}$ se llama hermitiana si $\boldsymbol{A}^* = \boldsymbol{A}$, o equivalentemente si $\bar{\boldsymbol{A}} = \boldsymbol{A}^T$.*

Las matrices hermitianas son fáciles de reconocer porque los elementos en la diagonal principal deben ser reales, y el "reflejo" de cada elemento no diagonal en la diagonal principal debe ser el conjugado de ese elemento.

Ejemplo 2.23 $\begin{bmatrix} 3 & i & 2+i \\ -i & -2 & -7 \\ 2-i & -7 & 1 \end{bmatrix}$ *es hermitiana, mientras que* $\begin{bmatrix} 1 & i \\ i & -2 \end{bmatrix}$ *y* $\begin{bmatrix} 1 & i \\ -i & i \end{bmatrix}$ *no lo son.*

El siguiente Teorema extiende el Teorema 2.3 y proporciona una caracterización muy útil de las matrices hermitianas en términos del producto interno estándar en $\mathbb{C}^n$.

Teorema 2.23 *Una matriz $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ es hermitiana si y sólo si*

$$\langle \boldsymbol{A}\boldsymbol{z}, \boldsymbol{w}\rangle = \langle \boldsymbol{z}, \boldsymbol{A}\boldsymbol{w}\rangle,$$

para todos las n-tuplas $\boldsymbol{z}$ y $\boldsymbol{w}$ en $\mathbb{C}^n$.

Demostración: *Si $\boldsymbol{A}$ es hermitiana, tenemos $\boldsymbol{A}^T = \bar{\boldsymbol{A}}$. Si $\boldsymbol{z}$ y $\boldsymbol{w}$ son columnas en $\mathbb{C}^n$, entonces $\langle \boldsymbol{z}, \boldsymbol{w}\rangle = \boldsymbol{z}^T\bar{\boldsymbol{w}}$, por lo que*

$$\langle \boldsymbol{A}\boldsymbol{z}, \boldsymbol{w}\rangle = (\boldsymbol{A}\boldsymbol{z})^T\bar{\boldsymbol{w}} = \boldsymbol{z}^T\boldsymbol{A}^T\bar{\boldsymbol{w}} = \boldsymbol{z}^T\bar{\boldsymbol{A}}\bar{\boldsymbol{w}} = \boldsymbol{z}^T(\overline{\boldsymbol{A}\boldsymbol{w}}) = \langle \boldsymbol{z}, A\boldsymbol{w}\rangle.$$

Para demostrar el recíproco, sea $\boldsymbol{e}_j$ la columna j de la matriz identidad. Si $\boldsymbol{A} = [a_{ij}]$, la condición nos da

$$\bar{a}_{ij} = \langle \boldsymbol{e}_i, \boldsymbol{A}\boldsymbol{e}_j\rangle = \langle A\boldsymbol{e}_i, \boldsymbol{e}_j\rangle = a_{ji}.$$

Por lo tanto, $\bar{\boldsymbol{A}} = \boldsymbol{A}^T$, así que $\boldsymbol{A}$ es hermitiana.

Sea $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ una matriz compleja. Al igual que en el caso real, un número complejo λ se llama un valor propio de $\boldsymbol{A}$ si $\boldsymbol{A}\boldsymbol{x} = \lambda\boldsymbol{x}$ se cumple para alguna columna $\boldsymbol{x} \neq \boldsymbol{0}$ en $\mathbb{C}^n$. En este caso, $\boldsymbol{x}$ se llama un vector propio de $\boldsymbol{A}$ correspondiente a λ. El polinomio característico $f_{\boldsymbol{A}}(x)$ se define por

$$f_{\boldsymbol{A}}(x) = \det(x\boldsymbol{I} - \boldsymbol{A}).$$

Este polinomio tiene coeficientes complejos (posiblemente no reales).

Es en este punto donde la ventaja de trabajar con números complejos se hace evidente. Los números reales son incompletos en el sentido de que el polinomio característico de una matriz real puede no tener todas sus raíces reales. Sin embargo, esta dificultad no ocurre con los números complejos. El llamado teorema fundamental del álgebra asegura que todo polinomio de grado positivo con coeficientes complejos tiene una raíz compleja. Por lo tanto, toda matriz cuadrada compleja $\boldsymbol{A}$ tiene un valor propio (complejo). De hecho, $f_{\boldsymbol{A}}(x)$ se factoriza completamente como sigue:

$$f_{\boldsymbol{A}}(x) = (x-\lambda_1)(x-\lambda_2)\cdots(x-\lambda_n)$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_n$ son los valores propios de $\boldsymbol{A}$ (con posibles repeticiones debido a raíces múltiples).

El siguiente resultado muestra que, para las matrices hermitianas, los eigenvalores son realmente números reales. Dado que las matrices simétricas reales son hermitianas, esto vuelve a demostrar el Teorema 2.19. También extiende el Teorema 2.21, que afirma que los eigenvectores de una matriz simétrica real correspondientes a eigenvalores distintos son realmente ortogonales. En el contexto complejo, dos n-tuplas $\boldsymbol{z}$ y $\boldsymbol{w}$ en $\mathbb{C}^n$ se dicen ortogonales si $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = 0$.

Teorema 2.24 *Sea $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ una matriz hermitiana.*

1. *Los valores propios de $\boldsymbol{A}$ son reales.*
2. *Los vectores propios de $\boldsymbol{A}$ correspondientes a valores propios distintos son ortogonales.*

Demostración: *Sean λ y μ valores propios de $\boldsymbol{A}$ con vectores propios (no nulos por definición) $\boldsymbol{z}$ y $\boldsymbol{w}$. Entonces $\boldsymbol{Az} = \lambda \boldsymbol{z}$ y $\boldsymbol{Aw} = \mu \boldsymbol{w}$, por lo que el Teorema 2.23 da*

$$\lambda\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \langle \lambda\boldsymbol{z}, \boldsymbol{w} \rangle = \langle \boldsymbol{Az}, \boldsymbol{w} \rangle = \langle \boldsymbol{z}, \boldsymbol{Aw} \rangle = \langle \boldsymbol{z}, \mu\boldsymbol{w} \rangle = \bar{\mu}\langle \boldsymbol{z}, \boldsymbol{w} \rangle \tag{2.4.1}$$

Si $\mu = \lambda$ y $\boldsymbol{w} = \boldsymbol{z}$, esto se convierte en $\lambda\langle \boldsymbol{z}, \boldsymbol{z} \rangle = \bar{\lambda}\langle \boldsymbol{z}, \boldsymbol{z} \rangle$. Debido a que $\langle \boldsymbol{z}, \boldsymbol{z} \rangle = \|\boldsymbol{z}\|^2 \neq 0$, esto implica $\lambda = \bar{\lambda}$. Así que λ es real, demostrando (1). De manera similar, μ es real, por lo que la ecuación (2.4.1) da $\lambda\langle \boldsymbol{z}, \boldsymbol{w} \rangle = \mu\langle \boldsymbol{z}, \boldsymbol{w} \rangle$. Si $\lambda \neq \mu$, esto implica $\langle \boldsymbol{z}, \boldsymbol{w} \rangle = 0$, demostrando (2).

El teorema de los ejes principales (Teorema 2.20) afirma que toda matriz simétrica real $\boldsymbol{A}$ es ortogonalmente diagonalizable, es decir, $\boldsymbol{P}^T\boldsymbol{AP}$ es diagonal donde $\boldsymbol{P}$ es una matriz ortogonal $\left(\boldsymbol{P}^{-1} = \boldsymbol{P}^T\right)$. El siguiente teorema identifica los análogos complejos de estas matrices reales ortogonales.

Teorema 2.25 *Las siguientes afirmaciones son equivalentes para una matriz compleja $\boldsymbol{A} \in \mathbb{C}^{n\times n}$.*

1. *$\boldsymbol{A}$ es invertible y $\boldsymbol{A}^{-1} = \boldsymbol{A}^*$.*
2. *Las filas de $\boldsymbol{A}$ forman un conjunto ortonormal en $\mathbb{C}^n$.*
3. *Las columnas de $\boldsymbol{A}$ forman un conjunto ortonormal en $\mathbb{C}^n$.*

Demostración: *Si $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{c}_1 & \boldsymbol{c}_2 & \cdots & \boldsymbol{c}_n \end{bmatrix}$ es una matriz compleja con la j-ésima columna $\boldsymbol{c}_j$, entonces $\boldsymbol{A}^T\bar{\boldsymbol{A}} = [\langle \boldsymbol{c}_i, \boldsymbol{c}_j \rangle]$, como en el Teorema 2.8. Ahora, (1) $\Leftrightarrow$ (2) se tiene, y (1) $\Leftrightarrow$ (3) se demuestra de la misma manera.*

Definición 2.18 (Matrices unitarias) *Una matriz $\boldsymbol{U} \in \mathbb{C}^{n\times n}$ se llama unitaria si $\boldsymbol{U}^{-1} = \boldsymbol{U}^*$.*

Así, una matriz real es unitaria si y sólo si es ortogonal.

Ejemplo 2.24 *La matriz $\boldsymbol{A} = \begin{bmatrix} 1+i & 1 \\ 1-i & i \end{bmatrix}$ tiene columnas ortogonales, pero las filas no son ortogonales. Al normalizar las columnas da la matriz unitaria $\frac{1}{2}\begin{bmatrix} 1+i & \sqrt{2} \\ 1-i & \sqrt{2}i \end{bmatrix}$.*

V. Osores

Dada una matriz real simétrica $\boldsymbol{A}$, el algoritmo de diagonalización conduce a un procedimiento para encontrar una matriz ortogonal $\boldsymbol{P}$ tal que $\boldsymbol{P}^T\boldsymbol{A}\boldsymbol{P}$ sea diagonal (ver ejemplo 2.20). El siguiente ejemplo ilustra el Teorema 2.24 y muestra que la técnica funciona para matrices complejas.

Ejemplo 2.25 *Considere la matriz hermitiana* $\boldsymbol{A} = \begin{bmatrix} 3 & 2+i \\ 2-i & 7 \end{bmatrix}$. *Encuentre los valores propios de* $\boldsymbol{A}$, *encuentre dos vectores propios ortonormales, y así encuentre una matriz unitaria* $\boldsymbol{U}$ *tal que* $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}$ *sea diagonal.*

Sol: *El polinomio característico de* $\boldsymbol{A}$ *viene dado por*

$$f_{\boldsymbol{A}}(x) = \det(x\boldsymbol{I} - \boldsymbol{A}) = \det \begin{bmatrix} x-3 & -2-i \\ -2+i & x-7 \end{bmatrix} = (x-2)(x-8)$$

Por lo tanto, los valores propios son 2 y 8 (ambos reales, como se esperaba), y los vectores propios correspondientes son $\begin{bmatrix} 2+i \\ -1 \end{bmatrix}$ *y* $\begin{bmatrix} 1 \\ 2-i \end{bmatrix}$ *(ortogonales, como se esperaba). Cada uno tiene una longitud de* $\sqrt{6}$, *por lo que, como en el algoritmo de diagonalización (real), sea* $U = \frac{1}{\sqrt{6}} \begin{bmatrix} 2+i & 1 \\ -1 & 2-i \end{bmatrix}$ *la matriz unitaria con los vectores propios normalizados como columnas. Entonces* $U^*AU = \begin{bmatrix} 2 & 0 \\ 0 & 8 \end{bmatrix}$ *es diagonal.*

2.4.2 Diagonalización unitaria

Una matriz $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ se dice unitariamente diagonalizable si $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U}$ es diagonal para alguna matriz unitaria $\boldsymbol{U}$. Como sugiere el ejemplo 2.25, vamos a demostrar que cada matriz hermitiana es unitariamente diagonalizable. Sin embargo, con solo un poco de esfuerzo adicional, podemos obtener un teorema muy importante que tiene este resultado como una consecuencia directa.

Una matriz compleja se llama triangular superior si cada entrada por debajo de la diagonal principal es cero. Debemos el siguiente teorema a Issai Schur.[3]

Teorema 2.26 (Teorema de Schur) *Si* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *es cualquier matriz compleja, existe una matriz unitaria* $\boldsymbol{U}$ *tal que*

$$\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U} = \boldsymbol{T}$$

es triangular superior. Además, las entradas en la diagonal principal de $\boldsymbol{T}$ *son los valores propios* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *de* $\boldsymbol{A}$ *(incluyendo multiplicidades).*

Demostración: *Usamos inducción sobre* n. *Si* $n = 1$, $\boldsymbol{A}$ *ya es triangular superior. Si* $n > 1$, *asumimos que el teorema es válido para matrices complejas de* $(n-1)\times(n-1)$. *Sea* λ_1 *un valor propio de* $\boldsymbol{A}$, *y sea* $\boldsymbol{y}_1$ *un vector propio con* $\|\boldsymbol{y}_1\| = 1$. *Entonces* $\boldsymbol{y}_1$ *es parte de una base de* $\mathbb{C}^n$ *(por el análogo del Teorema*

[3] Issai Schur (1875-1941) fue un matemático alemán que realizó trabajos fundamentales en la teoría de representaciones de grupos como matrices.

Versión en revisión

V. Osores

2.7), así que el proceso de Gram-Schmidt (análogo complejo) proporciona $\boldsymbol{y}_2, \ldots, \boldsymbol{y}_n$ de manera que $\{\boldsymbol{y}_1, \boldsymbol{y}_2, \ldots, \boldsymbol{y}_n\}$ es una base ortonormal de $\mathbb{C}^n$. Si $\boldsymbol{U}_1 = \begin{bmatrix} \boldsymbol{y}_1 & \boldsymbol{y}_2 & \cdots & \boldsymbol{y}_n \end{bmatrix}$ es la matriz con estos vectores como sus columnas, entonces (ver Lema 2.1)

$$\boldsymbol{U}_1^* \boldsymbol{A} \boldsymbol{U}_1 = \begin{bmatrix} \lambda_1 & X_1 \\ 0 & A_1 \end{bmatrix}$$

en forma de bloques. Ahora aplicamos la hipótesis de inducción para encontrar una matriz unitaria de orden $(n-1) \times (n-1)$ $\boldsymbol{W}_1$ tal que $\boldsymbol{W}_1^* \boldsymbol{A}_1 \boldsymbol{W}_1 = \boldsymbol{T}_1$ sea triangular superior. Entonces $\boldsymbol{U}_2 = \begin{bmatrix} 1 & 0 \\ 0 & W_1 \end{bmatrix}$ es una matriz unitaria de orden $n \times n$. Por lo tanto, $\boldsymbol{U} = \boldsymbol{U}_1 \boldsymbol{U}_2$ es unitaria (usando el Teorema 2.25), y

$$\begin{aligned} \boldsymbol{U}^* \boldsymbol{A} \boldsymbol{U} =& \boldsymbol{U}_2^* \left(\boldsymbol{U}_1^* \boldsymbol{A} \boldsymbol{U}_1 \right) \boldsymbol{U}_2 \\ =& = \begin{bmatrix} 1 & 0 \\ 0 & \boldsymbol{W}_1^* \end{bmatrix} \begin{bmatrix} \lambda_1 & \boldsymbol{X}_1 \\ 0 & \boldsymbol{A}_1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & \boldsymbol{W}_1 \end{bmatrix} = \begin{bmatrix} \lambda_1 & \boldsymbol{X}_1 \boldsymbol{W}_1 \\ 0 & \boldsymbol{T}_1 \end{bmatrix} \end{aligned}$$

es triangular superior. Finalmente, $\boldsymbol{A}$ y $\boldsymbol{U}^* \boldsymbol{A} \boldsymbol{U} = \boldsymbol{T}$ tienen los mismos valores propios porLema 2.3, y son los elementos diagonales de $\boldsymbol{T}$ porque $\boldsymbol{T}$ es triangular superior. ■

El hecho de que las matrices similares tengan la misma traza y determinante da la siguiente consecuencia del teorema de Schur.

Corolario 2.2 *Sea $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ una matriz compleja, y sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ los valores propios de $\boldsymbol{A}$, incluyendo multiplicidades. Entonces*

$$\det A = \lambda_1 \lambda_2 \cdots \lambda_n \quad \text{y} \quad \operatorname{tr} A = \lambda_1 + \lambda_2 + \cdots + \lambda_n.$$

El teorema de Schur afirma que toda matriz compleja puede ser "triangularizada unitariamente". Sin embargo, no podemos sustituir "diagonalizada unitariamente" aquí. De hecho, si $\boldsymbol{A} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix}$, no existe ninguna matriz compleja invertible $\boldsymbol{U}$ tal que $\boldsymbol{U}^{-1} \boldsymbol{A} \boldsymbol{U}$ sea diagonal. Sin embargo, la situación es mucho mejor para las matrices hermitianas.

Teorema 2.27 (Teorema Espectral) *Si $\boldsymbol{A}$ es hermitiana, existe una matriz unitaria $\boldsymbol{U}$ tal que $\boldsymbol{U}^* \boldsymbol{A} \boldsymbol{U}$ es diagonal.*

Demostración: *Por el teorema de Schur, sea $\boldsymbol{U}^* \boldsymbol{A} \boldsymbol{U} = \boldsymbol{T}$ triangular superior donde $\boldsymbol{U}$ es unitaria. Dado que $\boldsymbol{A}$ es hermitiana, esto da*

$$\boldsymbol{T}^* = \left(\boldsymbol{U}^* \boldsymbol{A} \boldsymbol{U} \right)^* = \boldsymbol{U}^* \boldsymbol{A}^* \boldsymbol{U}^{**} = \boldsymbol{U}^* \boldsymbol{A} \boldsymbol{U} = \boldsymbol{T}.$$

Esto significa que $\boldsymbol{T}$ es tanto triangular superior como inferior. Por lo tanto, T es en realidad diagonal.

El teorema de los ejes principales afirma que una matriz real $\boldsymbol{A}$ es simétrica si y sólo si es diagonalizable ortogonalmente (es decir, $\boldsymbol{P}^T \boldsymbol{A} \boldsymbol{P}$ es diagonal para alguna matriz ortogonal real $\boldsymbol{P}$). El Teorema 2.27 es el análogo complejo de la mitad de este resultado. Sin embargo, el inverso es falso para matrices complejas: Existen matrices diagonalizables unitariamente que no son hermitianas.

Ejemplo 2.26 *Muestre que la matriz no hermitiana* $\boldsymbol{A} = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ *es diagonalizable unitariamente.*

Sol: *El polinomio característico es* $f_{\boldsymbol{A}}(x) = x^2 + 1$. *Por lo tanto, los valores propios son* i *y* $-i$, *y es fácil verificar que* $\begin{bmatrix} i \\ -1 \end{bmatrix}$ *y* $\begin{bmatrix} -1 \\ i \end{bmatrix}$ *son los vectores propios correspondientes. Además, estos vectores propios son ortogonales y ambos tienen longitud* $\sqrt{2}$, *por lo que* $\boldsymbol{U} = \frac{1}{\sqrt{2}} \begin{bmatrix} i & -1 \\ -1 & i \end{bmatrix}$ *es una matriz unitaria tal que* $U^*AU = \begin{bmatrix} i & 0 \\ 0 & -i \end{bmatrix}$ *es diagonal.*

Hay una manera muy simple de caracterizar aquellas matrices complejas que son diagonalizables unitariamente. Con este fin, una matriz compleja $\boldsymbol{N}$ de $n \times n$ se llama normal si $\boldsymbol{N}\boldsymbol{N}^* = \boldsymbol{N}^*\boldsymbol{N}$. Está claro que toda matriz hermitiana o unitaria es normal, al igual que la matriz $\begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}$ en el ejemplo 2.26. De hecho, tenemos el siguiente resultado.

Teorema 2.28 *Una matriz compleja* $\boldsymbol{A} \in \mathbb{C}^{n \times n}$ *es diagonalizable unitariamente si y sólo si* $\boldsymbol{A}$ *es normal.*

Demostración: *Supongamos primero que* $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U} = \boldsymbol{D}$, *donde* $\boldsymbol{U}$ *es unitaria y* $\boldsymbol{D}$ *es diagonal. Entonces* $\boldsymbol{D}\boldsymbol{D}^* = \boldsymbol{D}^*\boldsymbol{D}$ *como se puede verificar fácilmente. Debido a que* $\boldsymbol{D}\boldsymbol{D}^* = \boldsymbol{U}^*(\boldsymbol{A}\boldsymbol{A}^*)\boldsymbol{U}$ *y* $\boldsymbol{D}^*\boldsymbol{D} = \boldsymbol{U}^*(\boldsymbol{A}^*\boldsymbol{A})\boldsymbol{U}$, *se sigue por cancelación que* $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A}$.

Por el contrario, supongamos que $\boldsymbol{A}$ *es normal, es decir,* $\boldsymbol{A}\boldsymbol{A}^* = \boldsymbol{A}^*\boldsymbol{A}$. *Por el teorema de Schur, sea* $\boldsymbol{U}^*\boldsymbol{A}\boldsymbol{U} = \boldsymbol{T}$, *donde* $\boldsymbol{T}$ *es triangular superior y* $\boldsymbol{U}$ *es unitaria. Entonces* $\boldsymbol{T}$ *también es normal:*

$$\boldsymbol{T}\boldsymbol{T}^* = \boldsymbol{U}^*(\boldsymbol{A}\boldsymbol{A}^*)\boldsymbol{U} = \boldsymbol{U}^*(\boldsymbol{A}^*\boldsymbol{A})\boldsymbol{U} = \boldsymbol{T}^*\boldsymbol{T}.$$

Por lo tanto, basta mostrar que una matriz triangular superior normal $\boldsymbol{T}$ *de* $n \times n$ *debe ser diagonal. Procedemos por inducción en* n; *está claro para* $n = 1$. *Si* $n > 1$ *y* $\boldsymbol{T} = [t_{ij}]$, *entonces al igualar las entradas* $(1,1)$ *en* $\boldsymbol{T}\boldsymbol{T}^*$ *y* $\boldsymbol{T}^*\boldsymbol{T}$ *se obtiene*

$$|t_{11}|^2 + |t_{12}|^2 + \cdots + |t_{1n}|^2 = |t_{11}|^2.$$

Esto implica $t_{12} = t_{13} = \cdots = t_{1n} = 0$, *por lo que* $\boldsymbol{T} = \begin{bmatrix} t_{11} & 0 \\ 0 & T_1 \end{bmatrix}$ *en forma de bloque. Por lo tanto,* $\boldsymbol{T}^* = \begin{bmatrix} \bar{t}_{11} & 0 \\ 0 & \boldsymbol{T}_1^* \end{bmatrix}$ *así que* $\boldsymbol{T}\boldsymbol{T}^* = \boldsymbol{T}^*\boldsymbol{T}$ *implica* $\boldsymbol{T}_1\boldsymbol{T}_1^* = \boldsymbol{T}_1\boldsymbol{T}_1^*$. *Por lo tanto,* $\boldsymbol{T}_1$ *es diagonal por inducción, y la demostración está completa.*

Concluimos esta sección usando el teorema de Schur (Teorema 2.26) para demostrar un teorema famoso sobre matrices. Recordemos que el polinomio característico de una matriz cuadrada $\boldsymbol{A}$ se define por $f_{\boldsymbol{A}}(x) = \det(x\boldsymbol{I} - \boldsymbol{A})$, y que los valores propios de $\boldsymbol{A}$ son las raíces de $f_{\boldsymbol{A}}(x)$.

Teorema 2.29 (Teorema de Cayley-Hamilton) [a] *Si* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ *es una matriz compleja, entonces* $f_{\boldsymbol{A}}(\boldsymbol{A}) = 0$*; es decir,* $\boldsymbol{A}$ *es una raíz de su polinomio característico.*

[a] Nombrado en honor a los matemáticos ingleses Arthur Cayley (1821-1895) y William Rowan Hamilton (1805-1865), un matemático irlandés famoso por su trabajo sobre la dinámica física.

Demostración: *Si* $p(x)$ *es cualquier polinomio con coeficientes complejos, entonces* $p\left(\boldsymbol{P}^{-1}\boldsymbol{A}\boldsymbol{P}\right) = \boldsymbol{P}^{-1}p(\boldsymbol{A})\boldsymbol{P}$ *para cualquier matriz compleja invertible* $\boldsymbol{P}$*. Por lo tanto, por el teorema de Schur, podemos suponer que* $\boldsymbol{A}$ *es triangular superior. Entonces los valores propios* $\lambda_1, \lambda_2, \ldots, \lambda_n$ *de* $\boldsymbol{A}$ *aparecen a lo largo de la diagonal principal, por lo que*

$$f_{\boldsymbol{A}}(x) = (x-\lambda_1)(x-\lambda_2)(x-\lambda_3)\cdots(x-\lambda_n).$$

Así,

$$f_{\boldsymbol{A}}(\boldsymbol{A}) = (\boldsymbol{A}-\lambda_1\boldsymbol{I})(\boldsymbol{A}-\lambda_2\boldsymbol{I})(\boldsymbol{A}-\lambda_3\boldsymbol{I})\cdots(\boldsymbol{A}-\lambda_n\boldsymbol{I}).$$

Note que cada matriz $\boldsymbol{A}-\lambda_i\boldsymbol{I}$ *es triangular superior. Ahora observe:*

1. $\boldsymbol{A}-\lambda_1\boldsymbol{I}$ *tiene la primera columna cero porque la columna 1 de* $\boldsymbol{A}$ *es* $(\lambda_1, 0, 0, \ldots, 0)^T$.

2. *Entonces* $(\boldsymbol{A}-\lambda_1\boldsymbol{I})(\boldsymbol{A}-\lambda_2\boldsymbol{I})$ *tiene las primeras dos columnas cero porque la segunda columna de* $(\boldsymbol{A}-\lambda_2\boldsymbol{I})$ *es* $(b, 0, 0, \ldots, 0)^T$ *para alguna constante* b.

3. *A continuación,* $(\boldsymbol{A}-\lambda_1\boldsymbol{I})(\boldsymbol{A}-\lambda_2\boldsymbol{I})(\boldsymbol{A}-\lambda_3\boldsymbol{I})$ *tiene las primeras tres columnas cero porque la columna 3 de* $(\boldsymbol{A}-\lambda_3\boldsymbol{I})$ *es* $(c, d, 0, \ldots, 0)^T$ *para algunas constantes* c *y* d.

Continuando de esta manera, vemos que $(\boldsymbol{A}-\lambda_1\boldsymbol{I})(\boldsymbol{A}-\lambda_2\boldsymbol{I})(\boldsymbol{A}-\lambda_3\boldsymbol{I})\cdots(\boldsymbol{A}-\lambda_n\boldsymbol{I})$ *tiene todas las* n *columnas cero; es decir,* $f_{\boldsymbol{A}}(A) = 0$.

2.5 Normas de vectores y matrices

Definición 2.19 *Sea* $\boldsymbol{A} \in \mathbb{C}^{n\times n}$*. Se define el espectro de* $\boldsymbol{A}$*, denotado* $\sigma(\boldsymbol{A})$*, como el conjunto de todos los valores propios de* $\boldsymbol{A}$*. Además, se llama radio espectral de* $\boldsymbol{A}$ *a*

$$r_\sigma(\boldsymbol{A}) := \max_{\lambda\in\sigma(\boldsymbol{A})} |\lambda|.$$

Definición 2.20 *Sea* V *un espacio vectorial sobre el cuerpo* $\mathbb{C}$*. Se llama norma de vector a toda aplicación* $\|\cdot\| : V \to \mathbb{R}_0^+$ *tal que para todo* $\boldsymbol{x}, \boldsymbol{y} \in V$ *y* $\lambda \in \mathbb{C}$ *se verifica:*

1. $\|\boldsymbol{x}\| > 0$ *si* $\boldsymbol{x} \neq \boldsymbol{0}$ *y* $\|\boldsymbol{x}\| = 0$ *si y sólo si* $\boldsymbol{x} = \boldsymbol{0}$.
2. $\|\lambda\boldsymbol{x}\| = |\lambda|\|\boldsymbol{x}\|$.
3. $\|\boldsymbol{x}+\boldsymbol{y}\| \leqslant \|\boldsymbol{x}\| + \|\boldsymbol{y}\|$.

V. Osores

Damos a continuación algunos ejemplos de normas para el espacio $\mathbb{C}^n$:

$$\|\boldsymbol{x}\|_1 := \sum_{i=1}^{n} |x_i|$$
$$\|\boldsymbol{x}\|_2 := (\boldsymbol{x}^*\boldsymbol{x})^{1/2} = \left(\sum_{i=1}^{n} |x_i|^2\right)^{1/2}$$
$$\|\boldsymbol{x}\|_\infty := \max_{1 \leqslant i \leqslant n} |x_i|$$

a las que nos referimos como "norma 1", "norma 2" y "norma ∞", respectivamente; en general, para $p \in [1, \infty)$ definimos la "norma p"

$$\|\boldsymbol{x}\|_p := \left(\sum_{i=1}^{n} |x_i|^p\right)^{1/p}.$$

Teorema 2.30 *Una norma $\|\cdot\| : \mathbb{K}^n \to \mathbb{R}_0^+$ es una función continua.*

Demostración: *Tomamos en cuenta que para $\boldsymbol{x} = (\xi_1, \dots, \xi_n)^{\boldsymbol{T}}, \boldsymbol{y} = (\eta_1, \dots, \eta_n)^{\boldsymbol{T}}$*

$$\begin{aligned} \left|\|\boldsymbol{x}\| - \|\boldsymbol{y}\|\right| &= \left|\|\boldsymbol{x}\| - \|-\boldsymbol{y}\|\right| \leqslant \|\boldsymbol{x} - \boldsymbol{y}\| \\ &\leqslant \sum_{i=1}^{n} |\xi_i - \eta_i| \, \|\boldsymbol{e}_i\| \leqslant n \max_{1 \leqslant i \leqslant n} \|\boldsymbol{e}_i\| \max_{1 \leqslant i \leqslant n} |\xi_i - \eta_i| \end{aligned}$$

y aplicamos la definición de la continuidad. ■

Teorema 2.31 *Si $\|\cdot\|$ y $\|\cdot\|^*$ son dos normas sobre $\mathbb{K}^n$, entonces existen constantes $m, M > 0$ tales que*

$$\forall \boldsymbol{x} \in \mathbb{K}^n : \quad m\|\boldsymbol{x}\| \leqslant \|\boldsymbol{x}\|^* \leqslant M\|\boldsymbol{x}\|. \tag{2.5.1}$$

Demostración: *Sean $\|\cdot\| := \|\cdot\|_\infty$ y $\|\cdot\|^*$ arbitraria (el caso general sigue por transitividad). Sea*

$$\mathcal{S} := \left\{ \boldsymbol{x} \in \mathbb{K}^n | \boldsymbol{x} = (\xi_1, \dots, \xi_n)^{\boldsymbol{T}}, \max_{1 \leqslant i \leqslant n} |\xi_i| = 1 \right\}$$

El conjunto $\mathcal{S}$ es compacto. Como $\|\cdot\|^*$ es continua, existen $\boldsymbol{x}_m, \boldsymbol{x}_M \in \mathcal{S}$ tales que

$$\|\boldsymbol{x}_m\|^* = \min_{\boldsymbol{x} \in \mathcal{S}} \|\boldsymbol{x}\|^* =: m, \quad \|\boldsymbol{x}_M\|^* = \max_{\boldsymbol{x} \in \mathcal{S}} \|\boldsymbol{x}\|^* =: M$$

luego

$$\forall \boldsymbol{x} \in \mathbb{K}^n, \quad \boldsymbol{x} \neq \boldsymbol{0} : \quad m \leqslant \left\| \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|_\infty} \right\|^* \leqslant M$$

lo que implica (2.5.1) en virtud de la homogeneidad (item 2. en la Definición 2.20). (Para $\boldsymbol{x} = \boldsymbol{0}$, (2.5.1) es trivial.) ■

Definición 2.21 *En el espacio $\mathbb{C}^{n \times n}$ se llama norma de matriz a toda aplicación $\|\cdot\| : \mathbb{C}^{n \times n} \to \mathbb{R}_0^+$ tal que para todas matrices $\boldsymbol{A}, \boldsymbol{B} \in \mathbb{C}^{n \times n}$ y todo $\lambda \in \mathbb{C}$ se verifica:*

1. $\|\boldsymbol{A}\| > 0$ si $\boldsymbol{A} \neq \boldsymbol{0}$ y $\|\boldsymbol{A}\| = 0$ si y sólo si $\boldsymbol{A} = \boldsymbol{0}$.

2. $\|\lambda \boldsymbol{A}\| = |\lambda| \|\boldsymbol{A}\|$.
3. $\|\boldsymbol{A}+\boldsymbol{B}\| \leqslant \|\boldsymbol{A}\| + \|B\|$.
4. $\|\boldsymbol{AB}\| \leq \|\boldsymbol{A}\| \|\boldsymbol{B}\|$.

En correspondencia con cada norma vectorial $\|\cdot\|$ de $\mathbb{C}^n$, se define una norma para matrices $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ por medio de la expresión

$$\|\boldsymbol{A}\| := \max_{\boldsymbol{x}\neq 0} \frac{\|\boldsymbol{Ax}\|}{\|\boldsymbol{x}\|} = \max_{\|\boldsymbol{x}\|=1} \|\boldsymbol{Ax}\|. \tag{2.5.2}$$

Esta norma matricial se dice inducida por la norma vectorial. En particular, las normas vectoriales 1,2 e ∞ inducen las siguientes normas matriciales, a las cuales igualmente nos referimos como “norma 1”, “norma 2” y “norma ∞”, respectivamente:

$$\begin{aligned}
\|\boldsymbol{A}\|_1 &:= \max_{\|\boldsymbol{x}\|_1=1} \|\boldsymbol{Ax}\|_1 := \max_{1\leqslant j\leqslant n} \sum_{i=1}^{n} |a_{ij}| \\
\|\boldsymbol{A}\|_2 &:= \max_{\|\boldsymbol{x}\|_2=1} \|\boldsymbol{Ax}\|_2 \\
\|\boldsymbol{A}\|_\infty &:= \max_{\|\boldsymbol{x}\|_\infty=1} \|\boldsymbol{Ax}\|_\infty = \max_{1\leqslant i\leqslant n} \sum_{j=1}^{n} |a_{ij}|
\end{aligned}$$

También se define sobre $\mathbb{C}^{n\times n}$ la siguiente norma, llamada norma de Frobenius, la cual no es inducida por una norma vectorial:

$$\|\boldsymbol{A}\|_{\boldsymbol{F}} := \left(\sum_{i,j=1}^{n} |a_{ij}|^2 \right)^{1/2}.$$

Definición 2.22 *Una norma de matriz se dice compatible con una norma de vector si para cada $\boldsymbol{A} \in \mathbb{C}^{n\times n}$ y para cada $\boldsymbol{x} \in \mathbb{C}^n$,*

$$\|\boldsymbol{Ax}\|_{vector} \leqslant \|\boldsymbol{A}\|_{matriz} \|\boldsymbol{x}\|_{vector}.$$

De acuerdo a la definición 2.20, cada norma matricial inducida por una norma vectorial es compatible con la norma vectorial que la induce. En particular, las normas matriciales 1, 2 e ∞ son compatibles con las correspondientes normas vectoriales 1,2 e ∞. Por otra parte, la norma de Frobenius, que como indicamos no es inducida por norma vectorial alguna, es compatible con la norma vectorial 2.

Teorema 2.32 *Si $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, entonces*

$$\|\boldsymbol{A}\|_2 = \sqrt{r_\sigma (\boldsymbol{A}^*\boldsymbol{A})}.$$

Demostración: *Es claro que $\boldsymbol{A}^*\boldsymbol{A}$ es una matriz hermitiana. Por el Teorema del Eje Principal sabemos que $\boldsymbol{A}^*\boldsymbol{A}$ tiene n vectores propios que forman una base ortonormal de $\mathbb{C}^n$. Veremos a continuación que los valores propios de $\boldsymbol{A}^*\boldsymbol{A}$ son además no negativos. En efecto, si λ es un valor propio de $\boldsymbol{A}^*\boldsymbol{A}$ y $\boldsymbol{v}$ es un vector propio asociado, entonces*

$$A^*\boldsymbol{Av} = \lambda \boldsymbol{v}$$

y además,

$$\|\boldsymbol{A}\boldsymbol{v}\|_2^2 = (\boldsymbol{A}\boldsymbol{v})^*(\boldsymbol{A}\boldsymbol{v}) = \boldsymbol{v}^*\left(\boldsymbol{A}^*\boldsymbol{A}\right)\boldsymbol{v} = \boldsymbol{v}^*(\lambda\boldsymbol{v}) = \lambda\|\boldsymbol{v}\|_2^2$$

Como $\|\boldsymbol{v}\| \neq 0$*, esta última relación implica que*

$$\lambda = \frac{\|\boldsymbol{A}\boldsymbol{v}\|_2^2}{\|\boldsymbol{v}\|_2^2} \geqslant 0 \tag{2.5.3}$$

Ahora sean $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \geqslant \lambda_n$ *los valores propios de* $\boldsymbol{A}^*\boldsymbol{A}$ *y* $\left\{\boldsymbol{v}^{(1)}, \boldsymbol{v}^{(2)}, \ldots, \boldsymbol{v}^{(n)}\right\}$ *un conjunto de vectores propios asociados que forman una base ortonormal de* $\mathbb{C}^n$*. Entonces, para* $\boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\}$ *existen escalares* $\alpha_1, \ldots, \alpha_n$ *tales que*

$$\boldsymbol{x} = \sum_{j=1}^{n} \alpha_j \boldsymbol{v}^{(j)} \tag{2.5.4}$$

Por otro lado,

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = (\boldsymbol{A}\boldsymbol{x})^*\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}^*\left(\boldsymbol{A}^*\boldsymbol{A}\right)\boldsymbol{x} \tag{2.5.5}$$

Reemplazando (2.5.4) *en* (2.5.5) *y reordenando,*

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \left(\sum_{j=1}^{n} \bar{\alpha}_j \boldsymbol{v}^{(j)*}\right)\left(\boldsymbol{A}^*\boldsymbol{A}\right)\left(\sum_{i=1}^{n} \alpha_i \boldsymbol{v}^{(i)}\right) = \sum_{j=1}^{n}\sum_{i=1}^{n} \lambda_i \bar{\alpha}_j \alpha_i \boldsymbol{v}^{(j)*}\boldsymbol{v}^{(i)}$$

y como $\boldsymbol{v}^{(j)*}\boldsymbol{v}^{(i)} = \delta_{ij}$*, entonces*

$$\|\boldsymbol{A}\boldsymbol{x}\|_2^2 = \sum_{i=1}^{n} |\alpha_i|^2 \lambda_i \leqslant \lambda_1 \sum_{i=1}^{n} |\alpha_i|^2 \tag{2.5.6}$$

En forma análoga calculamos

$$\|\boldsymbol{x}\|_2^2 = \boldsymbol{x}^*\boldsymbol{x} = \sum_{j=1}^{n}\sum_{i=1}^{n} \bar{\alpha}_i \alpha_j \boldsymbol{v}^{(j)*}\boldsymbol{v}^{(i)} = \sum_{i=1}^{n} |\alpha_i|^2 \tag{2.5.7}$$

De (2.5.6) *y* (2.5.7) *concluimos que*

$$\forall \boldsymbol{x} \in \mathbb{C}^n \backslash \{\boldsymbol{0}\}: \quad \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leqslant \sqrt{\lambda_1}$$

lo cual equivale a

$$\|\boldsymbol{A}\|_2 := \max_{\boldsymbol{x}\neq 0} \frac{\|\boldsymbol{A}\boldsymbol{x}\|_2}{\|\boldsymbol{x}\|_2} \leqslant \sqrt{\lambda_1} \tag{2.5.8}$$

Para mostrar que la cota $\sqrt{\lambda_1}$ *se alcanza, basta exhibir un vector no nulo para el cual la igualdad se cumpla en* (2.5.8)*. Con esta finalidad sea* $\boldsymbol{v}_1$ *un vector propio asociado a* λ_1*. Entonces, a partir de* (2.5.3) *obtenemos inmediatamente que*

$$\sqrt{\lambda_1} = \frac{\|\boldsymbol{A}\boldsymbol{v}_1\|_2}{\|\boldsymbol{v}_1\|_2}$$

esto es, el máximo de (2.5.8) *se alcanza en* $\boldsymbol{x} = \boldsymbol{v}_1$ *y es igual a* $\sqrt{\lambda_1}$*. Notando que* $\lambda_1 = r_\sigma\left(\boldsymbol{A}^*\boldsymbol{A}\right)$ *concluimos la demostración.* ■

Corolario 2.3 *Si $\boldsymbol{A}$ es hermitiana, entonces $\|\boldsymbol{A}\|_2 = r_\sigma(\boldsymbol{A})$.*

Demostración: *Como $\|\boldsymbol{A}\|_2 = \sqrt{r_\sigma(\boldsymbol{A}^*\boldsymbol{A})}$ y $\boldsymbol{A}^* = \boldsymbol{A}$, entonces $\|\boldsymbol{A}\|_2 = \sqrt{r_\sigma(\boldsymbol{A}^2)}$. En virtud de*

$$r_\sigma\left(\boldsymbol{A}^2\right) = \left(r_\sigma(\boldsymbol{A})\right)^2$$

inmediatamente se llega a

$$\|\boldsymbol{A}\|_2 = \sqrt{\left(r_\sigma(\boldsymbol{A})\right)^2} = r_\sigma(\boldsymbol{A})$$

que es lo que se quería demostrar. ∎

Teorema 2.33 *Sea $\|\cdot\|$ alguna norma vectorial sobre $\mathbb{C}^n$ y $\|\cdot\|$ la norma matricial inducida. En este caso,*

$$\forall \boldsymbol{B} \in \mathbb{C}^{n\times n}: \quad r_\sigma(\boldsymbol{B}) \leqslant \|\boldsymbol{B}\|.$$

Demostración: *Sea λ un valor propio de $\boldsymbol{B}$ con $|\lambda| = r_\sigma(\boldsymbol{B})$ y $\boldsymbol{x} \neq \boldsymbol{0}$ un vector propio asociado. Entonces*

$$\|\lambda\boldsymbol{x}\| = |\lambda|\|\boldsymbol{x}\| = r_\sigma(\boldsymbol{B})\|\boldsymbol{x}\| = \|\boldsymbol{B}\boldsymbol{x}\| \leqslant \|\boldsymbol{B}\|\|\boldsymbol{x}\|,$$

es decir,

$$\|\boldsymbol{B}\| \geqslant \frac{\|\boldsymbol{B}\boldsymbol{x}\|}{\|\boldsymbol{x}\|} = r_\sigma(\boldsymbol{B}).$$

∎

Teorema 2.34 *Sea $\boldsymbol{B} \in \mathbb{C}^{n\times n}$ una matriz arbitraria, con $\varepsilon > 0$ arbitrario dado. Entonces existe una norma vectorial $\|\cdot\|_B$ sobre $\mathbb{C}^n$ tal que para la norma matricial asociada,*

$$\|\boldsymbol{B}\|_{\boldsymbol{B}} \leqslant r_\sigma(\boldsymbol{B}) + \varepsilon.$$

Demostración: *Según el Teorema **??** (sobre la forma normal de Schur), existe una matriz $\boldsymbol{U}$ unitaria tal que*

$$\boldsymbol{U}^*\boldsymbol{B}\boldsymbol{U} = \begin{bmatrix} \lambda_1 & * & \cdots & * \\ 0 & \ddots & \ddots & \vdots \\ \vdots & \ddots & \ddots & * \\ 0 & \cdots & 0 & \lambda_n \end{bmatrix} =: (\varrho_{ik}).$$

Para los elementos arriba de la diagonal (ϱ_{ik} con $k > i$) sabemos que

$$\varrho_{ik} = \sum_{s=1}^{n}\sum_{l=1}^{n} \bar{\nu}_{li}\beta_{ls}\nu_{sk}, \quad \boldsymbol{U} =: (\nu_{sk})$$

Eso significa

$$|\varrho_{ik}| \leqslant n^2\beta, \quad \beta := \max_{1\leqslant l,s\leqslant n} |\beta_{ls}|.$$

Sean

$$\delta := \min\left\{\frac{\varepsilon}{n^3(\beta+1)}, 1\right\}, \quad \boldsymbol{D} := \operatorname{diag}\left(1, \delta, \ldots, \delta^{n-1}\right).$$

En este caso,

$$\boldsymbol{D}^{-1}\boldsymbol{U}^*\boldsymbol{B}\boldsymbol{U}\boldsymbol{D} = \left(\varrho_{ik}\delta^{k-i}\right),$$

entonces

$$\begin{aligned}
\left\|\boldsymbol{D}^{-1}\boldsymbol{U}^*\boldsymbol{B}\boldsymbol{U}\boldsymbol{D}\right\|_\infty &= \max_{1\leqslant i\leqslant n}\sum_{k=i}^{n}\left|\varrho_{ik}\delta^{k-i}\right| \\
&\leqslant \max_{1\leqslant i\leqslant n}|\varrho_{ii}| + \delta\max_{1\leqslant i\leqslant n}\sum_{k=i+1}^{n}|\varrho_{ik}| \\
&\leqslant r_\sigma(\boldsymbol{B}) + \delta(n-1)n^2\beta \\
&\leqslant r_\sigma(\boldsymbol{B}) + \varepsilon\frac{n^2(n-1)\beta}{n^3(\beta+1)} < r_\sigma(\boldsymbol{B}) + \varepsilon
\end{aligned}$$

Por otro lado,

$$\left\|\boldsymbol{D}^{-1}\boldsymbol{U}^*\boldsymbol{B}\boldsymbol{U}\boldsymbol{D}\right\|_\infty = \max_{\boldsymbol{x}\neq 0}\frac{\left\|\boldsymbol{D}^{-1}\boldsymbol{U}^*\boldsymbol{B}\boldsymbol{U}\boldsymbol{D}\boldsymbol{x}\right\|_\infty}{\|\boldsymbol{x}\|_\infty} = \max_{\boldsymbol{y}\neq 0}\frac{\left\|\boldsymbol{D}^{-1}\boldsymbol{U}^*\boldsymbol{B}\boldsymbol{y}\right\|_\infty}{\left\|\boldsymbol{D}^{-1}\boldsymbol{U}^*\boldsymbol{y}\right\|_\infty}$$

Entonces la norma deseada es $\|\boldsymbol{x}\|_B := \left\|\boldsymbol{D}^{-1}\boldsymbol{U}^*\boldsymbol{x}\right\|_\infty$. ■

CAPÍTULO 3

Formas Canónicas

3.1 Preliminares

3.1.1 Forma normal de Smith

Supongamos que una matriz A de orden $m \times n$ es llevada a una matriz B (escrito $A \to B$) mediante una serie de k operaciones elementales por filas. Denotemos por $E_1, E_2, \ldots, E_k$ las matrices elementales correspondientes. La reducción se convierte en

$$A \to E_1 A \to E_2 E_1 A \to E_3 E_2 E_1 A \to \cdots \to E_k E_{k-1} \cdots E_2 E_1 A = B.$$

En otras palabras,

$$A \to UA = B \quad \text{donde} \quad U = E_k E_{k-1} \cdots E_2 E_1$$

La matriz $U = E_k E_{k-1} \cdots E_2 E_1$ es invertible, siendo un producto de matrices invertibles. Además, U se puede calcular sin encontrar los E_i de la siguiente manera: si la serie de operaciones mencionada que lleva $A \to B$ se realiza en I_m en lugar de A, el resultado es $I_m \to UI_m = U$. Por lo tanto, esta serie de operaciones lleva la matriz bloque $\begin{bmatrix} A & I_m \end{bmatrix} \to \begin{bmatrix} B & U \end{bmatrix}$. Esto, junto con la discusión anterior, prueba el Teorema 3.1

Teorema 3.1 *Supongamos que A es de orden $m \times n$ y $A \to B$ mediante operaciones elementales por filas.*

1. *$B = UA$ donde U es una matriz invertible $m \times m$.*
2. *U se puede calcular mediante $\begin{bmatrix} A & I_m \end{bmatrix} \to \begin{bmatrix} B & U \end{bmatrix}$ utilizando las operaciones que llevan $A \to B$.*
3. *$U = E_k E_{k-1} \cdots E_2 E_1$ donde $E_1, E_2, \ldots, E_k$ son las matrices elementales correspondientes (en orden) a las operaciones elementales por filas que llevan A a B.*

Ejemplo 3.1 *Sea $A = \begin{bmatrix} 2 & 3 & 1 \\ 1 & 2 & 1 \end{bmatrix}$. Exprese la forma escalonada reducida por filas R de A tal que*

$R = UA$ donde U es invertible.

Sol: *Reduce la matriz doble* $\begin{bmatrix} A & I \end{bmatrix} \to \begin{bmatrix} R & U \end{bmatrix}$ *de la siguiente manera:*

$$\begin{bmatrix} A & I \end{bmatrix} = \left[\begin{array}{ccc|cc} 2 & 3 & 1 & 1 & 0 \\ 1 & 2 & 1 & 0 & 1 \end{array}\right] \to \left[\begin{array}{ccc|cc} 1 & 2 & 1 & 0 & 1 \\ 2 & 3 & 1 & 1 & 0 \end{array}\right] \to \left[\begin{array}{ccc|cc} 1 & 2 & 1 & 0 & 1 \\ 0 & -1 & -1 & 1 & -2 \end{array}\right]$$
$$\to \left[\begin{array}{ccc|cc} 1 & 0 & -1 & 2 & -3 \\ 0 & 1 & 1 & -1 & 2 \end{array}\right]$$

Por lo tanto, $R = \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 1 \end{bmatrix}$ *y* $U = \begin{bmatrix} 2 & -3 \\ -1 & 2 \end{bmatrix}$.

Sea A de orden $m \times n$, de rango r, y sea R la forma escalonada reducida por filas de A. El Teorema 3.1 muestra que $R = UA$ donde U es invertible, y que U se puede encontrar a partir de $\begin{bmatrix} A & I_m \end{bmatrix} \to \begin{bmatrix} R & U \end{bmatrix}$.

La matriz R tiene r unos principales (ya que el rango de A es r), por lo que, como R es reducida, la matriz de orden $n \times m$, R^T contiene cada fila de I_r en las primeras r columnas. Por lo tanto, las operaciones por filas llevarán $R^T \to \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n\times m}$. Por lo tanto, el Teorema 3.1 (nuevamente) muestra que $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n\times m} = U_1 R^T$ donde U_1 es una matriz $n \times n$ invertible. Escribiendo $V = U_1^T$, obtenemos

$$UAV = RV = RU_1^T = \left(U_1 R^T\right)^T = \left(\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n\times m}\right)^T = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}$$

Además, la matriz $U_1 = V^T$ se puede calcular mediante $\begin{bmatrix} R^T & I_n \end{bmatrix} \to \left[\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n\times m} V^T\right]$. Esto prueba el Teorema 3.2

Teorema 3.2 *Sea A una matriz $m \times n$ de rango r. Existen matrices invertibles U y V de tamaño $m \times m$ y $n \times n$, respectivamente, tales que*

$$UAV = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{m\times n}$$

Además, si R es la forma escalonada reducida por filas de A, entonces:

1. *U se puede calcular mediante* $\begin{bmatrix} A & I_m \end{bmatrix} \to \begin{bmatrix} R & U \end{bmatrix}$;

2. *V se puede calcular mediante* $\begin{bmatrix} R^T & I_n \end{bmatrix} \to \left[\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}_{n\times m} V^T\right]$.

Si A es de orden $m \times n$ de rango r, la matriz $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ se llama la forma normal de Smith de la matriz A. Mientras que la forma escalonada reducida por filas de A es la matriz "más sencilla" a la que se puede llevar A mediante operaciones por filas, la forma canónica de Smith es la matriz "más sencilla" a la que se puede llevar A mediante operaciones por filas y columnas. Esto se debe a que hacer operaciones por filas a R^T equivale a hacer operaciones por columnas a R y luego transponer.

Ejemplo 3.2 *Dada la matriz* $A = \begin{bmatrix} 1 & -1 & 1 & 2 \\ 2 & -2 & 1 & -1 \\ -1 & 1 & 0 & 3 \end{bmatrix}$, *encuentre las matrices invertibles U y V tales que* $UAV = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$, *donde* $r = \operatorname{rank} A$.

Sol: *La matriz U y la forma escalonada reducida por filas R de A se calculan mediante la reducción por filas* $\begin{bmatrix} A & I_3 \end{bmatrix} \to \begin{bmatrix} R & U \end{bmatrix}$:

$$\left[\begin{array}{rrrr|rrr} 1 & -1 & 1 & 2 & 1 & 0 & 0 \\ 2 & -2 & 1 & -1 & 0 & 1 & 0 \\ -1 & 1 & 0 & 3 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{rrrr|rrr} 1 & -1 & 0 & -3 & -1 & 1 & 0 \\ 0 & 0 & 1 & 5 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 1 & 1 \end{array}\right].$$

Por lo tanto

$$R = \begin{bmatrix} 1 & -1 & 0 & -3 \\ 0 & 0 & 1 & 5 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad y \quad U = \begin{bmatrix} -1 & 1 & 0 \\ 2 & -1 & 0 \\ -1 & 1 & 1 \end{bmatrix}.$$

En particular, $r = \operatorname{rank} R = 2$. *Ahora reducimos por filas* $\begin{bmatrix} R^T & I_4 \end{bmatrix} \to \begin{bmatrix} \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} & V^T \end{bmatrix}$:

$$\left[\begin{array}{rrr|rrrr} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ -1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ -3 & 5 & 0 & 0 & 0 & 0 & 1 \end{array}\right] \to \left[\begin{array}{rrr|rrrr} 1 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 & -5 & 1 \end{array}\right]$$

de donde

$$V^T = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 \\ 3 & 0 & -5 & -1 \end{bmatrix} \quad \textit{por lo tanto} \quad V = \begin{bmatrix} 1 & 0 & 1 & 3 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & -5 \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

Entonces $UAV = \begin{bmatrix} I_2 & 0 \\ 0 & 0 \end{bmatrix}$ *como se puede verificar fácilmente.*

A continuación $C_B(\boldsymbol{v})$ denota el vector de coordenadas de $\boldsymbol{v}$ con respecto a la base B.

V. Osores

Teorema 3.3 *Sea $T : V \to V$ un operador lineal donde $\dim V = n$, y sea B una base ordenada de V.*

1. $C_B(T(\boldsymbol{v})) = M_B(T)C_B(\boldsymbol{v})$ *para todo* $\boldsymbol{v}$ *en* V.
2. *Si* $S : V \to V$ *es otro operador en* V, *entonces* $M_B(ST) = M_B(S)M_B(T)$.
3. T *es un isomorfismo si y solo si* $M_B(T)$ *es invertible. En este caso* $M_D(T)$ *es invertible para toda base ordenada* D *de* V.
4. *Si* T *es un isomorfismo, entonces* $M_B\left(T^{-1}\right) = [M_B(T)]^{-1}$.
5. *Si* $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$, *entonces* $M_B(T) = \left[\begin{array}{cccc} C_B\left[T\left(\boldsymbol{b}_1\right)\right] & C_B\left[T\left(\boldsymbol{b}_2\right)\right] & \cdots & C_B\left[T\left(\boldsymbol{b}_n\right)\right]\end{array}\right]$.

Si B y D son dos bases ordenadas de V, y si tomamos $T = 1_V$ en el Teorema **??**, obtenemos

$$C_D(\boldsymbol{v}) = M_{DB}\left(1_V\right) C_B(\boldsymbol{v}) \quad \text{para todo } \boldsymbol{v} \text{ en } V.$$

Definición 3.1 **(Matriz de cambio $P_{D\leftarrow B}$ para bases B y D)** *Con esto en mente, definimos la matriz de cambio $P_{D\leftarrow B}$ por*

$$P_{D\leftarrow B} = M_{DB}\left(1_V\right) \quad \text{para cualquier base ordenada } B \text{ y } D \text{ de } V.$$

Esto prueba la ecuación (3.1.1) en el siguiente teorema:

Teorema 3.4 *Sean $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$ y D bases ordenadas de un espacio vectorial V. Entonces la matriz de cambio $P_{D\leftarrow B}$ se da en términos de sus columnas por*

$$P_{D\leftarrow B} = \left[\begin{array}{cccc} C_D\left(\boldsymbol{b}_1\right) & C_D\left(\boldsymbol{b}_2\right) & \cdots & C_D\left(\boldsymbol{b}_n\right)\end{array}\right]$$

y tiene la propiedad de que

$$C_D(\boldsymbol{v}) = P_{D\leftarrow B}C_B(\boldsymbol{v}) \text{ para todo } \boldsymbol{v} \text{ en } V \tag{3.1.1}$$

Además, si E es otra base ordenada de V, tenemos:

1. $P_{B\leftarrow B} = I_n$.
2. $P_{D\leftarrow B}$ *es invertible y* $\left(P_{D\leftarrow B}\right)^{-1} = P_{B\leftarrow D}$.
3. $P_{E\leftarrow D}P_{D\leftarrow B} = P_{E\leftarrow B}$.

Ejemplo 3.3 *En $\boldsymbol{P}_2$, encuentra $P_{D\leftarrow B}$ si $B = \left\{1, x, x^2\right\}$ y $D = \left\{1, (1-x), (1-x)^2\right\}$. Luego usa esto para expresar $p = p(x) = a + bx + cx^2$ como un polinomio en potencias de $(1-x)$.*

Sol: *Para calcular la matriz de cambio $P_{D\leftarrow B}$, expresa 1, x, x^2 en la base D:*

$$\begin{aligned} 1 &= 1 + 0(1-x) + 0(1-x)^2 \\ x &= 1 - 1(1-x) + 0(1-x)^2 \\ x^2 &= 1 - 2(1-x) + 1(1-x)^2 \end{aligned}$$

Entonces, la matriz de cambio es:

$$P_{D\leftarrow B} = \left[C_D(1), C_D(x), C_D(x^2)\right] = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix}$$

Dado que $C_B(p) = \begin{bmatrix} a \\ b \\ c \end{bmatrix}$, *tenemos que:*

$$C_D(p) = P_{D\leftarrow B} C_B(p) = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} a \\ b \\ c \end{bmatrix} = \begin{bmatrix} a+b+c \\ -b-2c \\ c \end{bmatrix}$$

Por lo tanto, $p(x) = (a+b+c) - (b+2c)(1-x) + c(1-x)^2$.

Teorema 3.5 (Teorema de similitud) *Sean B_0 y B dos bases ordenadas de un espacio vectorial finito dimensional V. Si $T : V \to V$ es cualquier operador lineal, las matrices $M_B(T)$ y $M_{B_0}(T)$ de T con respecto a estas bases son similares. Más precisamente,*

$$M_B(T) = P^{-1} M_{B_0}(T) P,$$

donde $P = P_{B_0\leftarrow B}$ es la matriz de cambio de B a B_0.

Demostración: *Sea $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$ y B_0 dos bases ordenadas de un espacio vectorial V. Un operador $T : V \to V$ tiene matrices diferentes $M_B[T]$ y $M_{B_0}[T]$ con respecto a B y B_0. Ahora podemos determinar cómo se relacionan estas matrices. El Teorema 3.4 afirma que*

$$C_{B_0}(\boldsymbol{v}) = P_{B_0\leftarrow B} C_B(\boldsymbol{v}) \text{ para todo } \boldsymbol{v} \text{ en } V$$

Por otro lado, el Teorema 3.3 da

$$C_B[T(\boldsymbol{v})] = M_B(T) C_B(\boldsymbol{v}) \text{ para todo } \boldsymbol{v} \text{ en } V$$

Combinando estos (y escribiendo $P = P_{B_0\leftarrow B}$ para conveniencia) obtenemos

$$\begin{aligned} P M_B(T) C_B(\boldsymbol{v}) &= P C_B[T(\boldsymbol{v})] \\ &= C_{B_0}[T(\boldsymbol{v})] \\ &= M_{B_0}(T) C_{B_0}(\boldsymbol{v}) \\ &= M_{B_0}(T) P C_B(\boldsymbol{v}) \end{aligned}$$

Esto se cumple para todo $\boldsymbol{v}$ *en* V. *Debido a que* $C_B(\boldsymbol{b}_j)$ *es la* j*-ésima columna de la matriz identidad, se sigue que*

$$PM_B(T) = M_{B_0}(T)P$$

Además, P *es invertible (de hecho,* $P^{-1} = P_{B \leftarrow B_0}$ *por el Teorema 3.4), por lo que esto da*

$$M_B(T) = P^{-1}M_{B_0}(T)P$$

Esto afirma que $M_{B_0}(T)$ *y* $M_B(T)$ *son matrices similares, y prueba el Teorema 3.5.*

Ejemplo 3.4 *Sea* $T : \mathbb{R}^3 \to \mathbb{R}^3$ *definida por* $T(a,b,c) = (2a-b, b+c, c-3a)$. *Si* B_0 *denota la base canónica de* $\mathbb{R}^3$ *y* $B = \{(1,1,0),(1,0,1),(0,1,0)\}$, *encuentra una matriz invertible* P *tal que* $P^{-1}M_{B_0}(T)P = M_B(T)$.

Sol: *Tenemos:*

$$M_{B_0}(T) = \begin{bmatrix} C_{B_0}(2,0,-3) & C_{B_0}(-1,1,0) & C_{B_0}(0,1,1) \end{bmatrix} = \begin{bmatrix} 2 & -1 & 0 \\ 0 & 1 & 1 \\ -3 & 0 & 1 \end{bmatrix}$$

$$M_B(T) = \begin{bmatrix} C_B(1,1,-3) & C_B(2,1,-2) & C_B(-1,1,0) \end{bmatrix} = \begin{bmatrix} 4 & 4 & -1 \\ -3 & -2 & 0 \\ -3 & -3 & 2 \end{bmatrix}$$

$$P = P_{B_0 \leftarrow B} = \begin{bmatrix} C_{B_0}(1,1,0) & C_{B_0}(1,0,1) & C_{B_0}(0,1,0) \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}$$

El lector puede verificar que $P^{-1}M_{B_0}(T)P = M_B(T)$; *equivaléntemente, que* $M_{B_0}(T)P = PM_B(T)$.

Teorema 3.6 *Sea* A *una matriz de orden* n *y sea* E *la base canónica de* $\mathbb{R}^n$.

1. *Sea* A' *similar a* A, *digamos* $A' = P^{-1}AP$, *y sea* B *la base ordenada de* $\mathbb{R}^n$ *formada por las columnas de* P *en orden. Entonces* $T_A : \mathbb{R}^n \to \mathbb{R}^n$ *es lineal y*

$$M_E(T_A) = A \quad y \quad M_B(T_A) = A'$$

2. *Si* B *es cualquier base ordenada de* $\mathbb{R}^n$, *sea* P *la matriz (invertible) cuyas columnas son los vectores en* B *en orden. Entonces*

$$M_B(T_A) = P^{-1}AP$$

Ejemplo 3.5 *Dada* $A = \begin{bmatrix} 10 & 6 \\ -18 & -11 \end{bmatrix}$, $P = \begin{bmatrix} 2 & -1 \\ -3 & 2 \end{bmatrix}$, *y* $D = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix}$, *verifica que* $P^{-1}AP = D$ *y usa este hecho para encontrar una base* B *de* $\mathbb{R}^2$ *tal que* $M_B(T_A) = D$.

Sol: *$P^{-1}AP = D$ se cumple si $AP = PD$; esta verificación queda como ejercicio para el lector. Sea B la base compuesta por las columnas de P en orden, es decir, $B = \left\{ \begin{bmatrix} 2 \\ -3 \end{bmatrix}, \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right\}$. Entonces, el Teorema 3.6 da que $M_B(T_A) = P^{-1}AP = D$. Más explícitamente,*

$$M_B(T_A) = \left[C_B \left(T_A \begin{bmatrix} 2 \\ -3 \end{bmatrix} \right) \quad C_B \left(T_A \begin{bmatrix} -1 \\ 2 \end{bmatrix} \right) \right] = \left[C_B \begin{bmatrix} 2 \\ -3 \end{bmatrix} \quad C_B \begin{bmatrix} -2 \\ 4 \end{bmatrix} \right] = \begin{bmatrix} 1 & 0 \\ 0 & -2 \end{bmatrix} = D.$$

Sea A una matriz $n \times n$. Como en el ejemplo 3.5, el Teorema 3.6 proporciona una nueva forma de encontrar una matriz invertible P tal que $P^{-1}AP$ sea diagonal. La idea es encontrar una base $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$ de $\mathbb{R}^n$ tal que $M_B(T_A) = D$ sea diagonal y tomar $P = \begin{bmatrix} \boldsymbol{b}_1 & \boldsymbol{b}_2 & \cdots & \boldsymbol{b}_n \end{bmatrix}$ como la matriz con los $\boldsymbol{b}_j$ como columnas. Entonces, por el Teorema 3.6,

$$P^{-1}AP = M_B(T_A) = D.$$

Como se mencionó anteriormente, esto convierte el problema algebraico de diagonalizar A en el problema geométrico de encontrar la base B.

El Teorema 3.6 permite deducir hechos sobre matrices a partir de las propiedades correspondientes de los operadores.

3.2 Formas canónicas de Jordan

Dada una matriz A, el efecto de una secuencia de operaciones por filas sobre A es producir UA donde U es invertible. Bajo esta operación de "equivalencia por filas" lo mejor que se puede lograr es la forma escalonada reducida por filas para A. Si también se permiten operaciones por columnas, el resultado es UAV donde tanto U como V son invertibles, y el mejor resultado bajo esta operación de "equivalencia" se llama la forma canónica de Smith de A (Teorema 3.2). Existen otros tipos de operaciones sobre una matriz y, en muchos casos, hay un resultado "canónico" que es el mejor posible.

Si A es cuadrada, la operación más importante de este tipo es posiblemente la "similitud", en la cual A se lleva a $U^{-1}AU$ donde U es invertible. En este caso decimos que las matrices A y B son similares, y escribimos $A \sim B$, cuando $B = U^{-1}AU$ para alguna matriz invertible U. Bajo similitud, las matrices canónicas, llamadas matrices canónicas de Jordan, son triangulares por bloques con bloques triangulares superiores "Jordan" en la diagonal principal. En este breve capítulo vamos a definir estos bloques de Jordan y probar que toda matriz es similar a una matriz canónica de Jordan.

Aquí está la clave del método. Sea $T : V \to V$ un operador en un espacio vectorial de dimensión n V, y supongamos que podemos encontrar una base ordenada B de V de manera que la matriz $M_B(T)$ sea lo más simple posible. Entonces, si B_0 es cualquier base ordenada de V, las matrices $M_B(T)$ y $M_{B_0}(T)$ son similares; es decir,

$$M_B(T) = P^{-1}M_{B_0}(T)P \quad \text{para alguna matriz invertible } P.$$

Además, $P = P_{B_0 \leftarrow B}$ se calcula fácilmente a partir de las bases B y D (Teorema 3.5). Esto, combinado con los subespacios invariantes y las sumas directas, nos permite calcular la forma canónica de

Jordan de cualquier matriz cuadrada A. A lo largo del camino derivamos una construcción explícita de una matriz invertible P tal que $P^{-1}AP$ es triangular por bloques.

Esta técnica es importante de muchas maneras. Por ejemplo, si queremos diagonalizar una matriz A de orden n, sea $T_A : \mathbb{R}^n \to \mathbb{R}^n$ el operador dado por $T_A(\boldsymbol{x}) = A\boldsymbol{x}$ para todo $\boldsymbol{x}$ en $\mathbb{R}^n$, y busquemos una base B de $\mathbb{R}^n$ tal que $M_B(T_A)$ sea diagonal. Si $B_0 = E$ es la base estándar de $\mathbb{R}^n$, entonces $M_E(T_A) = A$, por lo que

$$P^{-1}AP = P^{-1}M_E(T_A)\,P = M_B(T_A)\,,$$

y hemos diagonalizado A. Así, el problema "algebraico" de encontrar una matriz invertible P tal que $P^{-1}AP$ sea diagonal se convierte en el problema "geométrico" de encontrar una base B tal que $M_B(T_A)$ sea diagonal. Este cambio de perspectiva es una de las técnicas más importantes en álgebra lineal.

Teorema 3.7 *Sea $T : V \to V$ un operador lineal donde V tiene dimensión n. Supongamos que $V = U_1 \oplus U_2$ donde tanto U_1 como U_2 son T-invariantes. Si $B_1 = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k\}$ y $B_2 = \{\boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_n\}$ son bases de U_1 y U_2 respectivamente, entonces*

$$B = \{\boldsymbol{b}_1, \ldots, \boldsymbol{b}_k, \boldsymbol{b}_{k+1}, \ldots, \boldsymbol{b}_n\}$$

es una base de V, y $M_B(T)$ tiene la forma diagonal por bloques

$$M_B(T) = \begin{bmatrix} M_{B_1}(T) & 0 \\ 0 & M_{B_2}(T) \end{bmatrix}$$

donde $M_{B_1}(T)$ y $M_{B_2}(T)$ son las matrices de las restricciones de T a U_1 y a U_2 respectivamente.

3.2.1 Forma triangular por bloques

Hemos mostrado (Teorema 2.22) que cualquier matriz A de orden n con todos los valores propios reales es ortogonalmente similar a una matriz triangular superior U. El siguiente teorema muestra que U puede elegirse de una manera especial.

Teorema 3.8 (Teorema de triangulación por bloques) *Sea A una matriz $n \times n$ con todos sus valores propios reales y sea*

$$f_A(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los valores propios distintos de A. Entonces existe una matriz invertible P tal que

$$P^{-1}AP = \begin{bmatrix} U_1 & 0 & 0 & \cdots & 0 \\ 0 & U_2 & 0 & \cdots & 0 \\ 0 & 0 & U_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & U_k \end{bmatrix}$$

donde, para cada i, U_i es una matriz triangular superior $m_i \times m_i$ con todos los elementos en la diagonal principal iguales a λ_i.

Por ahora, nos centramos en un método para encontrar la matriz P. El concepto clave es el siguiente.

Definición 3.2 **(Espacios propios generalizados)** *Si A es como en el Teorema 3.8, el espacio propio generalizado $G_{\lambda_i}(A)$ se define por*

$$G_{\lambda_i}(A) = \operatorname{null}\left[(\lambda_i I - A)^{m_i}\right]$$

donde m_i es la multiplicidad de λ_i.

Observe que el espacio propio $E_{\lambda_i}(A) = \operatorname{null}(\lambda_i I - A)$ es un subespacio de $G_{\lambda_i}(A)$. Necesitamos tres resultados técnicos.

Lema 3.1 *Usando la notación del Teorema 3.8, tenemos* $\dim\left[G_{\lambda_i}(A)\right] = m_i$.

Demostración: *Escriba $A_i = (\lambda_i I - A)^{m_i}$ para conveniencia y sea P como en el Teorema 3.8. Los espacios $G_{\lambda_i}(A) = \operatorname{null}(A_i)$ y $\operatorname{null}\left(P^{-1}A_iP\right)$ son isomorfos vía $\boldsymbol{x} \leftrightarrow P^{-1}\boldsymbol{x}$, por lo que mostramos que $\dim\left[\operatorname{null}\left(P^{-1}A_iP\right)\right] = m_i$. Ahora $P^{-1}A_iP = \left(\lambda_i I - P^{-1}AP\right)^{m_i}$. Si usamos la forma de bloques en el Teorema 3.8, esto se convierte en*

$$\begin{aligned} P^{-1}A_iP &= \begin{bmatrix} \lambda_i I - U_1 & 0 & \cdots & 0 \\ 0 & \lambda_i I - U_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \lambda_i I - U_k \end{bmatrix}^{m_i} \\ &= \begin{bmatrix} (\lambda_i I - U_1)^{m_i} & 0 & \cdots & 0 \\ 0 & (\lambda_i I - U_2)^{m_i} & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & (\lambda_i I - U_k)^{m_i} \end{bmatrix} \end{aligned}$$

La matriz $(\lambda_i I - U_j)^{m_i}$ es invertible si $j \neq i$ y cero si $j = i$ (porque entonces U_i es una matriz triangular superior $m_i \times m_i$ con cada entrada en la diagonal principal igual a λ_i). Se sigue que $m_i = \dim\left[\operatorname{null}\left(P^{-1}A_iP\right)\right]$, como se requiere.

Lema 3.2 *Si P es como en el Teorema 3.8, denote las columnas de P de la siguiente manera:*

$$\boldsymbol{p}_{11}, \boldsymbol{p}_{12}, \ldots, \boldsymbol{p}_{1m_1}; \quad \boldsymbol{p}_{21}, \boldsymbol{p}_{22}, \ldots, \boldsymbol{p}_{2m_2}; \quad \cdots \quad ; \quad \boldsymbol{p}_{k1}, \boldsymbol{p}_{k2}, \ldots, \boldsymbol{p}_{km_k}$$

Entonces $\{\boldsymbol{p}_{i1}, \boldsymbol{p}_{i2}, \ldots, \boldsymbol{p}_{im_i}\}$ es una base de $G_{\lambda_i}(A)$.

Demostración: *Basta por el Lema 3.1 mostrar que cada $\boldsymbol{p}_{ij}$ está en $G_{\lambda_i}(A)$. Escriba la matriz en el Teorema 3.8 como $P^{-1}AP = \operatorname{diag}(U_1, U_2, \ldots, U_k)$. Entonces*

$$AP = P\operatorname{diag}(U_1, U_2, \ldots, U_k)$$

Comparando columnas da, sucesivamente:

$$\begin{array}{ll} A\boldsymbol{p}_{11} = \lambda_1 \boldsymbol{p}_{11}, & \textit{así que } (\lambda_1 I - A)\,\boldsymbol{p}_{11} = \mathbf{0} \\ A\boldsymbol{p}_{12} = u\boldsymbol{p}_{11} + \lambda_1 \boldsymbol{p}_{12}, & \textit{así que } (\lambda_1 I - A)^2\,\boldsymbol{p}_{12} = \mathbf{0} \\ A\boldsymbol{p}_{13} = w\boldsymbol{p}_{11} + v\boldsymbol{p}_{12} + \lambda_1 \boldsymbol{p}_{13} & \textit{así que } (\lambda_1 I - A)^3\,\boldsymbol{p}_{13} = \mathbf{0} \\ \vdots & \vdots \end{array}$$

donde u, v, w están en $\mathbb{R}$. En general, $(\lambda_1 I - A)^j\,\boldsymbol{p}_{1j} = \mathbf{0}$ para $j = 1, 2, \ldots, m_1$, por lo que $\boldsymbol{p}_{1j}$ está en $G_{\lambda_1}(A)$. Del mismo modo, $\boldsymbol{p}_{ij}$ está en $G_{\lambda_i}(A)$ para cada i y j.

Lema 3.3 *Si B_i es cualquier base de $G_{\lambda_i}(A)$, entonces $B = B_1 \cup B_2 \cup \cdots \cup B_k$ es una base de $\mathbb{R}^n$.*

Demostración: *Basta por el Lema 3.1 mostrar que B es independiente (l.i). Si una combinación lineal de B se anula, sea $\boldsymbol{x}_i$ la suma de los términos de B_i. Entonces $\boldsymbol{x}_1 + \cdots + \boldsymbol{x}_k = \mathbf{0}$. Pero $\boldsymbol{x}_i = \sum_j r_{ij}\boldsymbol{p}_{ij}$ por el Lema 3.2, así que $\sum_{i,j} r_{ij}\boldsymbol{p}_{ij} = \mathbf{0}$. Por lo tanto, cada $\boldsymbol{x}_i = \mathbf{0}$, así que cada coeficiente en $\boldsymbol{x}_i$ es cero.*

El Lema 3.2 sugiere un algoritmo para encontrar la matriz P en el Teorema 3.8. Observe que hay una cadena ascendente de subespacios que va de $E_{\lambda_i}(A)$ a $G_{\lambda_i}(A)$:

$$E_{\lambda_i}(A) = \operatorname{null}\left[(\lambda_i I - A)\right] \subseteq \operatorname{null}\left[(\lambda_i I - A)^2\right] \subseteq \cdots \subseteq \operatorname{null}\left[(\lambda_i I - A)^{m_i}\right] = G_{\lambda_i}(A).$$

Construimos una base para $G_{\lambda_i}(A)$ ascendiendo por esta cadena.

Definición 3.3 (Algoritmo de Triangulación) *Supongamos que A tiene polinomio característico*

$$f_A(x) = (x - \lambda_1)^{m_1}(x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$$

1. *Elige una base de $\operatorname{null}[(\lambda_1 I - A)]$; amplíala añadiendo vectores (posiblemente ninguno) para formar una base de $\operatorname{null}[(\lambda_1 I - A)^2]$; amplíala nuevamente para formar una base de $\operatorname{null}[(\lambda_1 I - A)^3]$, y así sucesivamente. Continúa hasta obtener una base ordenada $\{\boldsymbol{p}_{11}, \boldsymbol{p}_{12}, \ldots, \boldsymbol{p}_{1m_1}\}$ de $G_{\lambda_1}(A)$.*

2. *De manera similar a (1), elige una base $\{\boldsymbol{p}_{i1}, \boldsymbol{p}_{i2}, \ldots, \boldsymbol{p}_{im_i}\}$ de $G_{\lambda_i}(A)$ para cada i.*

3. *Sea $P = [\boldsymbol{p}_{11}\boldsymbol{p}_{12}\cdots\boldsymbol{p}_{1m_1};\ \ \boldsymbol{p}_{21}\boldsymbol{p}_{22}\cdots\boldsymbol{p}_{2m_2}; \cdots;\ \ \boldsymbol{p}_{k1}\boldsymbol{p}_{k2}\cdots\boldsymbol{p}_{km_k}]$ la matriz con estos vectores base (en orden) como columnas.*

Entonces $P^{-1}AP = \operatorname{diag}(U_1, U_2, \ldots, U_k)$ como en el Teorema 3.8.

Demostración: *El Lema 3.3 garantiza que $B = \{\boldsymbol{p}_{11}, \ldots, \boldsymbol{p}_{km_k}\}$ es una base de $\mathbb{R}^n$, y el Teorema 3.6 muestra que $P^{-1}AP = M_B(T_A)$. Ahora, $G_{\lambda_i}(A)$ es T_A-invariante para cada i porque*

$$(\lambda_i I - A)^{m_i}\,\boldsymbol{x} = \mathbf{0} \quad \textit{implica} \quad (\lambda_i I - A)^{m_i}(A\boldsymbol{x}) = A(\lambda_i I - A)^{m_i}\,\boldsymbol{x} = \mathbf{0}.$$

Por el Teorema 3.7 (y por inducción), tenemos

$$P^{-1}AP = M_B(T_A) = \operatorname{diag}(U_1, U_2, \ldots, U_k)$$

donde U_i es la matriz de la restricción de T_A a $G_{\lambda_i}(A)$, y queda por mostrar que U_i tiene la forma triangular superior deseada. Dado s, sea $\boldsymbol{p}_{ij}$ un vector base en $\operatorname{null}\left[(\lambda_i I - A)^{s+1}\right]$. *Entonces, $(\lambda_i I - A)\,\boldsymbol{p}_{ij}$ está en* $\operatorname{null}\left[(\lambda_i I - A)^{s}\right]$, *y por lo tanto es una combinación lineal de los vectores base $\boldsymbol{p}_{it}$ que vienen antes de $\boldsymbol{p}_{ij}$. Por lo tanto,*

$$T_A\left(\boldsymbol{p}_{ij}\right) = A\boldsymbol{p}_{ij} = \lambda_i \boldsymbol{p}_{ij} - (\lambda_i I - A)\,\boldsymbol{p}_{ij}$$

muestra que la columna de U_i correspondiente a $\boldsymbol{p}_{ij}$ tiene λ_i en la diagonal principal y ceros debajo de la diagonal principal. Esto es lo que queríamos.

Ejemplo 3.6 *Si* $A = \begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 2 & 0 & -1 \\ -1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 2 \end{bmatrix}$, *encuentra P tal que $P^{-1}AP$ sea triangular por bloques.*

Sol: $f_A(x) = \det[xI - A] = (x-2)^4$, *por lo que $\lambda_1 = 2$ es el único valor propio y estamos en el caso $k = 1$ del Teorema 3.8. Calculemos:*

$$(2I - A) = \begin{bmatrix} 0 & 0 & 0 & -1 \\ 0 & 0 & 0 & 1 \\ 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (2I - A)^2 = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & -2 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad (2I - A)^3 = 0_4$$

Mediante eliminación gaussiana encuentra una base $\{\boldsymbol{p}_{11}, \boldsymbol{p}_{12}\}$ de $\operatorname{null}(2I - A)$; *luego extiende de cualquier manera a una base $\{\boldsymbol{p}_{11}, \boldsymbol{p}_{12}, \boldsymbol{p}_{13}\}$ de* $\operatorname{null}[(2I - A)^2]$; *y finalmente consigue una base $\{\boldsymbol{p}_{11}, \boldsymbol{p}_{12}, \boldsymbol{p}_{13}, \boldsymbol{p}_{14}\}$ de* $\operatorname{null}\left[(2I - A)^3\right] = \mathbb{R}^4$. *Una elección es*

$$\boldsymbol{p}_{11} = \begin{bmatrix} 1 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{p}_{12} = \begin{bmatrix} 0 \\ 0 \\ 1 \\ 0 \end{bmatrix} \quad \boldsymbol{p}_{13} = \begin{bmatrix} 0 \\ 1 \\ 0 \\ 0 \end{bmatrix} \quad \boldsymbol{p}_{14} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix}.$$

Por lo tanto, $P = \begin{bmatrix} \boldsymbol{p}_{11} & \boldsymbol{p}_{12} & \boldsymbol{p}_{13} & \boldsymbol{p}_{14} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ *da* $P^{-1}AP = \begin{bmatrix} 2 & 0 & 0 & 1 \\ 0 & 2 & 1 & 0 \\ 0 & 0 & 2 & -2 \\ 0 & 0 & 0 & 2 \end{bmatrix}$

Ejemplo 3.7 *Si* $A = \begin{bmatrix} 2 & 0 & 1 & 1 \\ 3 & 5 & 4 & 1 \\ -4 & -3 & -3 & -1 \\ 1 & 0 & 1 & 2 \end{bmatrix}$, *encuentra P tal que $P^{-1}AP$ sea triangular por bloques.*

Sol: *Los valores propios son $\lambda_1 = 1$ y $\lambda_2 = 2$. En efecto,*

$$f_A(x) = \begin{vmatrix} x-2 & 0 & -1 & -1 \\ -3 & x-5 & -4 & -1 \\ 4 & 3 & x+3 & 1 \\ -1 & 0 & -1 & x-2 \end{vmatrix} = \begin{vmatrix} x-1 & 0 & 0 & -x+1 \\ -3 & x-5 & -4 & -1 \\ 4 & 3 & x+3 & 1 \\ -1 & 0 & -1 & x-2 \end{vmatrix}$$

$$= \begin{vmatrix} x-1 & 0 & 0 & 0 \\ -3 & x-5 & -4 & -4 \\ 4 & 3 & x+3 & 5 \\ -1 & 0 & -1 & x-3 \end{vmatrix} = (x-1)\begin{vmatrix} x-5 & -4 & -4 \\ 3 & x+3 & 5 \\ 0 & -1 & x-3 \end{vmatrix}$$

$$= (x-1)\begin{vmatrix} x-5 & -4 & 0 \\ 3 & x+3 & -x+2 \\ 0 & -1 & x-2 \end{vmatrix} = (x-1)\begin{vmatrix} x-5 & -4 & 0 \\ 3 & x+2 & 0 \\ 0 & -1 & x-2 \end{vmatrix}$$

$$= (x-1)(x-2)\begin{vmatrix} x-5 & -4 \\ 3 & x+2 \end{vmatrix} = (x-1)^2(x-2)^2$$

Resolviendo ecuaciones, encontramos que $\operatorname{null}(I-A) = \operatorname{span}\{\boldsymbol{p}_{11}\}$ *y* $\operatorname{null}(I-A)^2 = \operatorname{span}\{\boldsymbol{p}_{11}, \boldsymbol{p}_{12}\}$ *donde*

$$\boldsymbol{p}_{11} = \begin{bmatrix} 1 \\ 1 \\ -2 \\ 1 \end{bmatrix} \quad \boldsymbol{p}_{12} = \begin{bmatrix} 0 \\ 3 \\ -4 \\ 1 \end{bmatrix}$$

Dado que $\lambda_1 = 1$ tiene multiplicidad 2 como raíz de $f_A(x)$, $\dim G_{\lambda_1}(A) = 2$ *por el Lema 3.1. Dado que $\boldsymbol{p}_{11}$ y $\boldsymbol{p}_{12}$ se encuentran en $G_{\lambda_1}(A)$, tenemos* $G_{\lambda_1}(A) = \operatorname{span}\{\boldsymbol{p}_{11}, \boldsymbol{p}_{12}\}$. *En cuanto a $\lambda_2 = 2$, encontramos que* $\operatorname{null}(2I-A) = \operatorname{span}\{\boldsymbol{p}_{21}\}$ *y* $\operatorname{null}\left[(2I-A)^2\right] = \operatorname{span}\{\boldsymbol{p}_{21}, \boldsymbol{p}_{22}\}$ *donde*

$$\boldsymbol{p}_{21} = \begin{bmatrix} 1 \\ 0 \\ -1 \\ 1 \end{bmatrix} \quad y \quad \boldsymbol{p}_{22} = \begin{bmatrix} 0 \\ -4 \\ 3 \\ 0 \end{bmatrix}$$

De nuevo, $\dim G_{\lambda_2}(A) = 2$ *ya que λ_2 tiene multiplicidad 2, por lo que* $G_{\lambda_2}(A) = \operatorname{span}\{\boldsymbol{p}_{21}, \boldsymbol{p}_{22}\}$. *Por lo tanto,*

$$P = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 1 & 3 & 0 & -4 \\ -2 & -4 & -1 & 3 \\ 1 & 1 & 1 & 0 \end{bmatrix} \text{ da } P^{-1}AP = \begin{bmatrix} 1 & -3 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 3 \\ 0 & 0 & 0 & 2 \end{bmatrix}.$$

Si $p(x)$ es un polinomio y A es una matriz $n \times n$, entonces $p(A)$ es también una matriz $n \times n$ si interpretamos $A^0 = I_n$. Por ejemplo, si $p(x) = x^2 - 2x + 3$, entonces $p(A) = A^2 - 2A + 3I$. El Teorema 3.8 proporciona otra demostración del teorema de Cayley-Hamilton (ver también el Teorema 2.29). Como antes, sea $f_A(x)$ el polinomio característico de A.

Teorema 3.9 (Teorema de Cayley-Hamilton) *Si A es una matriz cuadrada con todos sus valores propios reales, entonces $f_A(A) = 0$.*

Demostración: *Como en el Teorema 3.8, escribimos $f_A(x) = (x - \lambda_1)^{m_1} \cdots (x - \lambda_k)^{m_k} = \Pi_{i=1}^{k}(x - \lambda_i)^{m_i}$, y escribimos*

$$P^{-1}AP = D = \operatorname{diag}(U_1, \ldots, U_k).$$

Por lo tanto,

$$f_A(U_i) = \Pi_{i=1}^{k}(U_i - \lambda_i I_{m_i})^{m_i} = 0 \text{ para cada } i$$

porque el factor $(U_i - \lambda_i I_{m_i})^{m_i} = 0$. De hecho, $U_i - \lambda_i I_{m_i}$ es de tamaño $m_i \times m_i$ y tiene ceros en la diagonal principal. Pero entonces

$$\begin{aligned} P^{-1}f_A(A)P = f_A(D) &= f_A[\operatorname{diag}(U_1, \ldots, U_k)] \\ &= \operatorname{diag}[f_A(U_1), \ldots, f_A(U_k)] \\ &= 0 \end{aligned}$$

Se deduce que $f_A(A) = 0$.

Ejemplo 3.8 *Si*

$$A = \begin{bmatrix} 1 & 3 \\ -1 & 2 \end{bmatrix}, \text{ entonces } f_A(x) = \det\begin{bmatrix} x-1 & -3 \\ 1 & x-2 \end{bmatrix} = x^2 - 3x + 5. \text{ Entonces}$$

$$f_A(A) = A^2 - 3A + 5I_2 = \begin{bmatrix} -2 & 9 \\ -3 & 1 \end{bmatrix} - \begin{bmatrix} 3 & 9 \\ -3 & 6 \end{bmatrix} + \begin{bmatrix} 5 & 0 \\ 0 & 5 \end{bmatrix} = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}$$

El Teorema 3.8 se refinará aún más en la próxima sección.

Ahora estamos en condiciones de demostrar el Teorema 3.8 (ver Apéndice A.1)

3.2.2 Forma canónica de Jordan

Dos matrices $m \times n$ A y B se llaman equivalentes por filas si A puede llevarse a B usando operaciones de fila y, equivalentemente, si $B = UA$ para alguna matriz invertible U. Sabemos que cada matriz $m \times n$ es equivalente por filas a una matriz única en forma escalonada reducida, y decimos que estas matrices escalonadas reducidas son formas canónicas para matrices $m \times n$ usando operaciones de fila. Si también permitimos operaciones de columna, entonces $A \to UAV = \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ para U y V invertibles, y las formas canónicas son las matrices $\begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix}$ donde r es el rango (esto es la forma normal de Smith y se discute en el Teorema 3.2). En esta sección, discutimos las formas canónicas para matrices cuadradas bajo semejanza: $A \to P^{-1}AP$.

Si A es una matriz $n \times n$ con eigenvalores reales distintos $\lambda_1, \lambda_2, \ldots, \lambda_k$, vimos en el Teorema 3.8 que A es semejante a una matriz triangular por bloques; más precisamente, existe una matriz invertible P

tal que

$$P^{-1}AP = \begin{bmatrix} U_1 & 0 & \cdots & 0 \\ 0 & U_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & U_k \end{bmatrix} = \operatorname{diag}(U_1, U_2, \ldots, U_k) \tag{11.2}$$

donde, para cada i, U_i es triangular superior con λ_i repetido en la diagonal principal. La forma canónica de Jordan es un refinamiento de este teorema. La prueba que dimos de (11.2) es matricial porque queríamos dar un algoritmo para encontrar la matriz P. Sin embargo, aquí vamos a emplear métodos abstractos. En consecuencia, reformulamos el Teorema 3.8 de la siguiente manera:

Teorema 3.10 *Sea $T : V \to V$ un operador lineal donde $\dim V = n$. Supongamos que $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los eigenvalores distintos de T, y que los λ_i son todos reales. Entonces existe una base F de V tal que $M_F(T) = \operatorname{diag}(U_1, U_2, \ldots, U_k)$ donde, para cada i, U_i es cuadrada, triangular superior, con λ_i repetido en la diagonal principal.*

Demostración: *Elija cualquier base $B = \{\boldsymbol{b}_1, \boldsymbol{b}_2, \ldots, \boldsymbol{b}_n\}$ de V y escriba $A = M_B(T)$. Dado que A tiene los mismos eigenvalores que T, el Teorema 3.8 muestra que existe una matriz invertible P tal que*

$$P^{-1}AP = \operatorname{diag}(U_1, U_2, \ldots, U_k)$$

donde los U_i son como en la declaración del teorema. Si $\boldsymbol{p}_j$ denota la columna j de P y $C_B : V \to \mathbb{R}^n$ es el isomorfismo de coordenadas, sea $\boldsymbol{f}_j = C_B^{-1}(\boldsymbol{p}_j)$ para cada j. Entonces $F = \{\boldsymbol{f}_1, \boldsymbol{f}_2, \ldots, \boldsymbol{f}_n\}$ es una base de V y $C_B(\boldsymbol{f}_j) = \boldsymbol{p}_j$ para cada j. Esto significa que $P_{B \leftarrow F} = [C_B(\boldsymbol{f}_j)] = [\boldsymbol{p}_j] = P$, y por lo tanto (por el Teorema 3.4) que $P_{F \leftarrow B} = P^{-1}$. Con esto, la columna j de $M_F(T)$ es

$$C_F(T(\boldsymbol{f}_j)) = P_{F \leftarrow B} C_B(T(\boldsymbol{f}_j)) = P^{-1} M_B(T) C_B(\boldsymbol{f}_j) = P^{-1} A \boldsymbol{p}_j$$

para todo j. Por lo tanto

$$M_F(T) = [C_F(T(\boldsymbol{f}_j))] = \left[P^{-1} A \boldsymbol{p}_j\right] = P^{-1} A [\boldsymbol{p}_j] = P^{-1}AP = \operatorname{diag}(U_1, U_2, \ldots, U_k)$$

como se requería.

Definición 3.4 **(Bloques de Jordan)** *Si $n \geq 1$, definimos el bloque de Jordan $J_n(\lambda)$ como la matriz $n \times n$ con λ en la diagonal principal, unos en la diagonal superior, y 0s en otros lugares. Tomamos $J_1(\lambda) = [\lambda]$.*

Por lo tanto

$$J_1(\lambda) = [\lambda], \quad J_2(\lambda) = \begin{bmatrix} \lambda & 1 \\ 0 & \lambda \end{bmatrix}, \quad J_3(\lambda) = \begin{bmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{bmatrix}, \quad J_4(\lambda) = \begin{bmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & \lambda \end{bmatrix}, \quad \ldots$$

Vamos a mostrar que el Teorema 3.10 se cumple con cada bloque U_i reemplazado por bloques de Jordan correspondientes a los eigenvalores. Resulta que todo depende del caso $\lambda = 0$. Un operador T se llama nilpotente si $T^m = 0$ para algún $m \geq 1$, y en este caso $\lambda = 0$ para cada eigenvalor λ de T. Además, la recíproca es cierta por el Teorema 3.8. Por lo tanto, el siguiente lema es crucial.

Lema 3.4 *Sea $T : V \to V$ un operador lineal donde $\dim V = n$, y asumamos que T es nilpotente; es decir, $T^m = 0$ para algún $m \geq 1$. Entonces V tiene una base B tal que*

$$M_B(T) = \operatorname{diag}(J_1, J_2, \ldots, J_k)$$

donde cada J_i es un bloque de Jordan correspondiente a $\lambda = 0$.[a]

[a] La recíproca también es cierta: Si $M_B(T)$ tiene esta forma para alguna base B de V, entonces T es nilpotente.

Demostración: *La demostración se presenta en el Apéndice A.2.*

Teorema 3.11 (Forma Canónica de Jordan real) *Sea $T : V \to V$ un operador lineal donde $\dim V = n$, y asumamos que $\lambda_1, \lambda_2, \ldots, \lambda_m$ son los eigenvalores distintos de T y que los λ_i son todos reales. Entonces existe una base E de V tal que*

$$M_E(T) = \operatorname{diag}(U_1, U_2, \ldots, U_k)$$

en forma de bloque. Además, cada U_j es a su vez diagonal por bloques:

$$U_j = \operatorname{diag}(J_1, J_2, \ldots, J_k)$$

donde cada J_i es un bloque de Jordan correspondiente a algún λ_i.

Demostración: *Sea $E = \{\boldsymbol{e}_1, \boldsymbol{e}_2, \ldots, \boldsymbol{e}_n\}$ una base de V como en el Teorema 3.10, y supongamos que U_i es una matriz $n_i \times n_i$ para cada i. Definimos*

$$E_1 = \{\boldsymbol{e}_1, \ldots, \boldsymbol{e}_{n_1}\}, \quad E_2 = \{\boldsymbol{e}_{n_1+1}, \ldots, \boldsymbol{e}_{n_2}\}, \quad \ldots, \quad E_k = \left\{\boldsymbol{e}_{n_{k-1}+1}, \ldots, \boldsymbol{e}_{n_k}\right\}$$

donde $n_k = n$, y definimos $V_i = \operatorname{span}\{E_i\}$ para cada i. Como la matriz $M_E(T) = \operatorname{diag}(U_1, U_2, \ldots, U_m)$ es diagonal por bloques, cada V_i es invariante bajo T y $M_{E_i}(T) = U_i$ para cada i. Supongamos que U_i tiene λ_i repetido en la diagonal principal, y consideramos la restricción $T : V_i \to V_i$. Entonces $M_{E_i}(T - \lambda_i I_{n_i})$ es una matriz nilpotente, por lo que $(T - \lambda_i I_{n_i})$ es un operador nilpotente en V_i. Entonces, el Lema 3.4 muestra que V_i tiene una base B_i tal que $M_{B_i}(T - \lambda_i I_{n_i}) = \operatorname{diag}(K_1, K_2, \ldots, K_{t_i})$ donde cada K_i es un bloque de Jordan correspondiente a $\lambda = 0$. Por lo tanto,

$$\begin{aligned} M_{B_i}(T) &= M_{B_i}(\lambda_i I_{n_i}) + M_{B_i}(T - \lambda_i I_{n_i}) \\ &= \lambda_i I_{n_i} + \operatorname{diag}(K_1, K_2, \ldots, K_{t_i}) = \operatorname{diag}(J_1, J_2, \ldots, J_k) \end{aligned}$$

donde $J_i = \lambda_i I_{f_i} + K_i$ es un bloque de Jordan correspondiente a λ_i (donde K_i es $f_i \times f_i$). Finalmente,

$$B = B_1 \cup B_2 \cup \cdots \cup B_k$$

es una base de V respecto a la cual T tiene la matriz deseada.

Corolario 3.1 *Si A es una matriz $n \times n$ con eigenvalores reales, existe una matriz invertible P tal que $P^{-1}AP = \operatorname{diag}(J_1, J_2, \ldots, J_k)$ donde cada J_i es un bloque de Jordan correspondiente a un eigenvalor λ_i.*

Demostración: *Aplicamos el Teorema 3.11 a la transformación* $T_A : \mathbb{R}^n \to \mathbb{R}^n$ *para encontrar una base* B *de* $\mathbb{R}^n$ *tal que* $M_B(T_A)$ *tenga la forma deseada. Si* P *es la matriz (invertible)* $n \times n$ *con los vectores de* B *como sus columnas, entonces* $P^{-1}AP = M_B(T_A)$ *por el Teorema 3.6.*

Por supuesto, si trabajamos sobre el campo $\mathbb{C}$ de números complejos en lugar de $\mathbb{R}$, el polinomio característico de una matriz (compleja) A se descompone completamente en un producto de factores lineales.

Teorema 3.12 (Forma Canónica de Jordan[2]) *Sea* $T : V \to V$ *un operador lineal donde* $\dim V = n$, *y supongamos que* $\lambda_1, \lambda_2, \ldots, \lambda_m$ *son los valores propios distintos de* T. *Entonces existe una base* F *de* V *tal que*

$$M_F(T) = \operatorname{diag}(U_1, U_2, \ldots, U_k)$$

en forma de bloques. Además, cada U_j *es diagonal por bloques:*

$$U_j = \operatorname{diag}\left(J_1, J_2, \ldots, J_{t_j}\right)$$

donde cada J_i *es un bloque de Jordan correspondiente a algún* λ_i.

Excepto por el orden de los bloques de Jordan J_i, la forma canónica de Jordan está determinada de manera única por el operador T. Es decir, para cada valor propio λ el número y tamaño de los bloques de Jordan correspondientes a λ están determinados de manera única. Así, por ejemplo, dos matrices (o dos operadores) son similares si y solo si tienen la misma forma canónica de Jordan. Omitimos la prueba de la unicidad; es mejor presentarla utilizando módulos en un curso de álgebra abstracta.

3.2.3 Motivación: ¿Cómo encontrar la factorización enunciada en el Corolario 3.1

Consideremos la matriz

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

la cuál no es diagonalizable. El polinomio característico viene dado por

$$p(\lambda) = (-\lambda - 2)\left(\lambda^2 - 2\lambda + 1\right) = (-\lambda - 2)(\lambda - 1)^2$$

de donde los autovalores son -2 y 1. Los subespacios propios de cada autovalor son

$$\begin{cases} E_{-2}(C) = \langle\{(1, -1, 0)\}\rangle \\ E_1(C) = \langle\{(0, -1, 1)\}\rangle \end{cases}$$

y por tanto la matriz no es diagonalizable porque las multiplicidades algebraica y geométrica no coinciden.

En esta sección vamos a ver que C es al menos semejante a una matriz que, aunque no es diagonal, es lo suficientemente sencilla como para poder usarla cómodamente. En este ejemplo concreto, veremos que la matriz C es semejante (es decir, que existe una matriz regular P tal que $C = PJP^{-1}$) a la matriz

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

La matriz J no es diagonal, pero solo tiene elementos distintos de cero en la diagonal principal, y 1s en la diagonal justo por encima de la diagonal principal. Este tipo de matrices son las que van a jugar el papel que jugaban las matrices diagonales en el tema anterior. En este ejemplo concreto podemos decir que al menos la matriz J, a pesar de no ser diagonal, no es tan mala. No es difícil ver (por inducción) que

$$J^k = \begin{bmatrix} 1^k & k1^{k-1} & 0 \\ 0 & 1^k & 0 \\ 0 & 0 & (-2)^k \end{bmatrix}$$

y que por tanto podemos calcular las potencias de C de una forma sencilla, ya que al igual que en secciones anteriores, se tiene

$$C^k = \left(PJP^{-1}\right)^k = PJ^kP^{-1}.$$

Ya veremos más adelante cómo encontrar esa matriz J, por ahora vamos a suponer que ya sabemos que la matriz J anterior es semejante a C. Nos vamos a concentrar en las siguientes líneas en encontrar la matriz P.

Si alguien nos asegura que la matriz C es semejante a J, entonces sabemos que lo que está ocurriendo es lo siguiente. Si denotamos por

$$f : \mathbb{R}^3 \to \mathbb{R}^3$$

al endomorfismo cuya matriz es C respecto de la base canónica, es decir, que f es la transformación lineal dada por

$$f(x, y, z) = (2y + 2z, 2x - z, -x - y),$$

entonces decir que C es semejante a J es lo mismo que decir que somos capaces de encontrar una base $B = \{v_1, v_2, v_3\}$ respecto de la cual la matriz de f es precisamente J.

Mirando la matriz J, ¿qué propiedades tienen los vectores v_1, v_2, v_3? Si la matriz de f respecto de la base B es J entonces sabemos que por definición las columnas de J son las coordenadas de $f(v_1), f(v_2), f(v_3)$ respecto de la base B respectivamente. Es decir,

$$\begin{cases} f(v_1) & = v_1 \\ f(v_2) & = v_1 + v_2 \\ f(v_3) & = -2v_3 \end{cases}.$$

Es evidente que v_1 y v_3 deben ser autovectores de f asociados a los autovalores 1 y -2 respectivamente. ¿Pero qué propiedad especial tiene v_2? Si manipulamos un poco la expresión anterior llegamos a que

$$f(v_2) = v_1 + v_2 \Leftrightarrow f(v_2) - v_2 = v_1 \Leftrightarrow (f - I)(v_2) = v_1.$$

Es decir, que cuando calculo la imagen de v_2 por medio de $f - I$ me da precisamente v_1. Si seguimos manipulando un poco, como sabemos que $(f - I)(v_2) = v_1$, al aplicar $f - I$ a los dos lados obtengo

$$(f - I)(v_2) = v_1 \Longrightarrow (f - I)[(f - I)(v_2)] = (f - I)(v_1) = 0$$

donde $(f-I)(v_1)=0$ porque ya sabíamos que v_1 era un autovector asociado al autovalor 1. Así que

$$(f-I)^2(v_2)=0 \Longrightarrow v_2 \in \operatorname{Ker}\left[(f-I)^2\right].$$

Así que ya sabemos dónde buscar v_2, en el núcleo de $(f-I)^2$ (espacio propio generalizado de $\lambda=1$). La matriz de $(f-I)$ respecto de la base canónica es evidentemente $C-I$, así que la matriz de $(f-I)^2$ debe ser $(C-I)^2$ (porque la matriz de la composición de dos aplicaciones es el producto de las matrices). Por tanto, ya podemos calcular todos los vectores $(x,y,z)\in\mathbb{R}^3$ que pertenecen al núcleo de $(f-I)^2$. Se tiene que satisfacer

$$\underbrace{\begin{bmatrix} 3 & -6 & -6 \\ -3 & 6 & 6 \\ 0 & 0 & 0 \end{bmatrix}}_{(C-I)^2}\begin{bmatrix} x \\ y \\ z \end{bmatrix}=\begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

es decir,

$$\left\{\begin{aligned} 3x-6y-6z&=0 \\ -3x+6y+6z&=0 \\ 0&=0 \end{aligned}\right. \rightarrow \left\{\begin{aligned} x&=2\alpha+2\beta \\ y&=\alpha \\ z&=\beta \end{aligned}\right.$$

y por lo tanto

$$\operatorname{Ker}\left((f-I)^2\right)=\langle\{(2,1,0),(2,0,1)\}\rangle.$$

El vector v_2 debe pertenecer a $\operatorname{Ker}\left((f-I)^2\right)=\langle\{(2,1,0),(2,0,1)\}\rangle$, así que tomemos por ejemplo el vector $v_2=(2,1,0)$. Lo otro que necesitamos de v_2 es que cuando le aplicamos $(f-I)$ nos debía quedar un autovector v_1 asociado al autovalor 1. Si calculamos las coordenadas de $(f-I)(v_2)$ respecto de la base canónica nos queda

$$\underbrace{\begin{bmatrix} -1 & 2 & 2 \\ 2 & -1 & -1 \\ -1 & -1 & -1 \end{bmatrix}}_{(C-I)}\begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}=\begin{bmatrix} 0 \\ 3 \\ -3 \end{bmatrix}$$

¿Es el vector $(0,3,-3)$ un autovector asociado al autovalor 1? Pues sí, porque ya sabíamos que $E_1(C)=\langle\{(0,-1,1)\}\rangle$. Así que podemos fijar $v_1=(0,3,-3)$.

Finalmente, como al vector v_3 lo único que le teníamos que pedir es que fuese un autovector en $E_{-2}(C)$, podemos fijar la siguiente base de $\mathbb{R}^3$,

$$B=\{\underbrace{(0,3,-3)}_{v_1},\underbrace{(2,1,0)}_{v_2},\underbrace{(1,-1,0)}_{v_3}\}.$$

¿Cuál es la matriz de f respecto de la base B? Para calcular la matriz de f en la base B debemos hacer el producto

$$P^{-1}AP$$

donde P es la matriz de cambio de base de B a la canónica, es decir,

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

Realizando el producto nos queda que

$$\underbrace{\begin{bmatrix} 0 & 2 & 1 \\ 3 & 1 & -1 \\ -3 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 & 2 & 2 \\ 2 & 0 & -1 \\ -1 & -1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 2 & 1 \\ 3 & 1 & -1 \\ -3 & 0 & 0 \end{bmatrix}}_{P^{-1}CP} = \underbrace{\begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -2 \end{bmatrix}}_{J}$$

El hecho de que nos haya quedado la matriz J no es una casualidad, ha ocurrido precisamente porque los vectores v_1, v_2, v_3 los hemos buscado precisamente de forma que satisfagan

$$\begin{cases} f(v_1) = (1,0,0)_B & = v_1 \\ f(v_2) = (1,1,0)_B & = v_1 + v_2 \\ f(v_3) = (0,0,-2)_B & = -2v_3. \end{cases}$$

¿Podíamos haber tomado como v_2 cualquier vector de

$$\operatorname{Ker}\left((f-I)^2\right) = \langle\{(2,1,0),(2,0,1)\}\rangle?$$

No (pero casi). Si nos damos cuenta, el vector $(0,-1,1)$, que es el generador de $E_1(C)$, pertenece a $\operatorname{Ker}\left((f-I)^2\right)$ ya que

$$(0,-1,1) = -(2,1,0) + (2,0,1)$$

es decir, $\operatorname{Ker}(f-I) \subset \operatorname{Ker}\left((f-I)^2\right)$. Si nos hubiese dado por escoger precisamente $v_2 = (0,-1,1)$ entonces habríamos tenido un problema, porque nosotros queríamos que $(f-I)(v_2)$ fuese un autovector v_1 asociado al autovalor 1 (pero no nulo, porque queremos usar v_1 como un vector de una base), y sin embargo

$$(f-I)(v_2) = 0$$

ya que el propio vector v_2 era un autovector asociado al autovalor 1. La única precaución que tenemos que tomar es precisamente la de escoger v_2 en $\operatorname{Ker}\left((f-I)^2\right)$ pero que no estuviese en $\operatorname{Ker}(f-I)$.

Definición 3.5 *Dada $A \in \mathcal{M}_n(\mathbb{K})$, si existe una matriz de Jordan $J \in \mathcal{M}_n(\mathbb{K})$ semejante a A, decimos que J es una forma canónica de Jordan de A.*

Nuestro objetivo es probar el siguiente teorema.

Teorema 3.13 *Sea $f : V \to V$ un operador lineal de un $\mathbb{K}$-espacio vectorial V. Supongamos que todas raíces $\lambda_1, \dots, \lambda_r$ del polinomio característico de f pertenecen a $\mathbb{K}$. Entonces existe una base B de V respecto de la cual f tiene una matriz de Jordan.*

O dicho en otras palabras, dada una matriz $A \in \mathbb{K}^n$, si todos sus autovalores pertenecen a $\mathbb{K}$ entonces A es semejante a una matriz de Jordan.

En la siguiente sección se redefinen conceptos previos, pero en función de operadores lineales.

3.2.4 Subespacios propios generalizados

En esta sección V denota una $\mathbb{K}$-espacio vectorial.

V. Osores

Definición 3.6 *Sea $f : V \to V$ un operador lineal y sea $\lambda \in \mathbb{K}$ un autovalor. Los subespacios propios generalizados de λ son*

$$E_\ell(\lambda) := \operatorname{Ker}(f - \lambda I)^\ell$$

para cada $\ell \in \mathbb{N}$.

Observaciones:

1. Por definición $E_1(\lambda)$ es el subespacio propio de λ.

2. Los subespacios propios generalizados forman una cadena

$$E_1(\lambda) \subset E_2(\lambda) \cdots \subset E_k(\lambda) \subset \cdots$$

En efecto, se tiene que dado $v \in E_k(\lambda)$ entonces $v \in E_{k+1}(\lambda)$ ya que

$$\begin{aligned} v \in E_k(\lambda) &\Longrightarrow (f - \lambda I)^k(v) = 0 \\ &\Longrightarrow (f - \lambda I)(f - \lambda I)^k(v) = (f - \lambda I)(0) = 0 \\ &\Longrightarrow (f - \lambda I)^{k+1}(v) = 0 \end{aligned}$$

y por tanto $v \in E_{k+1}(\lambda)$.

3. Esta cadena no puede ser infinita, es decir, a partir de un momento todos los subespacios propios generalizados deben ser iguales. Claro, cada $E_\ell(\lambda)$ es un subespacio de V y por tanto $\dim(E_\ell(\lambda)) \leq n$. Deducimos que sus dimensiones no pueden crecer indefinidamente, es decir que para algún ℓ vamos a tener que $\dim(E_\ell(\lambda)) = \dim(E_{\ell+1}(\lambda))$ y por tanto $E_\ell(\lambda) = E_{\ell+1}(\lambda)$. Es más, como $E_\ell(\lambda) = E_{\ell+1}(\lambda)$ vamos a tener que

$$E_\ell(\lambda) = E_{\ell+1}(\lambda) = E_{\ell+2}(\lambda) = \cdots .$$

Definición 3.7 *Al número natural ℓ más pequeño tal que $E_\ell(\lambda) = E_{\ell+1}(\lambda)$ se le llama índice de nilpotencia de λ, se denota por **nil**(λ), y al subespacio propio generalizado $E_\ell(\lambda)$ se le llama subespacio máximo de λ y se denota por $M(\lambda) = E_\ell(\lambda)$.*

Ejemplo 3.9 *Sea $f : \mathbb{R}^3 \to \mathbb{R}^3$ el operador lineal cuya matriz respecto de la base canónica es la matriz*

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

Ya sabemos que el polinomio característico es

$$p(\lambda) = (-\lambda - 2)\left(\lambda^2 - 2\lambda + 1\right) = (-\lambda - 2)(\lambda - 1)^2$$

y los autovalores son -2 y 1.

Calculemos primero los subespacios propios generalizados del autovalor $\lambda = -2$,

$$E_1(-2) = \langle\{(1,-1,0)\}\rangle.$$

Calculemos $E_2(-2)$. La matriz de $(f+2I)^2$ es $(C+2I)^2$ y por tanto debemos calcular todos los vectores $(x,y,z) \in \mathbb{R}^3$ que satisfacen

$$\underbrace{\begin{bmatrix} 6 & 6 & 6 \\ 9 & 9 & 0 \\ -6 & -6 & 3 \end{bmatrix}}_{(C+2I)^2} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

es decir,

$$\left\{ \begin{aligned} 6x+6y+6z &= 0 \\ 9x+9y &= 0 \\ -6x+-6y+3z &= 0 \end{aligned} \right. \Rightarrow \left\{ \begin{aligned} x+y+z=0 \\ x+y=0 \\ -2x+-2y+z=0 \end{aligned} \right. \Rightarrow \left\{ \begin{aligned} y=-x \\ z=0 \end{aligned} \right.$$

y por tanto

$$E_2(-2) = \operatorname{Ker}\left((f+2I)^2\right) = \langle\{(1,-1,0)\}\rangle.$$

Como $E_1(-2) = E_2(-2)$ el índice de nilpotencia de $\lambda = -2$ es 1 y su subespacio maximal es $M(-2) = E_1(-2) = \langle\{(1,-1,0)\}\rangle$.

Calculemos ahora los subespacios propios generalizados del autovalor $\lambda = 1$. Ya vimos que $E_1(1) = \langle\{(0,-1,1)\}\rangle$ y $E_2(1) = \operatorname{Ker}\left((f-I)^2\right) = \langle\{(2,1,0),(2,0,1)\}\rangle$. Observe que, como ya vimos en teóricamente, el subespacio $E_1(1)$ está contenido en el subespacio $E_2(1)$.

Calculemos ahora $E_3(1)$. La matriz de $(f-I)^3$ es $(C-I)^3$ y por tanto debemos calcular todos los vectores $(x,y,z) \in \mathbb{R}^3$ que satisfacen

$$\underbrace{\begin{bmatrix} -9 & 18 & 18 \\ 9 & -18 & -18 \\ 0 & 0 & 0 \end{bmatrix}}_{(C-I)^3} \begin{bmatrix} x \\ y \\ z \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$$

es decir,

$$\left\{ \begin{aligned} -9x+18y+18z &= 0 \\ 9x-18y-18z &= 0 \\ 0 &= 0 \end{aligned} \right. \to \left\{ \begin{aligned} x-2y-2z &= 0 \\ 0 &= 0 \\ 0 &= 0 \end{aligned} \right.$$

y por tanto

$$E_3(1) = \operatorname{Ker}\left((f-I)^3\right) = \langle\{(2,1,0),(2,0,1)\}\rangle.$$

Como $E_2(1) = E_3(1)$ el índice de nilpotencia de $\lambda = 1$ es 2 y su subespacio maximal es $M(1) = E_2(1)$.

Los subespacios máximos van a jugar el papel que jugaban en el tema anterior los subespacios propios (es decir, el de los autovectores). De hecho, en esta ocasión vamos a probar que los subespacios máximos tienen siempre las siguientes propiedades.

Teorema 3.14 (subespacios máximos) *Sea V un $\mathbb{K}$-espacio vectorial y sea $f : V \to V$ un operador lineal. Supongamos que todas las raíces $\lambda_1, \ldots, \lambda_r$ del polinomio característico de f pertenecen a $\mathbb{K}$. Entonces se tiene que:*

a) $V = M(\lambda_1) \oplus \cdots \oplus M(\lambda_r)$.

b) $\operatorname{mult}_a(\lambda_i) = \dim(M(\lambda_i))$ *para todo* $i = 1, \ldots, r$.

Lo primero que observamos es que si en el teorema anterior estamos trabajando sobre $\mathbb{K} = \mathbb{C}$ entonces la hipótesis de que todos los autovalores pertenecen a $\mathbb{C}$ se satisface siempre. Esa hipótesis solo supone una limitación cuando estamos trabajando sobre $\mathbb{R}$.

3.2.5 Diagramas de puntos

Sea $f : V \to V$ un operador lineal que tiene un único autovalor $\lambda \in \mathbb{K}$, como el polinomio característico tiene coeficientes reales, entonces el conjugado de esa raíz también sería ser una raíz, pero hemos supuesto que solo tenemos uno. Y supongamos que f tiene respecto de una cierta base $B = \{v_1, \ldots, v_n\}$ una matriz J de Jordan. Intercambiando la posición de los vectores, es evidente que podemos suponer que los bloques de Jordan que aparecen en J van de mayor a menor orden. Por ejemplo, la matriz

$$J = \left[\begin{array}{ccc:cc:cc:c} \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & \lambda & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{array}\right]$$

A continuación vamos a mostrar cómo asociar un diagrama de puntos a la matriz J, de forma que nos permita entender mejor su comportamiento. Expliquemos el método a través de un ejemplo usando la matriz J anterior. Es decir, que tenemos un operador lineal $f : \mathbb{K}^8 \to \mathbb{K}^8$ y una base $B = \{v_1, \ldots, v_8\}$ respecto de la cual su matriz asociada es J. El primer bloque nos indica que v_1 es un autovector, es decir $v_1 \in \operatorname{Ker}(f - \lambda I) = E_1(\lambda)$, y que se tienen las dos siguientes igualdades

$$\begin{cases} f(v_2) = v_1 + \lambda v_2 \Longleftrightarrow v_1 = (f - I\lambda)(v_2) \\ f(v_3) = v_2 + \lambda v_3 \Longleftrightarrow v_2 = (f - I\lambda)(v_3) \end{cases}$$

(note que esta elección es un caso mas delicado que en 3.2). Es decir, que v_1 lo obtenemos aplicando al vector v_2 la aplicación $f - \lambda I$ y a su vez v_2 lo obtenemos aplicando al vector v_3 la aplicación $f - \lambda I$. Dicho esto, escribimos el siguiente diagrama de puntos

$$\bullet_{v_1} \quad \longleftarrow \quad \bullet_{v_2} \quad \longleftarrow \quad \bullet_{v_3}$$

Los puntos representan, de izquierda a derecha, los vectores v_1, v_2, v_3 y la flecha $\longleftarrow$ representa que el vector de la izquierda se calcula aplicando la función $f - \lambda I$ al vector de la derecha. Observemos que

$v_2 \in E_2(\lambda) = \mathrm{Ker}(f - \lambda I)^2$ dado que

$$(f - I\lambda)^2\,(v_2) = (f - \lambda I)(f - \lambda I)\,(v_2) = (f - \lambda I)\,(v_1) = 0,$$

pero $v_2 \notin E_1(\lambda) = \mathrm{Ker}(f - \lambda I)$, ya que en caso contrario se tendría que $v_1 = (f - \lambda I)\,(v_2) = 0$, lo cual es absurdo porque v_1 forma parte de una base. Razonando de forma parecida tenemos que $v_3 \in E_3(\lambda) = \mathrm{Ker}(f - \lambda I)^3$ pero $v_3 \notin E_2(\lambda) = \mathrm{Ker}(f - \lambda I)^2$. Esto lo podemos representar poniendo encima de cada punto

$$\begin{array}{ccccc} E_1(\lambda) & \hookrightarrow & E_2(\lambda) & \hookrightarrow & E_3(\lambda) \\ \bullet_{v_1} & \longleftarrow & \bullet_{v_2} & \longleftarrow & \bullet_{v_3} \end{array}$$

Ahora pasamos al siguiente bloque de orden 2. Razonando de forma parecida a como hemos hecho anteriormente, tenemos que $v_4 \in E_1(\lambda)$ es un autovector y $v_4 = (f - \lambda I)\,(v_5)$. Y de nuevo $v_5 \in E_2(\lambda)$ pero $v_5 \notin E_1(\lambda)$. Para representar este nuevo bloque escribimos una línea de puntos debajo de la anterior,

$$\begin{array}{ccccc} E_1(\lambda) & \hookrightarrow & E_2(\lambda) & \hookrightarrow & E_3(\lambda) \\ \bullet_{v_1} & \longleftarrow & \bullet_{v_2} & \longleftarrow & \bullet_{v_3} \\ \bullet_{v_4} & \longleftarrow & \bullet_{v_5} & & \end{array}$$

Lo mismo hacemos con el último bloque de orden 2 y el bloque de orden 1, quedando finalmente el diagrama como

$$\begin{array}{ccccc} E_1(\lambda) & \hookrightarrow & E_2(\lambda) & \hookrightarrow & E_3(\lambda) \\ \bullet_{v_1} & \longleftarrow & \bullet_{v_2} & \longleftarrow & \bullet_{v_3} \\ \bullet_{v_4} & \longleftarrow & \bullet_{v_5} & & \\ \bullet_{v_6} & \longleftarrow & \bullet_{v_7} & & \\ \bullet_{v_8} & & & & \end{array}$$

De esta forma hemos llegado a un diagrama de puntos, tal que:

1. Los puntos representan exactamente $\dim(V)$ vectores linealmente independientes de V, que son por tanto una base de V,

2. La flecha $\longleftarrow$ simboliza que el vector que representa el punto de la izquierda de la flecha se obtiene aplicando a $(f - \lambda I)$ al vector que representa el punto de la derecha de la flecha.

3. Los puntos en la columna debajo de $E_i(\lambda)$ representan vectores que pertenecen a $E_i(\lambda)$ (en realidad solo es necesario para $i = 1$, para el resto vendrá dado por 1. y 2.

Tenemos una propiedad más fuerte que 3.: Los puntos en la columna debajo de $E_i(\lambda)$ representan vectores que pertenecen a $E_i(\lambda)$, y ninguno de ellos pertenece a $E_{i-1}(\lambda)$.

De hecho, trabajando un poquito más, podemos llegar a que: Ninguna combinación lineal de los vectores que están representados por puntos que están debajo de un $E_i(\lambda)$ pertenece a $E_{i-1}(\lambda)$, excepto la combinación nula.

En efecto, ya hemos argumentado que el punto que aparece en la columna de la derecha representa un vector en $E_3(\lambda)$ que no pertenece a $E_2(\lambda)$, y por tanto ningún múltiplo suyo (excepto el nulo) pertenecerá a $E_2(\lambda)$. Veamos ahora que en nuestro ejemplo los tres puntos que aparecen en la columna central representan tres vectores que pertenecen a $E_2(\lambda)$ con la propiedad de que ninguna combinación lineal de ellos no nula pertenece a $E_1(\lambda)$. Si existieran $a, b, c \in \mathbb{K}$ tales que

$$av_2 + bv_5 + cv_7 \in E_1(\lambda)$$

entonces

$$a(f - \lambda I)(v_2) + b(f - \lambda I)(v_5) + c(f - \lambda I)(v_7) = 0$$

es decir

$$av_1 + bv_4 + cv_6 = 0$$

de lo que deducimos $a = b = c = 0$ porque los vectores v_1, v_4, v_6 son linealmente independientes. Observe que este argumento va a funcionar en general, porque dados los vectores representados por puntos en una columna debajo de $E_i(\lambda)$, si una combinación de ellos pertenece a $E_{i-1}(\lambda)$, entonces la combinación con los mismos coeficientes de sus imágenes por medio de $(f - \lambda I)^{i-1}$ nos va a dar 0, lo cual implica que los coeficientes son todos cero ya que esas imágenes sabíamos que eran linealmente independientes.

El papel que ha jugado la base B es ciertamente auxiliar: solo hemos utilizado las propiedades que los vectores de dicha base deben tener si f respecto de ella tiene matriz J, no nos importan quienes son los vectores en concreto. Así pues, escribiremos el diagrama anterior como

$$\begin{array}{ccccc}
E_1(\lambda) & \hookrightarrow & E_2(\lambda) & \hookrightarrow & E_3(\lambda) \\
\bullet & \longleftarrow & \bullet & \longleftarrow & \bullet \\
\bullet & \longleftarrow & \bullet & & \\
\bullet & \longleftarrow & \bullet & & \\
\bullet & & & &
\end{array}$$

Eso sí, puesto que cada punto corresponde a un vector de la base, lo único que sí deberíamos tener presente en adelante es cuál es el orden que ocupan dichos vectores en la base. La respuesta es fácil: debemos empezar por la primera línea y ordenar de izquierda a derecha, para luego continuar con la segunda línea, de izquierda a derecha, y así sucesivamente.

El diagrama de puntos es útil porque nos permite calcular fácilmente las dimensiones de los subespacios propios generalizados. En efecto, si contamos los puntos es evidente que $\dim(E_3(\lambda)) = 8$. El punto que aparece en la columna de la derecha representa un vector que no pertenece a $E_2(\lambda)$, y dado que en $E_2(\lambda)$ tenemos al menos 7 vectores linealmente independientes, deducimos $\dim(E_2(\lambda)) = 7$. Vayamos con la dimensión de $E_1(\lambda)$. Ya sabemos que los vectores que representan los puntos en la columna debajo de $E_1(\lambda)$ son linealmente independientes y por tanto al menos sabemos que $\dim(E_1(\lambda)) \geq 4$. Queremos ver que de hecho se da la igualdad. También sabemos que los tres puntos que aparecen en la columna central representan tres vectores de $E_2(\lambda)$ con la propiedad de que ninguna combinación lineal de ellos pertenece a $E_1(\lambda)$. Denotemos con W al subespacio que generan esos tres vectores. Por tanto sabemos que $E_1(\lambda) \bigoplus W = E_2(\lambda)$, de lo que deducimos que

$$\dim(E_1(\lambda)) + \dim(W) = \dim(E_2(\lambda)) \quad \Rightarrow \quad \dim(E_1(\lambda)) + 3 = 7$$

es decir, $\dim(E_1(\lambda)) = 4$. De lo anterior se concluye que los puntos de las columnas sumados de izquierda a derecha nos da la dimensión de los subespacios generalizados,

$$\dim(E_1(\lambda)) = 4, \quad \dim(E_2(\lambda)) = 7, \quad \dim(E_3(\lambda)) = 8.$$

Y esto de nuevo es general, solo hemos usado la propiedad de que ninguna combinación lineal de los vectores que están representados por puntos que están debajo de un $E_i(\lambda)$ pertenece a $E_{i-1}(\lambda)$, excepto la combinación nula. En particular, tenemos la siguiente consecuencia:

Corolario 3.2 *Dado f un operador lineal con un único autovalor, si f tiene matrices de Jordan respecto de dos bases distintas, entonces dichas matrices deben tener los mismos bloques de Jordan (salvo permutación).*

Demostración: *Si ordenamos los bloques de mayor a menor de dicha matrices y hacemos sus diagramas nos deben quedar iguales. En caso contrario, tendríamos dos dimensiones distintas para algún subespacio propio generalizado de f, lo cual es absurdo.*

Pero incluso más importante para nosotros es que de igual manera que hemos asociado a una matriz de Jordan un diagrama de puntos, podemos asociar a un diagrama de puntos una matriz de Jordan. Así por ejemplo, si nos dicen que tenemos un operador lineal $f : V \to V$, donde $\dim(V) = 8$, y que tenemos una base B de V de forma que asociando vectores a puntos podemos construir el siguiente diagrama con las reglas 1., 2. y 3. anteriores

$$\begin{array}{ccccc} E_1(\lambda) & \hookrightarrow & E_2(\lambda) & \hookrightarrow & E_3(\lambda) \\ \bullet & \longleftarrow & \bullet & \longleftarrow & \bullet \\ \bullet & \longleftarrow & \bullet & \longleftarrow & \bullet \\ \bullet & \longleftarrow & \bullet & & \end{array}$$

entonces f respecto de esa base B tiene matriz de Jordan

$$J = \left[\begin{array}{ccc:ccc:cc} \lambda & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \lambda & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & \lambda & 0 & 0 & 0 & 0 & 0 \\ \hdashline 0 & 0 & 0 & \lambda & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & \lambda & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & \lambda & 0 & 0 \\ \hdashline 0 & 0 & 0 & 0 & 0 & 0 & \lambda & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \lambda \end{array}\right]$$

En efecto, tenemos tantos bloques como líneas tiene nuestro diagrama, y el orden de cada bloque viene dado por el número de puntos en la línea. Además vemos que el número de puntos debajo de la columna $E_1(\lambda)$ debe ser igual a $\dim(E_1(\lambda))$, y el número de puntos debajo de la columna $E_i(\lambda)$ debe ser igual $\dim(E_{i-1}(\lambda)) - \dim(E_i(\lambda))$ para $i > 1$.

Con esta idea en la cabeza, veamos cómo construir una base B de cualquier operador lineal $f : V \to V$ con un único autovalor respecto de la cual la matriz de f es una matriz de Jordan.

Teorema 3.15 (Jordan con un único autovalor) *Sea $f : V \to V$ un operador lineal con un único autovalor $\lambda \in \mathbb{K}$. Entonces existe una base B respecto de la cual f tiene una matriz de Jordan.*

Demostración: *Pendiente...*

Ejemplo 3.10 *Sea el operador lineal $f : \mathbb{R}^5 \to \mathbb{R}^5$ cuya matriz respecto de la base canónica es*

$$A = \begin{bmatrix} -2 & 0 & 3 & 4 & 5 \\ 0 & -2 & 0 & 6 & 7 \\ 0 & 0 & -2 & 0 & 8 \\ 0 & 0 & 0 & -2 & 0 \\ 0 & 0 & 0 & 0 & -2 \end{bmatrix}$$

Evidentemente su único autovalor es $\lambda = -2$. Calculemos primero las dimensiones de los subespacios propios generalizados. Para ello tenemos que calcular las distintas potencias de $A - (-2)I$, que son

$$[A-(-2)I]^2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 24 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad [A-(-2)I]^3 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

Así pues,

$$\begin{cases} \dim\left(E_1(-2)\right) = 5 - \operatorname{rg}(A-(-2)I) = 2 \\ \dim\left(E_2(-2)\right) = 5 - \operatorname{rg}(A-(-2)I)^2 = 4 \\ \dim\left(E_3(-2)\right) = 5 - \operatorname{rg}(A-(-2)I)^3 = 5 \end{cases}$$

Luego, el diagrama de puntos será

$$\begin{array}{ccccc} E_1(-2) & \hookrightarrow & E_2(-2) & \hookrightarrow & E_3(-2) \\ \bullet & \longleftarrow & \bullet & \longleftarrow & \bullet \\ \bullet & \longleftarrow & \bullet & & \end{array}$$

y la matriz A será semejante a la matriz de Jordan

$$J = \left[\begin{array}{ccc:cc} -2 & 1 & 0 & 0 & 0 \\ 0 & -2 & 1 & 0 & 0 \\ 0 & 0 & -2 & 0 & 0 \\ \hdashline 0 & 0 & 0 & -2 & 1 \\ 0 & 0 & 0 & 0 & -2 \end{array}\right].$$

Calculemos una base $B = \{v_1, v_2, v_3, v_4, v_5\}$ de $\mathbb{R}^5$ respecto de la cual f tiene la matriz J anterior. Primero encontramos una base de $E_2(-2)$, es decir, buscamos $(x, y, z, t, w) \in \mathbb{R}^5$ tales que

$$\underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 24 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{[A-(-2)I]^2} \begin{bmatrix} x \\ y \\ z \\ t \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

y por tanto obtenemos

$$E_2(-2) = \langle \{(1,0,0,0,0), (0,1,0,0,0), (0,0,1,0,0), (0,0,0,1,0)\} \rangle.$$

A continuación, buscamos un vector que añadido a la base anterior de $E_2(-2)$ que hemos calculado, nos de una base de $E_3(-2) = \mathbb{R}^5$. Basta tomar el vector

$$v_3 = (0, 0, 0, 0, 1)$$

Este vector será el que representa el punto debajo de $E_3(-2)$, por eso lo hemos denotado con v_3. A continuación, calculamos su imágen por medio de $f - (-2)I$, obteniendo

$$v_2 = (f - (-2)I)(v_3) = (5, 7, 8, 0, 0).$$

Ahora calculamos una base de $E_1(-2)$, es decir, buscamos $(x, y, z, t, w) \in \mathbb{R}^5$ tales que

$$\underbrace{\begin{bmatrix} 0 & 0 & 3 & 4 & 5 \\ 0 & 0 & 0 & 6 & 7 \\ 0 & 0 & 0 & 0 & 8 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{[A-(-2)I]} \begin{bmatrix} x \\ y \\ z \\ t \\ w \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

obteniendo

$$E_1(-2) = \langle \{(1, 0, 0, 0, 0), (0, 1, 0, 0, 0)\} \rangle.$$

Ya sabemos que el vector v_2 es independiente de $(1, 0, 0, 0, 0)$ y $(0, 1, 0, 0, 0)$, y necesitamos un vector que unido a estos tres vectores nos de una base de $E_2(-2)$. Podemos tomar

$$v_5 = (0, 0, 0, 1, 0).$$

Por último, calculamos las imágenes de v_2 y v_5 por medio de $f - (-2)I$ para obtener los vectores v_1 y v_4 respectivamente,

$$\begin{cases} v_1 = (f - (-2)I)(v_2) = (24, 0, 0, 0, 0) \\ v_4 = (f - (-2)I)(v_5) = (4, 6, 0, 0, 0) \end{cases}.$$

Finalmente, hemos llegado a que la matriz de f respecto de la base

$$B = \{\underbrace{(24, 0, 0, 0, 0)}_{v_1}, \underbrace{(5, 7, 8, 0, 0)}_{v_2}, \underbrace{(0, 0, 0, 0, 1)}_{v_3}, \underbrace{(4, 6, 0, 0, 0)}_{v_4}, \underbrace{(0, 0, 0, 1, 0)}_{v_5}\}$$

es la matriz de Jordan J.

A continuación abordaremos el Teorema de Jordan en el caso general de un operador lineal $f : V \to V$ cualquiera (no necesariamente con un único autovalor). Necesitamos el siguiente resultado que nos va a permitir reducirlo todo al caso de tener un único autovalor (que es el caso que ya sabemos hacer).

Proposición 3.1 *Sea $f : V \to V$ un operador lineal, y sea $\lambda \in \mathbb{K}$ un autovalor. Entonces los subespacios propios generalizados $E_k(\lambda)$ son subespacios invariantes de f. En particular, el subespacio máximo $M(\lambda)$ es un subespacio invariante de f.*

Demostración: *Dado $v \in E_k(\lambda)$, queremos probar que $f(v) \in E_k(\lambda)$. Se tiene que*

$$(f - \lambda I)(v) = f(v) - \lambda v \Rightarrow f(v) = v + (f - \lambda I)(v).$$

Por un lado tenemos que $v \in E_k(\lambda)$ pero también $(f - \lambda I)(v) \in E_{k-1}(\lambda) \subset E_k(\lambda)$, y puesto que $E_k(\lambda)$ es un subespacio vectorial de V, deducimos que $f(v) = v + (f - \lambda I)(v) \in E_k(\lambda)$.

Teorema 3.16 *Sea $f : V \to V$ un operador lineal de un $\mathbb{K}$-espacio vectorial V. Supongamos que todas raíces $\lambda_1, \ldots, \lambda_r$ del polinomio característico de f pertenecen a $\mathbb{K}$. Entonces existe una base B de V respecto de la cual f tiene una matriz de Jordan.*

Demostración: *Consideremos para cada λ_i su subespacio máximo $M(\lambda_i)$. Por la proposición anterior, $M(\lambda_i)$ es un subespacio invariante de f. Por tanto, podemos considerar el operador lineal*

$$f|_{M(\lambda_i)} : M(\lambda_i) \to M(\lambda_i).$$

El único autovalor de $f|_{M(\lambda_i)}$ es λ_i, ya que si tuviese otro autovalor $\mu \neq \lambda_i$, entonces por definición existiría un vector $w \in M(\lambda_i)$ no nulo tal que $f|_{M(\lambda_i)}(w) = f(w) = \mu w$. Es decir μ debe ser un autovalor de f, por tanto $\mu = \lambda_j$ con $i \neq j$, y entonces $w \in E_1(\lambda_j) \subset M(\lambda_j)$, pero ya sabemos que $M(\lambda_i) \cap M(\lambda_j) = 0$ por el Teorema de los subespacios máximos, contradicción. Por lo visto en la sección anterior, podemos encontrar una base B_i de $M(\lambda_i)$ tal que $f|_{M(\lambda_i)}$ tiene una matriz de Jordan J_i. Ahora bien, si consideramos la unión de todas esas bases

$$B = B_1 \cup \cdots \cup B_r$$

nos proporciona una base de $M(\lambda_1) \oplus \cdots \oplus M(\lambda_r)$. Por el Teorema de los subespacios máximos también sabemos que $M(\lambda_1) \oplus \cdots \oplus M(\lambda_r) = V$ y por tanto B es una base de V. Finalmente, la matriz de f respecto de la base B es la matriz de Jordan,

$$\left[\begin{array}{c:c:c} J_1 & & \\ \hdashline & \ddots & \\ \hdashline & & J_r \end{array}\right].$$

Ejemplo 3.11 *Sea la matriz*

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

El polinomio característico es

$$\begin{aligned} p(\lambda) = \det(A - \lambda I) = & \begin{vmatrix} 1-\lambda & 0 & 0 & 0 \\ -1 & 2-\lambda & 1 & 0 \\ -1 & 0 & 2-\lambda & 0 \\ -1 & 0 & 1 & 1-\lambda \end{vmatrix} \\ =& (1-\lambda) \begin{vmatrix} 2-\lambda & 1 & 0 \\ 0 & 2-\lambda & 0 \\ 0 & 1 & 1-\lambda \end{vmatrix} \\ =& (1-\lambda)(2-\lambda) \begin{vmatrix} 2-\lambda & 0 \\ 1 & 1-\lambda \end{vmatrix} \\ =& (1-\lambda)^2(2-\lambda)^2. \end{aligned}$$

Por tanto los autovalores son $\lambda_1 = 1$ *y* $\lambda_2 = 2$ *ambos con multiplicidad algebraica 2.*

Calculamos primero $E_1(1)$*, es decir, los autovectores de* $\lambda_1 = 1$*. Buscamos* $(x, y, z, t) \in \mathbb{R}^4$ *tales que*

$$\begin{bmatrix} 0 & 0 & 0 & 0 \\ -1 & 1 & 1 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow \begin{cases} 0 = 0 \\ -x + y + z = 0 \\ -x + z = 0 \\ -x + z = 0 \end{cases}$$

de lo que deducimos que

$$E_1(1) = \langle \{(1, 0, 1, 0), (0, 0, 0, 1)\} \rangle.$$

Como la dimensión de $E_1(1)$ *coincide con la multiplicidad algebraica de* $\lambda_1 = 1$ *ya sabemos entonces que* $M(1) = E_1(1)$*.*

Ahora calculemos los subespacios propios de $\lambda_2 = 2$*. Calculamos primero* $E_1(2)$*, es decir, los autovectores de* $\lambda_2 = 2$*. Buscamos* $(x, y, z, t) \in \mathbb{R}^4$ *tales que*

$$\begin{bmatrix} -1 & 0 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 0 \\ -1 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow \begin{cases} -x = 0 \\ -x + z = 0 \\ -x = 0 \\ -x + z - t = 0 \end{cases}$$

de lo que concluimos que

$$E_1(2) = \langle \{(0, 1, 0, 0)\} \rangle.$$

De forma inmediata ya sabemos entonces que $\dim(E_2(2)) = 2$ *y* $M(2) = E_2(2)$*. Pero busquemos una base de* $E_2(2)$*.*

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 1 \end{bmatrix}}_{(A-2I)^2} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow \begin{cases} x = 0 \\ 0 = 0 \\ x = 0 \\ x - z + t = 0 \end{cases}$$

de lo que deducimos que

$$E_2(2) = \langle \{(0, 0, 1, 1), (0, 1, 0, 0)\} \rangle.$$

Con los cálculos anteriores ya sabemos que la forma canónica de Jordan de A va a ser

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

Cálculo de la matriz P: Para los bloques de orden 1 correspondientes a $M(1)$ podemos tomar la base que hemos calculado anteriormente

$$\{(0,0,0,1),(1,0,1,0)\}.$$

Para el bloque de Jordan de orden 2 correspondiente a $M(2)$, debemos aplicar lo que se hizo antes. Así pues, buscamos una base de $E_1(2)$ y la completamos hasta obtener una base de $M(2) = E_2(2)$. Observamos que de hecho las mismas bases que nos han quedado después de los cálculos nos sirven. Tomamos por tanto el vector $(0,0,1,1)$ y calculamos su imagen por medio de $A-2I$, que da la casualidad que nos queda $(0,1,0,0)$. Tenemos por tanto que el operador que tiene matriz asociada A respecto de la canónica, tiene respecto de la base

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

la matriz J anterior. Así pues, dada la matriz de cambio de base de B a canónica,

$$P = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 \end{bmatrix}$$

tenemos la igualdad $A = PJP^{-1}$.

Ejemplo 3.12 *Sea la matriz*

$$A = \begin{bmatrix} 2 & 2 & -3 & 4 \\ -2 & 2 & 1 & 0 \\ 3 & 3 & -5 & 7 \\ 4 & 2 & -6 & 7 \end{bmatrix}.$$

El polinomio característico es

$$p(\lambda) = \det(A-\lambda I) = \begin{vmatrix} 2-\lambda & 2 & -3 & 4 \\ -2 & 2-\lambda & 1 & 0 \\ 3 & 3 & -5-\lambda & 7 \\ 4 & 2 & -6 & 7-\lambda \end{vmatrix} = (1-\lambda)^2(2-\lambda)^2.$$

Y por tanto los autovalores son $\lambda_1 = 1$ y $\lambda_2 = 2$ ambos con multiplicidad algebraica 2.

Calculamos primero $E_1(1)$, es decir, los autovectores de $\lambda_1 = 1$. Buscamos $(x,y,z,t) \in \mathbb{R}^4$ tales que

$$\begin{bmatrix} 1 & 2 & -3 & 4 \\ -2 & 1 & 1 & 0 \\ 3 & 3 & -6 & 7 \\ 4 & 2 & -6 & 6 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \rightarrow \begin{cases} x+2y-3z+4t=0 \\ -2x+y+z=0 \\ 3x+3y-6z+7t=0 \\ 4x+2y-6z+6t=0 \end{cases}$$

V. Osores

porlo tanto

$$E_1(1) = \langle\{(1,1,1,0)\}\rangle.$$

En particular eso nos indica que $\dim(E_2(1)) = 2$ *y* $M(1) = E_2(1)$. *Busquemos una base de* $E_2(2)$.

$$\underbrace{\begin{bmatrix} 4 & 3 & -7 & 7 \\ -1 & 0 & 1 & -1 \\ 7 & 5 & -12 & 12 \\ 6 & 4 & -10 & 10 \end{bmatrix}}_{(A-2I)^2} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

de lo que se conlcuye que

$$E_2(1) = \langle\{(1,1,0,-1),(0,0,1,1)\}\rangle.$$

Cálculo de los subespacios propios de $\lambda_2 = 2$. *Calculamos primero* $E_1(2)$, *es decir, los autovectores de* $\lambda_2 = 2$. *Buscamos* $(x,y,z,t) \in \mathbb{R}^4$ *tales que*

$$\begin{bmatrix} 0 & 2 & -3 & 4 \\ -2 & 0 & 1 & 0 \\ 3 & 3 & -7 & 7 \\ 4 & 2 & -6 & 5 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

de lo que deducimos que

$$E_1(2) = \langle\{(1,-1,2,2)\}\rangle.$$

De forma inmediata ya sabemos entonces que $\dim(E_2(2)) = 2$ *y* $M(2) = E_2(2)$. *Pero busquemos una base de* $E_2(2)$.

$$\underbrace{\begin{bmatrix} 3 & -1 & -1 & -1 \\ 3 & -1 & -1 & -1 \\ 1 & -1 & 1 & -2 \\ -2 & 0 & 2 & -1 \end{bmatrix}}_{(A-2I)^2} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

de lo que obtenemos

$$E_2(2) = \langle\{(1,2,1,0),(0,3,-1,-2)\}\rangle.$$

Con los cálculos anteriores ya sabemos que la forma canónica de Jordan de A *va a ser*

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

Cálculo de la matriz P: *Busquemos la base que corresponde al bloque de orden 2 de* $M(1)$. *Tomamos por ejemplo la base* $\{(1,1,1,0)\}$ *de* $E_1(1)$ *que habíamos calculado, y la completamos a una de* $E_2(1)$, *por ejemplo,* $\{(1,1,1,0),(0,0,1,1)\}$. *Tomamos el vector que no está en* $E_1(1)$, *es decir,* $(0,0,1,1)$, *y le aplicamos* $A - I$, *lo que casualmente da como resultado* $(1,1,1,0)$ *(sabíamos en cualquier caso que el resultado tenía que ser un múltiplo de* $(1,1,1,0)$*). Así pues, debemos tomar la base*

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

para el bloque de $M(1)$.

Para el bloque de Jordan de orden 2 correspondiente a $M(2)$, tomamos la base $\{(1,-1,2,2)\}$ de $E_1(1)$ que habíamos calculado, y la completamos hasta obtener una base de $E_2(2)$, por ejemplo, $\{(1,-1,2,2),(1,2,1,0)\}$. Tomamos el vector que no está en $E_1(2)$, es decir, $(1,2,1,0)$, y le aplicamos $A-2I$, lo que casualmente da como resultado $(1,-1,2,2)$. Así pues, debemos tomar la base

$$\{(1,-1,2,2),(1,2,1,0)\}$$

para el bloque de $M(2)$.

Tenemos por tanto que el operador lineal que tiene matriz asociada A respecto de la canónica, tiene respecto de la base

$$B = \{(1,1,1,0),(0,0,1,1),(1,-1,2,2),(1,2,1,0)\}$$

la matriz J anterior. Así pues, dada la matriz de cambio de base de B a la canónica,

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

tenemos la igualdad $A = PJP^{-1}$.

3.2.6 Aplicación a sistemas lineales discretos

Las matrices de Jordan tienen una aplicación importante en la solución de sistemas lineales discretos, especialmente cuando se trata de la descomposición de matrices y la resolución de ecuaciones en diferencias lineales. Aquí hay un ejemplo de cómo se aplican las matrices de Jordan en este contexto.

Consideremos un sistema lineal discreto que puede ser expresado en forma matricial

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k,$$

donde $\boldsymbol{x}_k$ es el vector de estado en el paso k y A es la matriz de transición del sistema. Supongamos que la matriz A se puede descomponer en su forma canónica de Jordan. Esto significa que existe una matriz invertible P y una matriz de Jordan J tal que:

$$A = PJP^{-1}.$$

La matriz de Jordan J tiene una estructura simple (no como la matriz diagonal) que facilita el análisis y la resolución del sistema.

Usando la descomposición de Jordan, podemos transformar el sistema original de la siguiente manera

$$\boldsymbol{x}_{k+1} = A\boldsymbol{x}_k = PJP^{-1}\boldsymbol{x}_k,$$

luego, si definimos un nuevo vector de estado $\boldsymbol{y}_k = P^{-1}\boldsymbol{x}_k$, el sistema se transforma en el siguiente

$$\boldsymbol{y}_{k+1} = J\boldsymbol{y}_k.$$

La ventaja de trabajar con J es que, debido a su estructura de Jordan, la solución del sistema transformado es más sencilla. Para una matriz de Jordan J, la solución general del sistema discreto es,

$$\boldsymbol{y}_k = J^k \boldsymbol{y}_0$$

donde $\boldsymbol{y}_0$ es el estado inicial en el sistema transformado. Finalmente, para obtener la solución en términos del sistema original, usamos la transformación inversa

$$\boldsymbol{x}_k = P\boldsymbol{y}_k.$$

De la ecuación anterior notamos la importancia del cálculo de J^k. A continuación se presentan algunas ideas.

3.2.7 Cálculo de potencias de matrices usando el Teorema de Jordan

Sea $A \in \mathbb{C}^{n\times n}$ una matriz. Queremos calcular sus potencias A^k de forma sencilla aplicando el Teorema de Jordan. Por dicho teorema, sabemos que existe una matriz de Jordan $J \in \mathbb{C}^{n\times n}$) y una matriz $P \in \mathbb{C}^{n\times n}$ regular tales que $A = PJP^{-1}$ y por tanto $A^k = PJ^kP^{-1}$. Así que todo se reduce a calcular las potencias de matrices de Jordan. De hecho, como dichas matrices son diagonales por bloques de la forma

$$J = \left[\begin{array}{c:c:c} J_1 & & \\ \hdashline & \ddots & \\ \hdashline & & J_r \end{array}\right]$$

donde cada J_i es un cierto bloque Jordan, se tiene que

$$J^k = \left[\begin{array}{c:c:c} J_1^k & & \\ \hdashline & \ddots & \\ \hdashline & & J_r^k \end{array}\right]$$

así que todo se reduce de nuevo a calcular las potencias de los bloques de Jordan. Un bloque de Jordan es de la forma

$$M = \begin{bmatrix} \lambda & 1 & 0 & \cdots & 0 \\ 0 & \lambda & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & & 1 \\ 0 & \cdots & \cdots & & \lambda \end{bmatrix} \in \mathbb{C}^{m\times m}$$

el cual podemos escribir de la forma

$$M = \underbrace{\begin{bmatrix} \lambda & 0 & 0 & \cdots & 0 \\ 0 & \lambda & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & & 0 \\ 0 & \cdots & \cdots & & \lambda \end{bmatrix}}_{\lambda I_m} + \underbrace{\begin{bmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ 0 & \cdots & \cdots & & 1 \\ 0 & \cdots & \cdots & & 0 \end{bmatrix}}_{N_m}.$$

El Teorema del binomio dice que dados dos números reales $a, b \in \mathbb{R}$ se tiene que

$$(a+b)^k = \sum_{j=0}^{k} \binom{k}{j} a^{k-j} b^j.$$

Si uno recupera la demostración y la vuelve a leer pensando que a y b son dos matrices cuadradas que conmutan, entonces descubrirá que la prueba sigue siendo válida. En nuestro caso, como evidentemente la matrices λI_m y N_m conmutan, se tiene que

$$M^k = (\lambda I_m + N_m)^k = \sum_{j=0}^{k} \binom{k}{j} \lambda^{k-j} I_m N_m^j = \sum_{j=0}^{k} \binom{k}{j} \lambda^{k-j} N_m^j.$$

Además, la matriz N_m^j es fácil de calcular: es una matriz con todas las entradas nulas excepto en la diagonal j-ésima por encima de la diagonal principal, cuyas entradas son todas 1. En particular, para $j \geq m$ se tiene que $N_m^j = 0$.

Ejemplo 3.13 *Sea*

$$M = \begin{bmatrix} 2 & 1 \\ 0 & 2 \end{bmatrix} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \end{bmatrix} + \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}.$$

Entonces se tiene que

$$\begin{aligned} M^k &= \sum_{j=0}^{k} \binom{k}{j} 2^{k-j} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^j \\ &= \binom{k}{0} 2^{k-0} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}^0 + \binom{k}{1} 2^{k-1} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \\ &= 2^k \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} + k2^{k-1} \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 2^k & k2^{k-1} \\ 0 & 2^k \end{bmatrix}. \end{aligned}$$

Ejemplo 3.14 *Calcule $\sum_{k=0}^{2000} (-1)^k M^k$ para la matriz*

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

Haciendo las cálculos llegamos a que la forma canónica de Jordan de la matriz anterior es

$$J = \begin{bmatrix} -1 & 0 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix} \text{ con matriz de paso } P = \begin{bmatrix} 2 & 0 & 1 \\ -1 & 1 & 0 \\ 1 & 1 & 0 \end{bmatrix}.$$

Luego, se tiene que

$$J^k = \begin{bmatrix} (-1)^k & 0 & 0 \\ 0 & 1 & k \\ 0 & 0 & 1 \end{bmatrix}$$

y por lo tanto

$$\sum_{k=0}^{2000}(-1)^k J^k = \begin{bmatrix} \sum_{k=0}^{2000}(-1)^k(-1)^k & 0 & 0 \\ 0 & \sum_{k=0}^{2000}(-1)^k & \sum_{k=0}^{2000}(-1)^k k \\ 0 & 0 & \sum_{k=0}^{2000}(-1)^k \end{bmatrix}$$
$$= \begin{bmatrix} 2001 & 0 & 0 \\ 0 & 1 & \sum_{k=0}^{2000}(-1)^k k \\ 0 & 0 & 1 \end{bmatrix}$$
$$= \begin{bmatrix} 2001 & 0 & 0 \\ 0 & 1 & 1000 \\ 0 & 0 & 1 \end{bmatrix}.$$

Finalmente

$$\sum_{k=0}^{2000}(-1)^k M^k = P\left(\sum_{k=0}^{2000}(-1)^k J^k\right)P^{-1} = \begin{bmatrix} 1 & -2000 & 2000 \\ 1000 & 2001 & -2000 \\ 1000 & 0 & 1 \end{bmatrix}.$$

Ejemplo 3.15 *Considere las matrices:*

$$A = \begin{bmatrix} 2 & 1 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{bmatrix}, \quad N = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

Muestre que $N^3 = 0$ y como aplicación, hallar A^n.

Sol: *Tenemos que*

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

Podemos escribir $A = 2I + N$. Además, $(2I)N = 2(IN) = 2(NI) = N(2I)$, es decir $2I$ y N conmutan, luego es aplicable el Teorema del binomio para hallar A^n. Como $N^3 = 0$, se verifica $N^4 = N^5 = \ldots = 0$, por tanto:

$$A^n = (2I+N)^n = \binom{n}{0}(2I)^n + \binom{n}{1}(2I)^{n-1}N + \binom{n}{2}(2I)^{n-2}N^2$$
$$= 2^n I + n2^{n-1}N + \frac{n(n-1)}{2}2^{n-2}N^2$$
$$= \begin{bmatrix} 2^n & 0 & 0 \\ 0 & 2^n & 0 \\ 0 & 0 & 2^n \end{bmatrix} + \begin{bmatrix} 0 & n2^{n-1} & 0 \\ 0 & 0 & n2^{n-1} \\ 0 & 0 & 0 \end{bmatrix} + \begin{bmatrix} 0 & 0 & \frac{n(n-1)2^{n-2}}{2} \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$
$$= \begin{bmatrix} 2^n & n2^n & \frac{n(n-1)2^{n-2}}{2} \\ 0 & 2^n & n2^n \\ 0 & 0 & 2^n \end{bmatrix}.$$

CAPÍTULO 4

Sistemas de ecuaciones en diferencias y estabilidad

En este capítulo presentamos algunos resultados básicos sobre las soluciones del sistema de ecuaciones en diferencias de la forma

$$\boldsymbol{x}_{n+1} = \boldsymbol{f}\left(\boldsymbol{x}_n\right), \quad n = 0, 1, \ldots$$

donde $\boldsymbol{f} : \mathbb{R}^k \to \mathbb{R}^k$ es una función dada.

4.1 Teoría lineal

Como en el caso bidimensional, primero consideramos el sistema homogéneo de la forma

$$\boldsymbol{z}_{n+1} = A\boldsymbol{z}_n \tag{4.1.1}$$

donde A es una matriz constante, real de orden k y $\boldsymbol{z}_n = \left(x_n^1, x_n^2, \ldots, x_n^k\right)^T$. Aquí $\boldsymbol{v}^T$ denota la transpuesta del vector $\boldsymbol{v}$. Todas las nociones básicas de solución, solución general y matriz fundamental son las mismas que en el caso bidimensional. Como en el caso bidimensional, la solución de (4.1.1) tiene la forma

$$\boldsymbol{z}_n = A^n \boldsymbol{z}_0.$$

El cálculo de A^n puede realizarse transformando la matriz A en la forma normal de Jordan y luego calculando la enésima potencia de la matriz transformada. En el caso general de una matriz de orden k, la forma normal de Jordan es más complicada ya que hay más posibilidades para la ubicación de los valores propios relativos al círculo unitario.

4.2 Sistema lineal no homogéneo

En esta sección consideramos el problema de valor inicial

$$\boldsymbol{x}_{n+1} = A\boldsymbol{x}_n + \boldsymbol{b}_n, \quad \boldsymbol{x}_0 = \boldsymbol{d} \quad n = 0, 1, \ldots \tag{4.2.1}$$

donde $k \in \{1, 2, \ldots\}$, $\boldsymbol{d} \in \mathbb{R}^k$, A es una matriz real de orden k, y $\boldsymbol{b}_n \in \mathbb{R}^k$ para $n = 0, 1, \ldots$

El PVI (4.2.1) tiene la única solución dada por

$$\boldsymbol{x}_n = A^n \boldsymbol{d} + \sum_{i=0}^{n} A^{n-i} \boldsymbol{b}_i, \quad n = 0, 1, \ldots . \tag{4.2.2}$$

En particular, si $\boldsymbol{b}_n = \boldsymbol{b}$, un vector constante, entonces obtenemos:

$$\boldsymbol{x}_n = A^n \boldsymbol{d} + \sum_{i=0}^{n} A^{n-i} \boldsymbol{b}, \quad n = 0, 1, \ldots .$$

Estos resultados básicos pueden extenderse a a sistemas lineales más generales de la forma

$$\boldsymbol{z}_{n+1} = A_n \boldsymbol{z}_n, \quad n = 0, 1, \ldots$$

y

$$\boldsymbol{z}_{n+1} = A_n \boldsymbol{z}_n + B_n, \quad \boldsymbol{z}_0 = \boldsymbol{d},$$

donde A_n es una matriz real de orden k para cada $n = 1, 2, \ldots$.

4.3 Estabilidad de sistemas lineales

Consideremos el sistema de ecuaciones en diferencias

$$\boldsymbol{x}_{n+1} = A_n \boldsymbol{x}_n + \boldsymbol{b}_n, \quad n = 0, 1, \ldots \tag{4.3.1}$$

donde $k \in \{1, 2, \ldots\}$, $\boldsymbol{x}_0 \in \mathbb{R}^k$, y donde para cada $n \geq 0$, A_n es una matriz real no singular de orden k y $\boldsymbol{b}_n \in \mathbb{R}^k$.

Definición 4.1 *Sea $\{\boldsymbol{x}_n\}_{n=n_0}^{\infty}$ una solución de (4.3.1). Decimos que $\{\boldsymbol{x}_n\}_{n=n_0}^{\infty}$ es una solución estable de (4.3.1) si dado $\varepsilon > 0$ y $n_1 \geq n_0$, existe $\delta = \delta(\varepsilon, n_1) > 0$ tal que si $\{\tilde{\boldsymbol{x}}_n\}_{n=n_1}^{\infty}$ es una solución de (4.3.1) con $\|\tilde{\boldsymbol{x}}_{n_1} - \boldsymbol{x}_{n_1}\| < \delta$, entonces $\|\tilde{\boldsymbol{x}}_n - \boldsymbol{x}_n\| < \varepsilon$ para todo $n \geq n_1$.*

Note que si $\{\tilde{\boldsymbol{x}}_n\}_{n=n_0}^{\infty}$ y $\{\boldsymbol{x}_n\}_{n=n_0}^{\infty}$ son dos soluciones de (4.3.1), entonces para todo $n \geq n_0$

$$\tilde{\boldsymbol{x}}_{n+1} - \boldsymbol{x}_{n+1} = A_n(\tilde{\boldsymbol{x}}_n - \boldsymbol{x}_n).$$

Así que si establecemos $\boldsymbol{y}_n = \tilde{\boldsymbol{x}}_n - \boldsymbol{x}_n$, entonces

$$\boldsymbol{y}_{n+1} = A_n \boldsymbol{y}_n, \quad n = n_0, n_0 + 1, \ldots .$$

Por lo tanto, es suficiente estudiar la estabilidad de la solución cero del último sistema. Suponiendo que $n_0 = 0$, vamos a estudiar la estabilidad de la solución cero del sistema de ecuaciones en diferencias:

$$\boldsymbol{x}_{n+1} = A_n \boldsymbol{x}_n, \quad n = 0, 1, \ldots \tag{4.3.2}$$

Definición 4.2 *Decimos que el sistema* (4.3.2) *es autónomo si* A_n *no depende de* n*; es decir, existe una matriz real constante de orden* k*,* A*, tal que*

$$\boldsymbol{x}_{n+1} = A\boldsymbol{x}_n, \quad n = 0, 1, \ldots \tag{4.3.3}$$

Definición 4.3 (Estabilidad para un sistema lineal)

(a) *La solución cero de* (4.3.2) *se llama estable si, dado* $\varepsilon > 0$ *y* $n_0 \geq 0$*, existe* $\delta = \delta(\varepsilon, n_0) > 0$ *tal que*

$$\|\boldsymbol{x}_{n_0}\| < \delta \quad \textit{implica} \quad \|\boldsymbol{x}_n\| < \varepsilon \quad \textit{para todo} \quad n \geq n_0.$$

(b) *La solución cero de* (4.3.2) *se llama localmente asíntoticamente estable si es estable, y si para todo* $n_0 \geq 0$*, existe* $K = K(n_0) > 0$ *tal que*

$$\|\boldsymbol{x}_{n_0}\| < K \quad \textit{implica} \quad \lim_{n\to\infty} \|\boldsymbol{x}_n\| = 0.$$

(c) *La solución cero de* (4.3.2) *se llama inestable si no es estable.*

(d) *La solución cero de* (4.3.2) *se llama atractor global si para cada solución* $\{\boldsymbol{x}_n\}_{n=n_0}^{\infty}$ *de* (4.3.2)*, tenemos*

$$\lim_{n\to\infty} \|\boldsymbol{x}_n\| = 0.$$

(e) *La solución cero de* (4.3.2) *se llama globalmente asíntoticamente estable si es estable y es un atractor global.*

El siguiente resultado proporciona una herramienta importante del álgebra lineal en nuestro estudio de la estabilidad.

Teorema 4.1 *Sea* A *una matriz constante de orden* k*. Entonces se cumplen las siguientes afirmaciones:*

1. $\lim_{n\to\infty} \|A^n\| = 0$ *si y sólo si el radio espectral de* A *es menor que 1, es decir,* $\rho(A) < 1$.
2. *Existe una constante* $M > 0$ *tal que* $\|A^n\| \leq M$ *para todo* $n \geq 0$ *si y sólo si cada eigenvalor* λ *de* A *satisface* $|\lambda| \leq 1$*, y cada eigenvalor* λ *de* A *con* $|\lambda| = 1$ *tiene multiplicidad 1.*
3. $\lim_{n\to\infty} \|A^n\| = \infty$ *si y sólo si existe un eigenvalor* λ *de* A *con* $|\lambda| > 1$*, o existe un eigenvalor* λ *de* A *con* $|\lambda| = 1$ *y index mayor que 1.*

Ahora podemos formular el resultado principal de esta sección.

Teorema 4.2 *Consideremos la ecuación en diferencias* (4.3.3)*. Entonces se cumplen las siguientes afirmaciones:*

1. *La solución cero de* (4.3.3) *es estable pero no localmente asintóticamente estable si y sólo si*

cada eigenvalor λ de A satisface $|\lambda| \leq 1$, cada eigenvalor λ de A con $|\lambda| = 1$ tiene index 1, y existe al menos un eigenvalor λ de A con $|\lambda| = 1$.

2. *La solución cero de (4.3.3) es globalmente asintóticamente estable si y sólo si $\rho(A) < 1$.*

3. *La solución cero de (4.3.3) es inestable si y sólo si existe un eigenvalor λ de A con $|\lambda| > 1$, o existe un eigenvalor λ de A con $|\lambda| = 1$ y index mayor que 1.*

4.4 Criterios de Routh-Hurwitz y Schur-Cohn

En esta sección se presentan los criterios de Routh-Hurwitz y Schur-Cohn, que son herramientas efectivas para determinar la estabilidad asintótica de sistemas lineales.

Teorema 4.3 (Criterio de Routh-Hurwitz) *Para números reales $a_0, a_1, \ldots, a_k$, y $a_k > 0$, sea*

$$P(\lambda) := a_k\lambda^k + a_{k-1}\lambda^{k-1} + \cdots + a_1\lambda + a_0. \tag{4.4.1}$$

Consideremos la ecuación polinómica

$$P(\lambda) = 0. \tag{4.4.2}$$

Para cada $n = 1, 2, \ldots, k$, sea Δ_n el menor principal de orden n de la matriz de orden k

$$A = \begin{bmatrix} a_{k-1} & a_k & 0 & \cdots & 0 \\ a_{k-3} & a_{k-2} & a_{k-1} & \cdots & 0 \\ a_{k-5} & a_{k-4} & a_{k-3} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_0 \end{bmatrix}.$$

Las siguientes afirmaciones son verdaderas:

1. *Una condición necesaria y suficiente para que todas las raíces de (4.4.2) tengan una parte real negativa es*
$$\Delta_n > 0 \text{ para } n = 1, 2, \ldots, k.$$

2. *Una condición necesaria y suficiente para la existencia de una raíz de (4.4.2) con una parte real positiva es*
$$\Delta_n < 0, \qquad \text{para algún } n \in \{1, 2, \ldots, k\}.$$

Ejemplo 4.1 *Muestre que una condición necesaria y suficiente para que todas las raíces del polinomio cuadrático con coeficientes reales*

$$\lambda^2 + p\lambda + q = 0$$

tengan una parte real negativa es

$$p > 0 \text{ y } q > 0.$$

Sol: *Note que*

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

y por lo tanto, una condición necesaria y suficiente para que todas las raíces de la ecuación tengan una parte real negativa es

$$p > 0 \text{ y } pq > 0$$

de lo cual se sigue el resultado.

El criterio de Routh-Hurwitz no es inmediatamente aplicable a los sistemas dinámicos discretos y a las ecuaciones en diferencias, pero puede ser utilizado junto con la transformación bilineal

$$z = \frac{\lambda+1}{\lambda-1} \quad \text{o} \quad \lambda = \frac{z+1}{z-1}.$$

Esta transformación mapea la región fuera del círculo unitario en el plano λ a la mitad real positiva de un plano z, y la región dentro del círculo unitario en el plano λ a la mitad real positiva de un plano z. Así tenemos el teorema.

Teorema 4.4 *Sea $P(z)$ el polinomio definido por (4.4.1), y sea $Q(z)$ el polinomio dado por*

$$Q(z) = (z-1)^k P\left(\frac{z+1}{z-1}\right).$$

Considere la ecuación

$$Q(z) = 0. \tag{4.4.3}$$

Las siguientes afirmaciones son verdaderas.

1. *Todas las raíces de (4.4.3) se encuentran en el disco unitario abierto $|\lambda| < 1$ si y sólo si todas las raíces de (4.4.2) se encuentran en la mitad izquierda abierta del plano $\mathrm{Re}(z) < 0$.*

2. *Existe una raíz λ de (4.4.3) con $|\lambda| > 1$ si y sólo si existe una raíz de (4.4.2) que se encuentra en la mitad derecha abierta del plano $\mathrm{Re}(z) > 0$.*

Una prueba alternativa, más directa, para la estabilidad de sistemas dinámicos discretos y ecuaciones en diferencias es proporcionada por el criterio de Schur-Cohn. Esta prueba, al igual que el criterio de Routh-Hurwitz, se expresa en términos de determinantes formados a partir de los coeficientes de la ecuación característica.

Teorema 4.5 (Criterio de Schur-Cohn) *Considere el polinomio*

$$Q(\lambda) = a_k\lambda^k + a_{k-1}\lambda^{k-1} + \cdots + a_1\lambda + a_0$$

V. Osores

donde $a_\ell \in \mathbb{R}$ *para* $0 \leq \ell \leq k$, *y* $a_k > 0$. *Para* $j = 1, 2, \dots, k$, *considere el determinante de orden* $2j$

$$D_j = \begin{vmatrix} a_0 & 0 & \cdots & 0 & | & a_k & a_{k-1} & \cdots & a_{k-j+1} \\ a_1 & a_0 & \cdots & 0 & | & 0 & a_k & \cdots & a_{k-j+2} \\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ a_{j-1} & a_{j-2} & \cdots & a_0 & | & 0 & 0 & \cdots & a_k \\ -- & -- & -- & - & -- & -- & -- & -- & \\ a_k & 0 & \cdots & 0 & | & a_0 & a_1 & \cdots & a_{j-1} \\ a_{k-1} & a_k & \cdots & 0 & | & 0 & a_0 & \cdots & a_{j-2} \\ \vdots & \vdots & \ddots & \vdots & | & \vdots & \vdots & \ddots & \vdots \\ a_{k-j+1} & a_{k-j+2} & \cdots & a_k & | & 0 & 0 & \cdots & a_0 \end{vmatrix} \tag{4.4.4}$$

Una condición necesaria y suficiente para que $Q(\lambda)$ *tenga todas sus raíces dentro del disco unitario es*

$$\begin{matrix} D_j < 0 & \text{para } j \text{ impar} \\ D_j > 0 & \text{para } j \text{ par} \end{matrix}, \quad j = 1, \dots, k.$$

A continuación, utilizamos el criterio de Schur-Cohn para derivar condiciones explícitas necesarias y suficientes para que la solución cero de ecuaciones de diferencias lineales homogéneas de segundo y tercer orden sea asintóticamente estable.

Ejemplo 4.2 *Considere la ecuación de diferencia lineal de segundo orden con coeficientes reales*

$$x_{n+2} + px_{n+1} + qx_n = 0, \quad n = 0, 1, \dots, \quad x_0, x_1 \in \mathbb{R}. \tag{4.4.5}$$

Muestre que una condición necesaria y suficiente para que la solución cero de la ecuación (4.4.5) *sea asintóticamente estable es que*

$$|p| < 1 + q < 2. \tag{4.4.6}$$

Sol: *(No olvidar escribir como un sistema lineal para identificar la matriz* A*). En efecto, los determinantes* (4.4.4) *son*

$$D_1 = \begin{vmatrix} q & 1 \\ 1 & q \end{vmatrix} = q^2 - 1,$$

$$D_2 = \begin{vmatrix} q & 0 & 1 & p \\ p & q & 0 & 1 \\ 1 & 0 & q & p \\ p & 1 & 0 & q \end{vmatrix} = (q-1)^2 \left((q+1)^2 - p\right).$$

Observa que $D_1 < 0$ *y* $D_2 > 0$ *pueden escribirse como*

$$q^2 < 1 \quad y \quad p^2 < (q+1)^2,$$

lo cual es equivalente a la condición (4.4.6). *De manera similar, se puede mostrar que una condición necesaria y suficiente para que todas las raíces de la ecuación cúbica con coeficientes reales*

$$\lambda^3 + a_1\lambda^2 + a_2\lambda + a_3 = 0$$

tengan una parte real negativa es

$$a_1 > 0, \quad a_2 > 0, \quad a_3 > 0, \quad y \quad a_1 a_2 > a_3.$$

Usando técnicas similares al ejemplo 4.2, se puede demostrar que una condición necesaria y suficiente para que todas las raíces de la ecuación cúbica con coeficientes reales

$$\lambda^3 + a_2\lambda^2 + a_1\lambda + a_0 = 0$$

estén en el disco abierto $|\lambda| < 1$ es

$$|a_2 + a_0| < 1 + a_1, \quad |a_2 - 3a_0| < 3 - a_1, \text{ y } a_0^2 + a_1 - a_0 a_2 < 1.$$

Dejamos los detalles de la prueba al lector.

4.5 Sistemas no lineales y estabilidad

En esta sección definimos la estabilidad para el sistema no lineal general de ecuaciones de diferencia de la forma

$$\boldsymbol{x}_{n+1} = \boldsymbol{f}\left(n, \boldsymbol{x}_n\right), \quad n = 0, 1, \ldots \tag{4.5.1}$$

donde $\boldsymbol{x}_n \in \mathbb{R}^k$ y $\boldsymbol{f} : \mathbb{Z} \times \mathbb{R}^k \to \mathbb{R}^k$ es continua en $\boldsymbol{x}$. La ecuación (4.5.1) se dice que es autónoma si la variable $\boldsymbol{n}$ no aparece explícitamente en el lado derecho de esta ecuación, es decir, si (4.5.1) tiene la forma

$$\boldsymbol{x}_{n+1} = \boldsymbol{f}\left(\boldsymbol{x}_n\right), \quad n = 0, 1, \ldots. \tag{4.5.2}$$

Un punto $\bar{\boldsymbol{x}}$ se llama punto de equilibrio de la ecuación (4.5.1) si $\bar{\boldsymbol{x}} = \boldsymbol{f}(n, \bar{\boldsymbol{x}})$ para todo $n \geq N$. Para todos los propósitos de la teoría de estabilidad podemos asumir que el punto de equilibrio es la solución cero. Para ver esto, definimos $\boldsymbol{y}_n = x_n - \bar{\boldsymbol{x}}$. Entonces (4.5.1) toma la forma

$$\boldsymbol{y}_{n+1} = \boldsymbol{f}\left(n, \boldsymbol{y}_n + \bar{\boldsymbol{x}}\right) - \bar{\boldsymbol{x}} = \boldsymbol{g}\left(n, \boldsymbol{y}_n\right),$$

donde la solución cero es el equilibrio correspondiente a la solución de equilibrio $\bar{\boldsymbol{x}}$ de (4.5.1). Sin embargo, consideraremos (4.5.1) ya que la transformación mencionada anteriormente puede ser técnicamente complicada.

Definición 4.4 (Estabilidad para sistemas no lineales)

(a) Una solución de equilibrio de la ecuación (4.5.1) *se llama estable (uniformemente estable) si dado $\varepsilon > 0$ y $n_0 \geq 0$, existe $\delta = \delta\left(\varepsilon, n_0\right) > 0$ (si δ es independiente de n_0) tal que*

$$\|\boldsymbol{x}_{n_0} - \bar{\boldsymbol{x}}\| < \delta \text{ implica } \|\boldsymbol{x}_n - \bar{\boldsymbol{x}}\| < \varepsilon \text{ para todo } n \geq n_0.$$

(b) Una solución de equilibrio de (4.5.1) *se llama exponencialmente estable si existen $\delta > 0, M >$*

$0, \eta > 0$ *tales que*

$$\|\boldsymbol{x}_{n_0} - \bar{\boldsymbol{x}}\| \leq M \|\boldsymbol{x}_0 - \bar{\boldsymbol{x}}\| \eta^{n-n_0}, \qquad \textit{siempre que } \|\boldsymbol{x}_0 - \bar{\boldsymbol{x}}\| < \delta.$$

(c) Una solución de equilibrio de (4.5.1) *se llama inestable si no es estable.*

(d) Una solución de equilibrio de (4.5.1) *se llama atractiva si existe* $K = K(n_0)$ *tal que* $\|\boldsymbol{x}_0 - \bar{\boldsymbol{x}}\| < K$ *implica* $\lim_{n\to\infty} \boldsymbol{x}_n = \bar{\boldsymbol{x}}$*, y uniformemente atractiva si la elección de* K *es independiente de* n_0.

(e) Una solución de equilibrio de (4.5.1) *se llama asintóticamente estable si es estable y atractiva, y uniformemente asintóticamente estable si es uniformemente estable y atractiva.*

(f) Una solución de equilibrio de (4.5.1) *se llama un atrayente global si*

$$\lim_{n\to\infty} \|\boldsymbol{x}_n\| = \bar{\boldsymbol{x}}$$

para cada solución $\{\boldsymbol{x}_n\}_{n=n_0}^{\infty}$ *de* (4.5.1).

(g) La solución de equilibrio de (4.5.1) *se llama globalmente asintóticamente estable si es estable y es un atrayente global.*

Se puede demostrar para sistemas autónomos que la estabilidad (respectivamente, estabilidad asintótica) es equivalente a la estabilidad uniforme (respectivamente, estabilidad asintótica uniforme) del sistema.

4.5.1 Estabilidad linealizada

El método de linealización es probablemente el primer método utilizado en la teoría de estabilidad. Fue originado por Lyapunov, Perron y Poincaré en el estudio de la estabilidad de ecuaciones diferenciales. En este enfoque consideramos (4.5.1) junto con su linealización

$$\boldsymbol{z}_{n+1} = D_{\boldsymbol{f}}(n, \bar{\boldsymbol{x}})\boldsymbol{z}_n = A(n)\boldsymbol{z}_n, \tag{4.5.3}$$

donde $A(n) = D_{\boldsymbol{f}}(n, \bar{\boldsymbol{x}})$ es la matriz jacobiana de $\boldsymbol{f}$ evaluada en una solución de equilibrio $\bar{\boldsymbol{x}}$ de (4.5.1). Así podemos reescribir (4.5.1) como:

$$\boldsymbol{x}_{n+1} = A(n)\boldsymbol{x}_n + \boldsymbol{g}(n, \boldsymbol{x}_n), \tag{4.5.4}$$

donde $\boldsymbol{g}(n, \boldsymbol{y}) = \boldsymbol{f}(n, \boldsymbol{y}) - A(n)\boldsymbol{y}$.

Note que las versiones autónomas de las ecuaciones (4.5.4) y (4.5.3) son, respectivamente,

$$\boldsymbol{y}_{n+1} = A\boldsymbol{y}_n + \boldsymbol{g}(\boldsymbol{y}_n) \tag{4.5.5}$$

y

$$\boldsymbol{z}_{n+1} = D_{\boldsymbol{f}}(\bar{\boldsymbol{x}})\boldsymbol{z}_n = A\boldsymbol{z}_n. \tag{4.5.6}$$

Necesitaremos la siguiente definición.

Definición 4.5 *Decimos que* $G(n, \boldsymbol{y}) = o(\boldsymbol{y})$ *cuando* $\|\boldsymbol{y}\| \to 0$ *si, dado* $\varepsilon > 0$, *existe* $\delta > 0$ *tal que* $\|G(n, \boldsymbol{y})\| \leq \varepsilon \|\boldsymbol{y}\|$ *siempre que* $\|\boldsymbol{y}\| < \delta$ *y* $n \in \mathbb{Z}^+$.

Ahora tenemos nuestros principales resultados de estabilidad, conocidos como los Teoremas de estabilidad linealizada.

Teorema 4.6 (Teorema de estabilidad linealizada) *Supongamos que* $\boldsymbol{g}(n, \boldsymbol{y}) = o(\boldsymbol{y})$ *cuando* $\|\boldsymbol{y}\| \to 0$. *Si el equilibrio cero de la ecuación* (4.5.3) *es uniformemente asintóticamente estable, entonces el equilibrio cero de la ecuación* (4.5.4) *es exponencialmente estable.*

En el caso especial del sistema (4.5.6) tenemos resultados más fuertes:

Teorema 4.7

1. *Si* $\rho(A) < 1$, *entonces el equilibrio cero de* (4.5.6) *es exponencialmente estable.*
2. *Si* $\|D_{\boldsymbol{f}}(0)\| < 1$, *entonces el equilibrio cero de* (4.5.6) *es exponencialmente estable.*

Un resultado parcialmente inverso de esto es el siguiente:

Teorema 4.8

1. *Si* $\rho(A) = 1$, *entonces el equilibrio cero de* (4.5.6) *puede ser estable o inestable.*
2. *Si* $\rho(A) > 1$, *y* $\boldsymbol{g}(\boldsymbol{y}) = o(\boldsymbol{y})$ *cuando* $\|\boldsymbol{y}\| \to 0$, *entonces el equilibrio cero de* (4.5.6) *es inestable.*

V. Osores

APÉNDICE A

Apéndice

A.1 Demostración del Teorema 3.8

La demostración del Teorema 3.8 requiere el siguiente hecho simple sobre bases, cuya demostración dejamos al lector.

Lema A.1 *Si $\{\boldsymbol{v}_1, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ es una base de un espacio vectorial V, también lo es $\{\boldsymbol{v}_1 + s\boldsymbol{v}_2, \boldsymbol{v}_2, \ldots, \boldsymbol{v}_n\}$ para cualquier escalar s.*

Demostración: (Demostración del Teorema 3.8) *Sea A como en el Teorema 3.8, y sea $T = T_A : \mathbb{R}^n \to \mathbb{R}^n$ la transformación de matriz inducida por A. Por conveniencia, llamamos matriz $\lambda - m$-ut a una matriz triangular superior de tamaño $m \times m$ donde cada entrada de la diagonal es igual a λ. Entonces debemos encontrar una base B de $\mathbb{R}^n$ tal que $M_B(T) = \operatorname{diag}(U_1, U_2, \ldots, U_k)$ donde U_i es una matriz $\lambda_i - m_i$-ut para cada i. Procedemos por inducción en n. Si $n = 1$, tomamos $B = \{\boldsymbol{v}\}$ donde $\boldsymbol{v}$ es cualquier vector propio de T.*

Si $n > 1$, sea $\boldsymbol{v}_1$ un vector propio de λ_1 de T, y sea $B_0 = \{\boldsymbol{v}_1, \boldsymbol{w}_1, \ldots, \boldsymbol{w}_{n-1}\}$ cualquier base de $\mathbb{R}^n$ que contenga a $\boldsymbol{v}_1$. Luego (ver Lema 2.4)

$$M_{B_0}(T) = \begin{bmatrix} \lambda_1 & X \\ 0 & A_1 \end{bmatrix}$$

en forma de bloque donde A_1 es de tamaño $(n-1) \times (n-1)$. Además, A y $M_{B0}(T)$ son similares, así que

$$f_A(x) = f_{M_{B_0}(T)}(x) = (x - \lambda_1) f_{A_1}(x)$$

Por lo tanto, $f_{A_1}(x) = (x - \lambda_1)^{m_1 - 1} (x - \lambda_2)^{m_2} \cdots (x - \lambda_k)^{m_k}$ así que (por inducción) sea

$$Q^{-1} A_1 Q = \operatorname{diag}(Z_1, U_2, \ldots, U_k)$$

donde Z_1 es una matriz $\lambda_1 - (m_1 - 1)$-ut y U_i es una matriz $\lambda_i - m_i$-ut para cada $i > 1$.

Si $P = \begin{bmatrix} 1 & 0 \\ 0 & Q \end{bmatrix}$, entonces $P^{-1} M B_0(T) = \begin{bmatrix} \lambda_1 & XQ \\ 0 & Q^{-1} A_1 Q \end{bmatrix} = A'$, digamos. Por lo tanto $A' \sim M_{B_0}(T) \sim A$ entonces por Teorema 3.6(2) hay una base B de $\mathbb{R}^n$ tal que $M_{B_1}(T_A) = A'$, es decir

$M_{B_1}(T) = A'$. Por lo tanto, $M_{B_1}(T)$ toma la forma de bloque

$$M_{B_1}(T) = \begin{bmatrix} \lambda_1 & XQ \\ 0 & \operatorname{diag}(Z_1, U_2, \ldots, U_k) \end{bmatrix} = \left[\begin{array}{cc|ccc} \lambda_1 & X_1 & & Y & \\ 0 & Z_1 & 0 & 0 & 0 \\ \hline & & U_2 & \cdots & 0 \\ & 0 & \vdots & & \vdots \\ & & 0 & \cdots & U_k \end{array}\right] \tag{11.1}$$

Si escribimos $U_1 = \begin{bmatrix} \lambda_1 & X_1 \\ 0 & Z_1 \end{bmatrix}$, la base B_1 cumple nuestras necesidades excepto que la matriz de fila Y puede no ser cero.

Remediamos este defecto de la siguiente manera. Observamos que el primer vector en la base B_1 es un vector propio de λ_1 de T, que continuamos denominando como $\boldsymbol{v}_1$. La idea es agregar múltiplos escalares adecuados de $\boldsymbol{v}_1$ a los otros vectores en B_1. Esto resulta en una nueva base por el Lema A.1, y los múltiplos se pueden elegir de manera que la nueva matriz de T sea igual que (11.1) excepto que $Y = 0$. Sean $\{\boldsymbol{w}_1, \ldots, \boldsymbol{w}_{m_2}\}$ los vectores en B_1 correspondientes a λ_2 (dando lugar a U_2 en (11.1)). Escribimos

$$U_2 = \begin{bmatrix} \lambda_2 & u_{12} & u_{13} & \cdots & u_{1_{m_2}} \\ 0 & \lambda_2 & u_{23} & \cdots & u_{2_{m_2}} \\ 0 & 0 & \lambda_2 & \cdots & u_{3_{m_2}} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & \lambda_2 \end{bmatrix} \quad y \quad Y = \begin{bmatrix} y_1 & y_2 & \cdots & y_{m_2} \end{bmatrix}$$

Reemplazamos primero $\boldsymbol{w}_1$ por $\boldsymbol{w}_1' = \boldsymbol{w}_1 + s\boldsymbol{v}_1$ donde s debe ser determinado. Luego (11.1) da

$$\begin{aligned} T(\boldsymbol{w}_1') &= T(\boldsymbol{w}_1) + sT(\boldsymbol{v}_1) \\ &= (y_1\boldsymbol{v}_1 + \lambda_2\boldsymbol{w}_1) + s\lambda_1\boldsymbol{v}_1 \\ &= y_1\boldsymbol{v}_1 + \lambda_2(\boldsymbol{w}_1' - s\boldsymbol{v}_1) + s\lambda_1\boldsymbol{v}_1 \\ &= \lambda_2\boldsymbol{w}_1' + [(y_1 - s(\lambda_2 - \lambda_1)]\boldsymbol{v}_1 \end{aligned}$$

Como $\lambda_2 \neq \lambda_1$ podemos elegir s tal que $T(\boldsymbol{w}_1') = \lambda_2\boldsymbol{w}_1'$. Similarmente, sea $\boldsymbol{w}_2' = \boldsymbol{w}_2 + t\boldsymbol{v}_1$ donde t debe ser elegido. Entonces, como antes,

$$\begin{aligned} T(\boldsymbol{w}_2') &= T(\boldsymbol{w}_2) + tT(\boldsymbol{v}_1) \\ &= (y_2\boldsymbol{v}_1 + u_{12}\boldsymbol{w}_1 + \lambda_2\boldsymbol{w}_2) + t\lambda_1\boldsymbol{v}_1 \\ &= u_{12}\boldsymbol{w}_1' + \lambda_2\boldsymbol{w}_2' + [(y_2 - u_{12}s) - t(\lambda_2 - \lambda_1)]\boldsymbol{v}_1 \end{aligned}$$

Nuevamente, t puede ser elegido para que $T(\boldsymbol{w}_2') = u_{12}\boldsymbol{w}_1' + \lambda_2\boldsymbol{w}_2'$. Continuamos de esta manera para eliminar $y_1, \ldots, y_{m_2}$. Este procedimiento también funciona para $\lambda_3, \lambda_4, \ldots$ y por lo tanto produce una nueva base B tal que $M_B(T)$ es como en (11.1) pero con $Y = 0$.

A.2 Prueba del Lema 3.4

Lema A.2 *Sea $T : V \to V$ un operador lineal donde $\dim V = n$, y supongamos que T es nilpotente; es decir, $T^m = 0$ para algún $m \geq 1$. Entonces V tiene una base B tal que*

$$M_B(T) = \operatorname{diag}(J_1, J_2, \ldots, J_k)$$

donde cada $J_i = J_{n_i}(0)$ es un bloque de Jordan correspondiente a $\lambda = 0$.

Demostración: *La demostración procede por inducción en n. Si $n = 1$, entonces T es un operador escalar, y por lo tanto $T = 0$ y el lema se cumple. Si $n \geq 1$, podemos suponer que $T \neq 0$, entonces $m \geq 1$ y podemos suponer que m está elegido de manera que $T^m = 0$, pero $T^{m-1} \neq 0$. Supongamos que $T^{m-1}\boldsymbol{u} \neq \boldsymbol{0}$ para algún $\boldsymbol{u}$ en V.*

Prop. $\{\boldsymbol{u}, T\boldsymbol{u}, T^2\boldsymbol{u}, \ldots, T^{m-1}\boldsymbol{u}\}$ es independiente.

Dem. Supongamos $a_0\boldsymbol{u} + a_1T\boldsymbol{u} + a_2T^2\boldsymbol{u} + \cdots + a_{m-1}T^{m-1}\boldsymbol{u} = \boldsymbol{0}$ donde cada a_i está en $\mathbb{R}$. Dado que $T^m = 0$, aplicando T^{m-1} da $\boldsymbol{0} = T^{m-1}\boldsymbol{0} = a_0T^{m-1}\boldsymbol{u}$, entonces $a_0 = 0$. Por lo tanto, $a_1T\boldsymbol{u} + a_2T^2\boldsymbol{u} + \cdots + a_{m-1}T^{m-1}\boldsymbol{u} = \boldsymbol{0}$ y aplicando T^{m-2} da $a_1 = 0$ de la misma manera. Continúa de esta manera para obtener $a_i = 0$ para cada i. Esto prueba la propiedad.

Ahora define $P = \operatorname{span}\{\boldsymbol{u}, T\boldsymbol{u}, T^2\boldsymbol{u}, \ldots, T^{m-1}\boldsymbol{u}\}$. Entonces P es un subespacio T- invariante (porque $T^m = 0$), y $T : P \to P$ es nilpotente con matriz $M_B(T) = J_m(0)$ donde $B = \{\boldsymbol{u}, T\boldsymbol{u}, T^2\boldsymbol{u}, \ldots, T^{m-1}\boldsymbol{u}\}$. Por lo tanto, estamos hechos, por inducción, si $V = P \oplus Q$ donde Q es T-invariante (entonces $\dim Q = n - \dim P < n$ porque $P \neq 0$ y $T : Q \to Q$ es nilpotente). Con esto en mente, elige un subespacio T- invariante Q de dimensión maximal tal que $P \cap Q = \{\boldsymbol{0}\}$. Suponemos que $V \neq P \oplus Q$ y buscamos una contradicción.

Elige $\boldsymbol{x} \in V$ tal que $\boldsymbol{x} \notin P \oplus Q$. Entonces $T^m\boldsymbol{x} = \boldsymbol{0} \in P \oplus Q$ mientras que $T^0\boldsymbol{x} = \boldsymbol{x} \notin P \oplus Q$. Por lo tanto, existe $k, 1 \leq k \leq m$, tal que $T^k\boldsymbol{x} \in P \oplus Q$ pero $T^{k-1}\boldsymbol{x} \notin P \oplus Q$. Escribe $\boldsymbol{v} = T^{k-1}\boldsymbol{x}$, de modo que

$$\boldsymbol{v} \notin P \oplus Q \quad y \quad T\boldsymbol{v} \in P \oplus Q$$

Sea $T\boldsymbol{v} = \boldsymbol{p} + \boldsymbol{q}$ con $\boldsymbol{p}$ en P y $\boldsymbol{q}$ en Q. Entonces $\boldsymbol{0} = T^{m-1}(T\boldsymbol{v}) = T^{m-1}\boldsymbol{p} + T^{m-1}\boldsymbol{q}$, así que, dado que P y Q son T-invariantes, $T^{m-1}\boldsymbol{p} = -T^{m-1}\boldsymbol{q} \in P \cap Q = \{\boldsymbol{0}\}$. Por lo tanto,

$$T^{m-1}\boldsymbol{p} = \boldsymbol{0}$$

Dado que $\boldsymbol{p} \in P$, tenemos $\boldsymbol{p} = a_0\boldsymbol{u} + a_1T\boldsymbol{u} + a_2T^2\boldsymbol{u} + \cdots + a_{m-1}T^{m-1}\boldsymbol{u}$ para $a_i \in \mathbb{R}$. Dado que $T^m = 0$, aplicando T^{m-1} da $\boldsymbol{0} = T^{m-1}\boldsymbol{p} = a_0T^{m-1}\boldsymbol{u}$, entonces $a_0 = 0$. Así que $\boldsymbol{p} = T(\boldsymbol{p}_1)$ donde

$$\boldsymbol{p}_1 = a_1\boldsymbol{u} + a_2T\boldsymbol{u} + \cdots + a_{m-1}T^{m-2}\boldsymbol{u} \in P$$

Si escribimos $\boldsymbol{v}_1 = \boldsymbol{v} - \boldsymbol{p}_1$, tenemos

$$T(\boldsymbol{v}_1) = T(\boldsymbol{v} - \boldsymbol{p}_1) = T\boldsymbol{v} - \boldsymbol{p} = \boldsymbol{q} \in Q$$

Dado que $T(Q) \subseteq Q$, se sigue que $T(Q + \mathbb{R}\boldsymbol{v}_1) \subseteq Q \subseteq Q + \mathbb{R}\boldsymbol{v}_1$. Además, $\boldsymbol{v}_1 \notin Q$ (de lo contrario, $\boldsymbol{v} = \boldsymbol{v}_1 + \boldsymbol{p}_1 \in P \oplus Q$, una contradicción). Por lo tanto, $Q \subset Q + \mathbb{R}\boldsymbol{v}_1$, así que, por la maximalidad de

Q, tenemos $(Q + \mathbb{R}\boldsymbol{v}_1) \cap P \neq \{\mathbf{0}\}$, digamos

$$\mathbf{0} \neq \boldsymbol{p}_2 = \boldsymbol{q}_1 + a\boldsymbol{v}_1 \quad \text{donde} \quad \boldsymbol{p}_2 \in P, \quad \boldsymbol{q}_1 \in Q, \quad y \quad a \in \mathbb{R}$$

Así que $a\boldsymbol{v}_1 = \boldsymbol{p}_2 - \boldsymbol{q}_1 \in P \oplus Q$. Pero dado que $\boldsymbol{v}_1 = \boldsymbol{v} - \boldsymbol{p}_1$, tenemos

$$a\boldsymbol{v} = a\boldsymbol{v}_1 + a\boldsymbol{p}_1 \in (P \oplus Q) + P = P \oplus Q$$

Dado que $\boldsymbol{v} \notin P \oplus Q$, esto implica que $a = 0$. Pero entonces $\boldsymbol{p}_2 = \boldsymbol{q}_1 \in P \cap Q = \{\mathbf{0}\}$, una contradicción. Esto completa la demostración.