Teresita Noriega Sánchez
Héctor de Arazoza Rodríguez

ÁLGEBRA

TOMO I

El primer enfoque es más algebraico y trabaja fundamentalmente con las matrices; el segundo, se dedica más al estudio de las aplicaciones lineales nilpotentes y los subespacios asociados a estas aplicaciones.

El programa vigente, en el Plan B, establece realizar el estudio a partir de las λ-matrices, tal como se hace en el capítulo 4. Hay varios argumentos que justifican este enfoque, uno de los cuales es el tiempo necesario para poder llegar al teorema de Jordán. En nuestra opinión el número de clases que se requiere para llegar al teorema de Jordán por cualquiera de las dos vías es aproximadamente el mismo. Cada uno de los enfoques puede ofrecer ventajas sobre el otro.

Desde el punto de vista del curso de Álgebra lineal, a partir de las λ-matrices no solo se obtiene el teorema de Jordán, sino también un método para saber si dos matrices de orden n son semejantes o no, sin tener que calcular los valores propios ni la forma de Jordán. Sin embargo, el método tiene un inconveniente: la forma de hallar la base en la cual se obtiene la representación de Jordán es bastante complicada. Por otra parte, el método de los endomorfismos nilpotentes no solo demuestra la existencia de la matriz de Jordán, sino que lo hace construyendo una base en la cual se obtiene dicha representación (prácticamente la construcción de esta base es la que se emplea al final del capítulo 4). Además, desde este punto de vista, el polinomio minimal (y otros conceptos) aparecen en forma natural. Sin embargo, el método de los endomorfismos nilpotentes no resuelve el problema de la equivalencia de dos matrices en forma general, si no es mediante la forma de Jordán.

Para el desarrollo posterior de un matemático, los dos métodos son importantes; el primero porque se puede generalizar a la descomposición de grupos; y el segundo por su enfoque funcional, que tiene sus homólogos en el teorema de la alternativa de Fredholm, y otros problemas de ecuaciones diferenciales e integrales. Es por eso que proponemos los dos métodos y consideramos que el profesor puede elegir uno de ellos de acuerdo con su interés y la perspectiva de desarrollo de sus alumnos, pero entendemos que el estudiante debe en algún momento leer y estudiar el otro método para así completar su formación.

Agradecemos la colaboración de los Licenciados Mayra Solana y Roberto Nuñez, quienes revisaron el manuscrito. También agradecemos el trabajo de los compañeros Tania Caldevilla y Mariano Blanco en la mecanografía del manuscrito.

Teresita Noriega Sánchez
Héctor de Arazoza Rodríguez

La Habana, setiembre de 1983

$\dim(V+W) = \dim V + \dim W - \dim(V \cap W)$

Indice

1 Espacios vectoriales

Introducción

La estructura de espacio vectorial desempeña una función muy importante en la matemática. Este concepto, que es el fundamental en la rama del Álgebra denominada Álgebra lineal, sirve de base a problemas de la más diversa índole; por tanto, es necesario dominar los problemas y los métodos del Álgebra lineal.

Para la comprensión de los temas que vamos a tratar, la base fundamental es el conocimiento de la teoría de los sistemas de ecuaciones lineales; es decir, es indispensable que el estudiante sepa resolver sistemas de ecuaciones lineales, y por ello recomendamos que revisen los problemas relativos a estos sistemas, estudiados en el primer curso de Álgebra y que se encuentran en el libro de A. Kurosh *Álgebra Superior*.

Realmente la estructura de espacio vectorial no es nueva para el estudiante, pues ya ha trabajado con ella, sin denominarla espacio vectorial, en algunos casos particulares. En el epígrafe que sigue analizaremos estos casos y así obtendremos una idea de qué vamos a definir como espacio vectorial.

1.1 Algunas estructuras conocidas

Vectores de $\mathbb{R}^3$

En los cursos de Geometría analítica y en los de Mecánica se emplea el conjunto de los vectores de $\mathbb{R}^3$ como el conjunto de segmentos orientados con punto inicial en el origen de coordenadas. Cada uno de estos vectores se identifican mediante las coordenadas cartesianas del punto final del vector. Así, un elemento del conjunto, el vector $\vec{v}$, se identifica con tres números (v_1, v_2, v_3) que son las coordenadas de su punto final. En este conjunto se realizan diversas operaciones, pero para nuestros objetivos son de interés la *suma de vectores* y el *producto de un vector por un escalar*. Las definiciones de estas dos operaciones son muy simples:

Si $\vec{v}$, $\vec{w}$ son los vectores y λ un escalar real, se definen:

$$\vec{v}+\vec{w}=(v_1+w_1,\ v_2+w_2,\ v_3+w_3)$$
$$\lambda\vec{v}=(\lambda v_1,\ \lambda v_2,\ \lambda v_3).$$

Como vemos, las operaciones se realizan componente a componente. Geométricamente, estas operaciones tienen una interpretación importante, que es la siguiente:

El vector suma $\vec{v}+\vec{w}$ se obtiene por la llamada "regla del paralelogramo"; el vector $\vec{v}+\vec{w}$ se encuentra en el plano determinado por los vectores $\vec{v}$ y $\vec{w}$.
Si $\lambda>0$, el vector $\lambda\vec{v}$ es un vector que se encuentra en la misma dirección del vector $\vec{v}$, y su magnitud es λ veces la magnitud de $\vec{v}$. Si $\lambda<0$, entonces $\lambda\vec{v}$ es un vector en la dirección opuesta a $\vec{v}$ y su magnitud es $|\lambda|$ veces la magnitud de $\vec{v}$.
Estas dos operaciones: suma de vectores y producto de un vector por un escalar, satisfacen las propiedades siguientes:

(1) La suma de vectores es conmutativa:

$\vec{v}+\vec{w}=\vec{w}+\vec{v}$ para todo $\vec{v},\vec{w}\in \mathbf{R}^3$.

(2) La suma de vectores es asociativa:

$\vec{u}+(\vec{v}+\vec{w})=(\vec{u}+\vec{v})+\vec{w}$ para todo $\vec{u}$, $\vec{v}$, $\vec{w}\in \mathbf{R}^3$.

(3) Existe un vector nulo $\vec{0}=(0,0,0)$, que cumple:

$\vec{v}+\vec{0}=\vec{v}$ para todo $\vec{v}\in \mathbf{R}^3$.

(4) Para todo vector $\vec{v}$ de $\mathbf{R}^3$ existe otro vector, $-\vec{v}$, que cumple:

$\vec{v}+(-\vec{v})=\vec{0}$.

(5) La suma de vectores y el producto por un escalar cumplen una propiedad distributiva:

$\lambda(\vec{v}+\vec{w})=\lambda\vec{v}+\lambda\vec{w}$ para todo $\vec{v},\vec{w}\in \mathbf{R}^3$, $\lambda\in \mathbf{R}$.

(6) La suma de escalares y el producto de un vector por un escalar cumplen una propiedad distributiva:

$(\alpha+\beta)\,\vec{v}=\alpha\vec{v}+\beta\vec{v}$ para todo $\vec{v}\in \mathbf{R}^3$, $\alpha,\beta\in \mathbf{R}$.

(7) Hay una propiedad asociativa entre el producto de un vector por un escalar y el producto de los escalares:

$\alpha(\beta)\ \vec{v}=\alpha\ (\beta\ \vec{v})$ para todo $\vec{v}\in \mathbf{R}^3$, $\alpha,\beta\in \mathbf{R}$.

(8) La unidad de los escalares cumple:

$1\cdot\vec{v}=\vec{v}$ para todo $\vec{v}\in \mathbf{R}^3$.

Matrices

Vamos a considerar el conjunto de las matrices de orden $m\times n$ con coeficientes complejos (podemos tomar también los coeficientes reales). Este conjunto se denota $M_{m,n}$ ($\mathbb{C}$). En este caso también están definidas diversas operaciones, pero vamos a fijar nuestra atención solamente en dos: la *suma de matrices* y el *producto de una matriz por un escalar*.

Si tenemos dos matrices A, $B \in M_{m,n}(\mathbb{C})$, se define la suma $A+B$ como la matriz $m \times n$ cuyo elemento (i, j), es decir, el elemento que está en la fila i y la columna j, se forma sumando el elemento (i, j) de A y el elemento (i, j) de B.

Si $A \in M_{m,n}$ ($\mathbb{C}$) y λ es un número complejo, la matriz λA se forma multiplicando cada uno de los elementos de la matriz A por el número complejo λ.

En definitiva, también en el conjunto de las matrices las operaciones se realizan componente a componente.

Las operaciones que acabamos de definir tienen las propiedades siguientes:

(1) La suma de matrices es conmutativa:

$A+B=B+A$ para toda $A, B \in M_{m,n}$ ($\mathbb{C}$).

(2) La suma de matrices es asociativa:

$A+(B+C)=(A+B)+C$ para toda $A,B,C \in M_{m,n}$ ($\mathbb{C}$).

(3) Existe una matriz nula, 0, aquella en que todos los elementos son nulos, que cumple:

$A+0=A$ para toda $A \in M_{m,n}$ ($\mathbb{C}$).

(4) Para toda matriz A de $M_{m,n}$ ($\mathbb{C}$) existe otra matriz $-A$, que se forma cambiando el signo a todos los elementos de A, que cumple:

$A+(-A)=0$.

(5) La suma de matrices y el producto de una matriz por un escalar cumplen una propiedad distributiva:

$\alpha(A+B)=\alpha A+\alpha B$ para toda $A,B \in M_{m,n}$ ($\mathbb{C}$), $\alpha \in \mathbb{C}$.

(6) La suma de escalares y el producto de una matriz por un escalar cumplen una propiedad distributiva:

$(\alpha+\beta)A=\alpha A+\beta A$ para toda $A \in M_{m,n}$ ($\mathbb{C}$); $\alpha, \beta \in \mathbb{C}$.

(7) Hay una propiedad asociativa entre el producto de una matriz por un escalar y el producto de dos escalares.

$(\alpha\beta)A=\alpha(\beta A)$ para toda $A \in M_{m,n}$ ($\mathbb{C}$); $\alpha, \beta \in \mathbb{C}$.

(8) La unidad de los escalares cumple:

$1 \cdot A=A$ para toda $A \in M_{m,n}$ ($\mathbb{C}$).

El conjunto solución de un sistema homogéneo de ecuaciones lineales

Vamos a considerar en $\mathbb{C}$, la matriz A de los coeficientes de un sistema homogéneo de m ecuaciones lineales con n incógnitas. En este caso, como es conocido, el sistema se puede escribir

$$AX=0,$$

donde X es una matriz $n\times 1$, conocida como *matriz de las incógnitas*.

Llamaremos S al conjunto de todas las soluciones de este sistema. Los elementos de S son matrices $n\times 1$ que satisfacen la ecuación, y como son matrices del mismo orden, se pueden sumar y multiplicar por un escalar; además, podemos comprobar que los resultados de estas operaciones son también soluciones de la ecuación.

Sean X, $Y\in S$. Comprobemos que $X+Y\in S$. Para que esto se cumpla debemos probar que $A(X+Y)=0$.

$$A(X+Y)=AX+AY=0+0=0.$$

Por tanto, $X+Y\in S$.
Sea $X\in S$, $a\in\mathbb{C}$. Comprobemos que $aX\in S$.

$$A(aX)=a(AX)=a(0)=0.$$

Por tanto, $aX\in S$.

Como resultado tenemos que en el conjunto S existen dos operaciones: la suma de dos soluciones de la ecuación y el producto de una solución por un escalar. Estas operaciones cumplen las propiedades siguientes:

(1) La suma en S es conmutativa:

$$X+Y=Y+X \text{ para toda } X,Y\in S.$$

(2) La suma en S es asociativa:

$$X+(Y+Z)=(X+Y)+Z \text{ para toda } X,Y,Z\in S.$$

(3) La solución trivial, que es un elemento de S, cumple:

$$X+0=X \text{ para toda } X\in S.$$

(4) Para toda solución X, existe otra solución, $-X$, que cumple:

$$X+(-X)=0.$$

(5) La suma en S y el producto de una solución por un escalar cumplen una propiedad distributiva:

$$a(X+Y)=aX+aY \text{ para toda } X,Y\in S;\ a\in\mathbb{C}.$$

(6) La suma de escalares y el producto de una solución por un escalar, cumplen una propiedad distributiva:

$$(a+\beta)X=aX+\beta X \text{ para toda } X\in S,\ a,\beta\in\mathbb{C}.$$

(7) Hay una propiedad asociativa entre el producto de una solución por un escalar y el producto de los escalares:

$(\alpha\beta)X=\alpha(\beta X)$ para toda $X\in S$; $\alpha, \beta\in \mathbb{C}$.

(8) La unidad de los escalares cumple:

$1 \cdot X=X$ para toda $X\in S$.

Para las propiedades (1), (2), (5), (6), (7) y (8) no es necesario verificar su cumplimiento, pues en definitiva la suma en S y el producto por un escalar en S son las mismas operaciones que definimos para las matrices y cumplen estas propiedades. La propiedad (3) es evidente, y la (4) es válida pues si X es una solución, $-X$ es también una solución.

Como vemos, el conjunto de los vectores de $\mathbb{R}^3$, el de las matrices y el de las soluciones de un sistema homogéneo de ecuaciones lineales, satisfacen propiedades que son comunes entre ellos; estas propiedades surgen de la existencia de dos operaciones: una suma y un producto por un escalar.

El conjunto de los escalares no es el mismo en los tres casos, pero las propiedades algebraicas que cumplen los escalares sí son las mismas. En este texto, el conjunto de los escalares, que denotaremos por K, será siempre uno de los cuerpos o campos numéricos siguientes:

$\mathbb{Q}$ conjunto de los números racionales

$\mathbb{R}$ conjunto de los números reales

$\mathbb{C}$ conjunto de los números complejos.

Si en cada uno de los conjuntos analizados sustituimos el nombre del conjunto en particular por un mismo nombre, podemos enunciar la definición de espacio vectorial. El estudiante pudiera preguntarse para qué se necesita esta definición. Veamos.

Al agrupar conjuntos de diversa índole pero que tienen estructuras algebraicas similares, es decir, en las cuales hay operaciones definidas que satisfacen propiedades similares, buscamos estudiar las propiedades comunes, que dependen no de cómo se definen las operaciones, sino de las propiedades que estas cumplen. Así podemos plantear problemas del tipo siguiente:

En $\mathbb{R}^3$ todo vector se puede escribir a partir de tres vectores que denotamos $\vec{i}$, $\vec{j}$ y $\vec{k}$. Esta propiedad de poder expresar los elementos de un conjunto a partir de un número finito de ellos, ¿es una propiedad particular de $\mathbb{R}^3$ o es una propiedad general de la estructura? A este problema y a otros similares daremos respuesta en este capítulo.

1.2 Definición de espacio vectorial sobre un cuerpo K

Antes de dar la definición de espacio vectorial debemos aclarar qué se entiende por *suma* y por *producto por un escalar*.

Por *suma* en un conjunto E se entiende una relación que a dos elementos (u, v) del conjunto E les hace corresponder un tercer elemento, $u+v$ de E, el cual se denomina *suma* de u y v.

Sea K el conjunto de los escalares ($K=\mathbb{Q}$ o $\mathbb{R}$ o $\mathbb{C}$). Por *producto* de un elemento v de un conjunto E por un escalar α de K se entiende una relación que a cada par (α, v), formado por un escalar y un elemento de E, le hace corresponder otro elemento, αv de E, el cual se denomina *producto* de α y v.

DEFINICIÓN 1.1

Sean K un conjunto de escalares y E un conjunto no vacío. Se dice que E es un *espacio vectorial sobre* K o un K-espacio, si en E existen dos operaciones: una suma y un producto por un escalar, que satisfacen las propiedades siguientes:

(1) La suma en E es conmutativa:
$u+v=v+u$ para todo $u, v \in E$.

(2) La suma en E es asociativa:
$u+(v+w)=(u+v)+w$ para todo $u, v, w \in E$.

(3) Existe en E un elemento único, que se denota por $0 \in E$ y se denomina *cero* del espacio, que cumple:
$u+0=u$ para todo $u \in E$

(4) Para todo elemento $u \in E$ existe un único elemento $(-u) \in E$ denominado *opuesto* de u, que cumple:
$u+(-u)=0$.

(5) La suma de E y el producto por un escalar cumplen la propiedad distributiva:
$\alpha(u+v)=\alpha u+\alpha v$ para todo $u, v \in E$, $\alpha \in K$.

(6) La suma en el conjunto de los escalares y el producto por un escalar cumplen la propiedad distributiva:
$(\alpha+\beta) u=\alpha u+\beta u$ para todo $u \in E$; $\alpha, \beta \in K$

(7) El producto por un escalar y el producto en el conjunto de los escalares, cumplen la propiedad asociativa:
$(\alpha\beta) u=\alpha(\beta u)$ para todo $u \in E$; $\alpha, \beta \in K$.

(8) La unidad del conjunto de los escalares cumple:
$1 \cdot u=u$ para todo $u \in E$.

Las propiedades 1 a 8 de la definición se denominan también los axiomas de espacio vectorial.

En rigor, el espacio vectorial es una terna: $(E, +, \cdot)$ formada por el conjunto E y las operaciones suma y producto por un escalar que satisfacen los axiomas. Usualmente se dice que E es el espacio vectorial pues se suponen incluidas en E las dos operaciones.

A los elementos de un espacio vectorial se les llama *vectores*.

Ejemplos

1. El conjunto de los vectores de $\mathbb{R}^3$ con la suma de vectores y el producto de un vector por un escalar, es un espacio vectorial.
2. El conjunto $M_{m,n}(\mathbb{C})$ de las matrices $m \times n$ con coeficientes complejos, con la suma de matrices y el producto de una matriz por un escalar, es un espacio vectorial sobre $\mathbb{C}$.
3. Sean A una matriz $m \times n$ con coeficientes complejos y $AX=0$ el sistema homogéneo de ecuaciones lineales que ella determina. El conjunto de las soluciones de este sistema es un espacio vectorial sobre $\mathbb{C}$.
4. $\mathbb{R}$ con la suma y el producto usuales, es un $\mathbb{R}$-espacio.
5. $\mathbb{C}$ con la suma y el producto de números complejos, es un $\mathbb{C}$-espacio.
6. Sea $K[x]$ el conjunto de los polinomios en una indeterminada con coeficientes en K ($K=\mathbb{Q}$, $\mathbb{R}$, $\mathbb{C}$). Entonces, con la suma de polinomios y el producto de un polinomio por un escalar, $K[x]$ es un espacio vectorial sobre K. Comprobemos que $K[x]$ satisface cada una de las propiedades dadas en la definición 1.1.

   Las propiedades (1), (2), (5), (6), (7) y (8) son propiedades conocidas de la suma y el producto de polinomios. El elemento que se plantea en la propiedad (3) es el polinomio nulo, y como para cada polinomio $p(x)$ se obtiene el polinomio $-p(x)$ cambiando el signo a todos sus coeficientes, también se satisface la propiedad (4). Por tanto, $K[x]$ es un K-espacio.
7. A partir del ejemplo 6 podemos construir otro ejemplo.

   Sea $K_n[x]$ el conjunto de los polinomios con coeficientes en K de grado menor que n ($n \in \mathbb{N}$) y que, además, contiene al polinomio nulo. Este conjunto, con la suma de polinomios y el producto por un escalar, es un espacio vectorial sobre K. Obsérvese que $K_n[x] \subset K[x]$ para todo valor de $n \in \mathbb{N}$.
8. Sea K un conjunto de escalares.

   Consideremos el conjunto $K^n = \{ (x_1, x_2, \ldots, x_n) / x_i \in K\}$ y definamos en K^n una suma y un producto por un escalar.

   Sean $x=(x_1,\ldots,x_n)$, $y=(y_1,\ldots,y_n) \in K^n$, $\lambda \in K$. Definimos:

   $x+y=(x_1+y_1, x_2+y_2, \ldots, x_n+y_n) \in K^n$
   $\lambda x=(\lambda x_1, \lambda x_2, \ldots, \lambda x_n) \in K^n$

   Comprobemos los axiomas de la definición.

   (1) Conmutatividad

   $$\begin{aligned} x+y &= (x_1+y_1, x_2+y_2, \ldots, x_n+y_n) \\ &= (y_1+x_1, y_2+x_2, \ldots, y_n+x_n) \\ &= y+x \end{aligned}$$

   (2) Asociatividad

   Sea $z=(z_1, z_2, \ldots, z_n)$.

   $$\begin{aligned} (x+y)+z &= ((x_1+y_1)+z_1,\ (x_2+y_2)+z_2, \ldots, (x_n+y_n)+z_n) \\ &= (x_1+(y_1+z_1),\ x_2+(y_2+z_2), \ldots, x_n+(y_n+z_n)) \\ &= x+(y+z) \end{aligned}$$

brexd

(3) Sea $0=(0,0,\ldots,0)$.

$$\begin{aligned} x+0 &= (x_1+0, x_2+0, \ldots, x_n+0) \\ &= (x_1, x_2, \ldots, x_n) \\ &= x \end{aligned}$$

(4) Para $x=(x_1,x_2,\ldots,x_n)$ tomemos $-x=(-x_1,-x_2,\ldots,-x_n)$.

$$\begin{aligned} x+(-x) &= (x_1+(-x_1), x_2+(-x_2), \ldots, x_n+(-x_n)) \\ &= (0,0,\ldots,0) \\ &= 0 \end{aligned}$$

(5) $x, y \in K^n$; $\alpha \in K$

$$\begin{aligned} \alpha(x+y) &= \alpha(x_1+y_1, x_2+y_2, \ldots, x_n+y_n) \\ &= (\alpha(x_1+y_1), \alpha(x_2+y_2), \ldots, \alpha(x_n+y_n)) \\ &= (\alpha x_1+\alpha y_1, \alpha x_2+\alpha y_2, \ldots, \alpha x_n+\alpha y_n) \\ &= \alpha x+\alpha y. \end{aligned}$$

(6) Para $x \in K^n$; $\alpha, \beta \in K$;

$$\begin{aligned} (\alpha+\beta)x &= ((\alpha+\beta)x_1, (\alpha+\beta)x_2, \ldots, (\alpha+\beta)^x{}_n) \\ &= (\alpha x_1+\beta x_1, \alpha x_2+\beta x_2, \ldots, \alpha x_n+\beta x_n) \\ &= \alpha x+\beta x \end{aligned}$$

(7) Para $x \in K^n$; $\alpha, \beta \in K$:

$$\begin{aligned} (\alpha\beta)x &= ((\alpha\beta)x_1, (\alpha\beta)x_2, \ldots, (\alpha\beta)x_n) \\ &= (\alpha(\beta x_1), \alpha(\beta x_2), \ldots, \alpha(\beta x_n)) \\ &= \alpha(\beta x_1, \beta x_2, \ldots, \beta x_n) \\ &= \alpha(\beta x) \end{aligned}$$

(8) Para todo $x \in K^n$ y tomando $1 \in K$:

$$1x=(1x_1, 1x_2, \ldots, 1x_n)=(x_1, x_2, \ldots, x_n)=x$$

Por tanto, K^n es un espacio vectorial sobre K.

Más adelante veremos que este ejemplo es fundamental.

Observemos que K^n puede ser interpretado como el conjunto de las matrices $n\times 1$ (o $1\times n$) y que las operaciones que acabamos de definir son las que corresponden a las matrices. Así, tenemos los espacios: $\mathbb{Q}^n$ que es un $\mathbb{Q}$-espacio, $\mathbb{R}^n$ que es un $\mathbb{R}$-espacio y $\mathbb{C}^n$ que es un $\mathbb{C}$-espacio.

9. Si consideramos en $\mathbb{R}$ la suma usual de números reales y $\mathbb{Q}$ como campo de escalares, y el producto de un elemento de $\mathbb{R}$ por un escalar de $\mathbb{Q}$ es el producto usual de números, entonces $\mathbb{R}$ es un espacio vectorial sobre $\mathbb{Q}$.

10. Análogamente al ejemplo anterior, si consideramos en $\mathbb{C}$ la suma usual de números complejos y $\mathbb{R}$ como campo de escalares, y el producto de un elemento de $\mathbb{C}$ por un escalar de $\mathbb{R}$ es el producto usual de números, entonces $\mathbb{C}$ es un espacio vectorial sobre $\mathbb{R}$.

11. Sea $C_a(\mathbb{R}, \mathbb{R})$ el conjunto de las funciones de $\mathbb{R}$ en $\mathbb{R}$ que son continuas en el punto $a \in \mathbb{R}$. Consideremos en este conjunto la suma usual de funciones y el producto de una función por un número real. La suma de dos funciones continuas en un punto a es también una función con-

tinua en ese punto, y lo mismo ocurre para el producto de una función continua en un punto por un escalar. Entonces, $C_a(\mathbb{R}, \mathbb{R})$ con estas operaciones usuales, es un espacio vectorial sobre $\mathbb{R}$.

12. Si consideramos el conjunto $C(\mathbb{R}, \mathbb{R})$ de las funciones de $\mathbb{R}$ en $\mathbb{R}$ continuas en todos los puntos del dominio, con las operaciones usuales con funciones, entonces tenemos en $C(\mathbb{R}, \mathbb{R})$ una estructura de espacio vectorial sobre $\mathbb{R}$.

13. Sea L_0 el conjunto de las funciones de $\mathbb{R}$ en $\mathbb{R}$ que tienen $\lim_{x\to 0} f(x)=0$.
Consideremos en este conjunto la suma de funciones y el producto de una función por un escalar. L_0 con estas operaciones tiene una estructura de espacio vectorial sobre $\mathbb{R}$.
Podemos generalizar el conjunto y considerar L_0^a como el conjunto de las funciones que en el punto a tienen límite cero. Este conjunto, con las mismas operaciones, tiene una estructura de espacio vectorial sobre $\mathbb{R}$.
Sin embargo, si cambiamos el valor del límite, es decir, si consideramos el conjunto F_a^a como el conjunto de las funciones de $\mathbb{R}$ en $\mathbb{R}$ que tienen límite a en el punto a, este conjunto no es un espacio vectorial para la suma y el producto usuales, pues si $f(x)$ tiene límite a en el punto a, entonces si tomamos un escalar cualquiera $\lambda\in\mathbb{R}$, la función $\lambda f(x)$, $\lambda\neq 1$, no está en L_a^a, pues $\lim_{x\to a} \lambda f(x)=\lambda a\neq a$.

14. Consideremos el conjunto $D_a(\mathbb{R}, \mathbb{R})$ de las funciones de $\mathbb{R}$ en $\mathbb{R}$ que son derivables en el punto a. Para la suma de funciones y el producto de una función por un escalar, este conjunto es un espacio vectorial sobre $\mathbb{R}$.

15. El axioma (3) de la definición 1.1 establece que un espacio vectorial tiene al menos un elemento. De acuerdo con esto, podemos presentar un último ejemplo de espacio vectorial, el más simple, que es el formado por un único elemento: el cero. Es decir, $\{0\}$ es un espacio vectorial sobre K y se denomina el *espacio trivial* sobre K.

1.3 Algunas propiedades elementales de los espacios vectoriales

A partir de la definición de espacio vectorial sobre K, podemos plantear algunas consecuencias inmediatas.

PROPOSICIÓN 1.1. Sea E un espacio vectorial sobre K. En E se cumplen:

(1) $a\cdot 0=0$ para todo $a\in K$.

(2) $0\cdot x=0$ para todo $x\in E$.

(3) Si $a\cdot x=0$, entonces $a=0$ o $x=0$.

(4) $-x=(-1)x$ para todo $x\in E$.

Demostración

Observemos ante todo que el símbolo 0 tiene dos interpretaciones: una como el *cero* del conjunto de los escalares y otra como el *vector* nulo.

Para (1):

La propiedad $a \cdot 0 = 0$ se demuestra probando que $a \cdot 0$ es el vector nulo de E. Este vector está caracterizado por el axioma (3) de la definición:

$x + 0 = x$ para todo $x \in E$.

Multiplicando ambos miembros de la igualdad por $a \in K$, obtenemos:

$a(x+0) = ax$,

de donde

$ax + a\,0 = ax$.

Como x es elemento de E y a es elemento de K, ax es elemento de E, y si lo denotamos por y, tenemos:

$y + a \cdot 0 = y$,

que no es más que la propiedad del vector cero, y como este es único, entonces,

$a \cdot 0 = 0$.

Para (2):

Aquí también tenemos que demostrar que $0x$ cumple la propiedad del neutro. Partimos de la identidad numérica

$1 + 0 = 1$.

Multiplicando por $x \in E$ y operando, obtenemos:

$(1+0)x = 1x$
$1x + 0x = 1x$
$x + 0x = x$ para todo $x \in E$.

Nuevamente esta es la propiedad que satisface el vector nulo, y por unicidad,

$0x = 0$.

Para (3):

Vamos a suponer que $a \neq 0$. Si a es un número no nulo, existe su inverso a^{-1} y $aa^{-1} = 1$. Entonces, si $ax = 0$, multiplicando ambos miembros por a^{-1} y operando, obtenemos:

$a^{-1}(ax) = a^{-1}(0) = 0$.

Pero

$$a^{-1}(ax)=(a^{-1}a)x=1x=x$$

y, por tanto, $x=0$, lo cual demuestra esta propiedad.
Para (4):

En este caso debemos demostrar que $(-1)x$ es el opuesto de x, basándonos en el axioma (4) de la definición. Para ello sumamos x y $(-1)x$:

$$x+(-1)x=(1)x+(-1)x=(1+(-1))x=(1-1)x=0x=0.$$

Por tanto, como el opuesto de un vector es único, tenemos:

$$(-1)x=-x.$$

Según esta propiedad, para encontrar, en la práctica, el opuesto de un vector dado, basta con multiplicar el vector por el escalar -1.

De estas propiedades y de los axiomas de espacio vectorial, resulta que en un espacio vectorial podemos seguir las mismas reglas operativas de cálculo que estamos acostumbrados a emplear en $\mathbf{R}^3$ y en otros ejemplos que vimos anteriormente.

1.4 Subespacio vectorial

Los ejemplos de $K[x]$ y $K_n[x]$ y los de las funciones continuas $C_a(\mathbf{R},\mathbf{R})$ y $C(\mathbf{R},\mathbf{R})$ nos pueden dar una idea de una definición natural de un subconjunto de un espacio vectorial que es también un espacio vectorial. $K_n[x]\subset K[x]$, $C(\mathbf{R},\mathbf{R})\subset C_a(\mathbf{R},\mathbf{R})$, o sea, para cada uno de estos pares de espacios se cumple que uno es subconjunto del otro, y cada uno es un espacio vectorial para las mismas operaciones.

La suma de polinomios que definimos en $K[x]$ es la misma que empleamos en $K_n[x]$. La suma de funciones que definimos en $C_a(\mathbf{R},\mathbf{R})$ se conserva en $C(\mathbf{R},\mathbf{R})$. Esto es válido también para el producto por un escalar. En estos casos se dice que $K_n[x]$ es un *subespacio* de $K[x]$ y que $C(\mathbf{R},\mathbf{R})$ es un *subespacio* de $C_a(\mathbf{R},\mathbf{R})$.

Nota: Entre los conjuntos $\mathbf{R}n[x]$, $K[x]$, $C(\mathbf{R},\mathbf{R})$ y $C_a(\mathbf{R},\mathbf{R})$ existen otras relaciones análogas a la explicada. Sugerimos al estudiante que las analice.

Tenemos, pues, un espacio vectorial E y un subconjunto no vacío de E que vamos a denotar por F. ¿Cuándo diremos que F es un subespacio de E?

DEFINICIÓN 1.2

Se dice que un subconjunto F de un K-espacio E es un *subespacio vectorial* de E si se cumple:

(1) La suma en E, restringida a los elementos de F, es una suma en F.

(2) El producto por un escalar de los elementos de E, restringido al subconjunto F, es un producto por un escalar en F.

(3) Para estas dos operaciones, suma y producto por un escalar, que se definen en F a partir de las de E, F es un espacio vectorial.

¿Qué significa esta definición desde el punto de vista práctico? ¿Cómo comprobamos que un subconjunto es o no es un subespacio vectorial de otro espacio? Vamos a analizar lo que implica la definición:

La suma en E es una suma en F.

Esto quiere decir que al sumar dos elementos de F el resultado está en F:

$$x, y \in F \Rightarrow x+y \in F.$$

El producto por un escalar en E, restringido a F, es un producto en F.

Al tomar un elemento de F y multiplicarlo por un escalar, el resultado es un elemento de F:

$$x \in F,\ \lambda \in K \Rightarrow \lambda x \in F.$$

F con estas dos operaciones es un espacio vectorial.

Hay que comprobar los axiomas de la definición de espacio vectorial.

(1) La suma es conmutativa.
No hay que probarlo pues si la suma es conmutativa en E (E es un K-espacio), es conmutativa para los elementos de un subconjunto de E.

(2) La suma es asociativa.
No es necesario comprobarlo pues si la suma es asociativa en E, también lo será en F.

(3) Existe un elemento 0 único.
Hay que comprobar que el 0 es uno de los elementos de F.

(4) Para todo $x \in F$ existe el opuesto.
Hay que comprobar para todo elemento x de F, que su opuesto $(-x)$ es también un elemento de F.

(5) Distributividad del producto por un escalar y la suma de vectores.
No hay que comprobarla porque si se cumple en E, se cumple en $F \subset E$.

(6) Distributividad del producto por un escalar y la suma de escalares.
No hay que comprobarla por la misma razón anterior.

(7) Asociatividad del producto.
Tampoco hay que comprobarla por las razones ya planteadas.

(8) Para $1 \in K$ se cumple:

$$1 \cdot u = u \text{ para todo } u \in F.$$

Este axioma se cumple, evidentemente, pues es válido para todos los elementos de E.

En la práctica basta comprobar lo siguiente:

F subconjunto de E (no vacío) es un subespacio de E si se cumple:

(1) $x, y \in F \Rightarrow x+y \in F$

(2) $x \in F,\ \lambda \in K \Rightarrow \lambda x \in F$

(3) $0 \in F$

(4) $x \in F \Rightarrow -x \in F$.

Ejemplos

1. $K_n[x]$ es un subespacio de $K[x]$.
   (1) La suma de dos polinomios de grado menor que n es un polinomio de grado menor que n.
   (2) El producto de un polinomio de grado menor que n por un escalar, es también un polinomio de grado menor que n.
   (3) $0 \in K_n[x]$ por definición de $K_n[x]$.
   (4) Si $f(x) \in K_n[x]$, entonces $-f(x)$ tiene el mismo grado que $f(x)$ y, por tanto, $-f(x) \in K_n[x]$.
2. $\mathbb{R}$ es un subespacio de $\mathbb{C}$ si se considera $\mathbb{C}$ como $\mathbb{R}$-espacio.
3. $\mathbb{R}$ no es un subespacio de $\mathbb{C}$ si se considera $\mathbb{C}$ como $\mathbb{C}$-espacio.
   Al tratar de comprobar (2), vemos que si tomamos un vector cualquiera $x \in \mathbb{R}$ y $a \in \mathbb{C}$, el producto ax no es siempre un número real. Por tanto, $\mathbb{R}$ no puede ser un subespacio de $\mathbb{C}$ como $\mathbb{C}$-espacio.
4. Si consideramos al conjunto de las funciones de $\mathbb{R}$ en $\mathbb{R}$, $\mathrm{F}(\mathbb{R}, \mathbb{R})$, este conjunto con la suma de funciones y el producto por un escalar es un espacio vectorial sobre $\mathbb{R}$.
   De los ejemplos de espacios vectoriales que hemos presentado hay varios que son subespacios de $\mathrm{F}(\mathbb{R}, \mathbb{R})$; en particular analizaremos el caso de L_0 (ejemplo 13). Realmente no es necesario comprobar nada, pues ya vimos que L_0 es un espacio vectorial para las mismas operaciones definidas en $\mathrm{F}(\mathbb{R}, \mathbb{R})$ y, por tanto, es un subespacio de $\mathrm{F}(\mathbb{R}, \mathbb{R})$ pero vamos a comprobar las cuatro condiciones dadas.
   (1) $f(x),\ g(x) \in L_0$ significa que
   $$\lim_{x\to 0} f(x) = 0, \quad \lim_{x\to 0} g(x) = 0.$$
   Para comprobar si $f(x) + g(x)$ está en L_0, hay que calcular
   $$\lim_{x\to 0} (f(x) + g(x)) = \lim_{x\to 0} f(x) + \lim_{x\to 0} g(x) = 0 + 0 = 0.$$
   Por tanto, $f(x) + g(x) \in L_0$
   (2) Sea $f(x) \in L_0$ y $a \in \mathbb{R}$. Para comprobar que $af(x) \in L_0$, hay que calcular:
   $$\lim_{x\to 0} a f(x) = a \lim_{x\to 0} f(x) = a\,0 = 0,$$
   lo que significa que $a f(x) \in L_0$.
   (3) El 0 de $\mathrm{F}(\mathbb{R}, \mathbb{R})$ es la función indénticamente nula, y esta función tiene por límite en el origen, 0. Por tanto, $0 \in L_0$.
   (4) Si $f(x) \in L_0$, entonces $\lim\limits_{x\to 0} -f(x) = -(0) = 0$. Por tanto, $-f(x) \in L_0$.

5. En $F(\mathbb{R},\mathbb{R})$ consideremos el subconjunto L_a de las funciones que tienen límite a cuando x tiende a cero. Veamos si este conjunto es un subespacio de $F(\mathbb{R},\mathbb{R})$.

(1) Sean $f(x)$, $g(x) \in L_a$, es decir,

$$\lim_{x\to 0} f(x) = a, \quad \lim_{x\to 0} g(x) = a.$$

Para comprobar si $f(x)+g(x)$ está en L_a hay que calcular:

$$\lim_{x\to 0} (f(x)+g(x)) = \lim_{x\to 0} f(x) + \lim_{x\to 0} g(x) = a+a=2\,a.$$

Pero para que $f(x)+g(x) \in L_a$ tendría que ser

$$\lim_{x\to 0} (f(x)+g(x)) = a \neq 2a \text{ si } a\neq 0.$$

Por tanto, la única posibilidad de que un subconjunto L_a de $F(\mathbb{R},\mathbb{R})$ sea un subespacio vectorial de $F(\mathbb{R},\mathbb{R})$ es el ejemplo que ya vimos de L_0.

6. Sea A una matriz compleja $m\times n$. Consideremos el sistema de ecuaciones compatibles $AX=B$,
y sea S_B el conjunto de las soluciones del sistema, que es un subconjunto de $\mathbb{C}^n$. Queremos saber en qué caso S_B es un subespacio de $\mathbb{C}^n$.
Veamos si se cumplen las condiciones:

(1) Sean X, $Y\in\mathbb{C}^n$ tales que $AX=B$ y $AY=B$. Entonces
$A(X+Y)=AX+AY=B+B=2B$.
Si $X+Y$ es una solución, tenemos:
$A(X+Y)=B\neq 2B$ si $B\neq 0$.
Por tanto, la única posibilidad es $B=0$.

(2) Sea $X\in S_B, \lambda\in\mathbb{C}$. Analicemos λX.
$A(\lambda X)=\lambda(AX)=\lambda B\neq B$ si $\lambda\neq 1$, $B\neq 0$.
Por tanto, aquí también la única posibilidad de que $\lambda X\in S_B$ para todo $\lambda\in\mathbb{C}$, es que $B=0$.

(3) $0\in S_B$, es decir, el vector nulo es solución del sistema $AX=B$. Esto ocurre solamente si $B=0$.

(4) Si $X\in S_B$, entonces
$A(-X)=-AX=-B\neq B$ si $B\neq 0$.
Nuevamente encontramos la condición $B=0$.

Como conclusión podemos decir que solamente los conjuntos de soluciones de los sistemas de ecuaciones lineales homogéneos son subespacios del espacio $\mathbb{C}^n$.

7. En todo espacio vectorial E siempre hay dos subespacios que se denominan *triviales;* estos son $\{0\}$ y el propio E.
Vamos a analizar de nuevo las condiciones que permiten comprobar si un conjunto dado es un subespacio vectorial de un espacio dado. Veamos primeramente las condiciones (3) y (4).

(3) El vector nulo, 0, pertenece a F.

Esta condición no hay que comprobarla si se cumple la condición (2):
$x \in F$, $\lambda \in K \Rightarrow \lambda x \in F$, pues de ser así, en particular para el escalar $\lambda=0$ tenemos:
$0x=0 \in F$.

(4) Para todo $x \in F \Rightarrow -x \in F$.
Esta condición tampoco hay que comprobarla, pues si se cumple (2), tomando en particular el escalar $\lambda=-1$ tenemos:
$x \in F \Rightarrow (-1)\ x=-x \in F$.

Por tanto, es suficiente comprobar (1) y (2) para afirmar que $F \subset E$ es un subespacio de E. Sin embargo, quisimos mantener (3) y (4) en las condiciones a comprobar pues estas pueden ayudar a decidir si un subconjunto *no* es un subespacio vectorial. Por ejemplo, en el caso de los sistemas $AX=B$, los no homogéneos se pueden eliminar inmediatamente pues el vector trivial o vector cero no es una solución de un sistema no homogéneo.

Vamos a establecer una propiedad equivalente a la definición que permitirá caracterizar los subespacios.

TEOREMA 1.1

Sean E un espacio vectorial sobre K y F un subconjunto no vacío de E. Entonces F es un subespacio de E si y solo si

$$x, y \in F,\ \alpha, \beta \in K \Rightarrow \alpha x+\beta y \in F.$$

Demostración

1) Probemos que:
Si F es un subespacio, entonces F cumple:
$x, y \in F$, $\alpha, \beta \in K \Rightarrow \alpha x+\beta y \in F$.
Si F es un subespacio, tenemos:
i) $u, v \in F \Rightarrow u+v \in F$
ii) $u \in F$, $\lambda \in K \Rightarrow \lambda u \in F$
En este caso; $x, y \in F$, $\alpha, \beta \in K \Rightarrow \alpha x \in F$, $\beta y \in F$ por ii).
Pero si $\alpha x \in F$, $\beta y \in F \Rightarrow \alpha x+\beta y \in F$ por i),
y esto demuestra la primera parte del teorema.

2) Probemos ahora que:
Si F cumple
$x, y \in F$, $\alpha, \beta \in K \Rightarrow \alpha x+\beta y \in F$, (*)
entonces F es un subespacio.
Tenemos que demostrar que F satisface:
i) $x, y \in F \Rightarrow x+y \in F$
ii) $x \in F$, $\lambda \in K \Rightarrow \lambda x \in F$

Como la implicación (*) se cumple para valores cualesquiera de $x, y \in F$ y $\alpha, \beta \in K$, se cumple también para los casos particulares: $\alpha=\beta=1$, $\beta=0$.
Así, en el caso $\alpha=\beta=1$ tenemos:
$x, y \in F \Rightarrow 1x+1y=x+y \in F$,

que es la implicación i), y en el caso $\beta=0$:
$x, y \in F$; $\alpha, 0 \in K \Rightarrow \alpha x + 0y = \alpha x \in F$,
que es la implicación ii). Con esto termina la demostración.

El tipo de expresión $\alpha x + \beta y$ es muy importante y vamos a emplearla frecuentemente en lo adelante. Ahora la utilizaremos para analizar nuevos ejemplos de subespacios.

Ejemplos

1. Analicemos los subespacios de $\mathbb{R}^3$. Ya conocemos dos de los subespacios de $\mathbb{R}^3$: $\{0\}$ y $\mathbb{R}^3$. Vamos pues, a analizar otros casos.

   Primer caso: $F_1=\{$un vector y sus múltiplos$\}$
   Sea $F_1 \subset \mathbb{R}^3$ y supongamos que $x \in F_1$, $x \neq 0$. Evidentemente, todos los múltiplos de x deben estar en F_1. Comprobemos que este conjunto, es decir, $\{\lambda x / \lambda \in K\}$ es un subespacio de $\mathbb{R}^3$.
   Según la caracterización dada en el teorema 1.1
   $u, v \in F_1$; $\alpha, \beta \in K \Rightarrow \alpha u + \beta v \in F_1$.

   Pero
   $u \in F_1 \Longleftrightarrow u = \lambda_1 x_1$, $\lambda_1 \in \mathbb{R}$,
   $v \in F_1 \Longleftrightarrow v = \lambda_2 x_1$, $\lambda_2 \in \mathbb{R}$.

   Por tanto,
   $$\begin{aligned} \alpha u + \beta v &= \alpha\ (\lambda_1 x_1) + \beta\,(\lambda_2 x_1) \\ &= (\alpha\lambda_1)\, x_1 + (\alpha\lambda_2)\, x_1 \\ &= (\alpha\lambda_1 + \alpha\lambda_2)\, x_1 \\ &= \lambda x_1 \in F_1,\ \lambda \in \mathbb{R}. \end{aligned}$$
   Luego, F_1 es un subespacio de $\mathbb{R}^3$.
   Los subconjuntos del tipo F_1, formados por los múltiplos de un vector no nulo, son subespacios de $\mathbb{R}^3$. Geométricamente, un subespacio del tipo F_1 es una *recta que pasa por el origen de coordenadas.*

   Segundo caso: $F_2=\{$dos vectores no colineales y sus múltiplos y sumas$\}$
   Sean x_1, x_2 dos vectores no colineales y $F_2=\{\lambda x_1 + \mu x_2 / \lambda, \mu \in K\}$. Este subconjunto es un subespacio de $\mathbb{R}^3$ pues se puede comprobar que satisface la caracterización dada en el teorema 1.1.
   Como $u, v \in F_2$; $\alpha, \beta \in K$, tenemos:
   $u = \lambda_1 x_1 + \lambda_2 x_2$, $v = \mu_1 x_1 + \mu_2 x_2$, $\lambda_1, \lambda_2, \mu_1, \mu_2 \in \mathbb{R}$.

   Entonces,
   $$\begin{aligned} \alpha u + \beta v &= \ \alpha\ (\lambda_1 x_1 + \lambda_2 x_2) + \beta\ (\mu_1 x_1 + \mu_2 x_2) \\ &= \alpha\lambda_1 x_1 + \alpha\lambda_2 x_2 + \beta\mu_1 x_1 + \beta\mu_2 x_2 \\ &= (\alpha\lambda_1 + \beta\mu_1)\ x_1 + (\alpha\lambda_2 + \beta\mu_2)\, x_2. \end{aligned}$$

   Por tanto, $\alpha u + \beta v \in F_2$ y F_2 es un subespacio.
   Este tipo de subespacio, que contiene a dos vectores no colineales y sus combinaciones, tiene también una interpretación geométrica: es un *plano,* determinado por los vectores x_1, x_2, *que pasa por el origen de coordenadas.*

Las posibilidades que tenemos para un subespacio de $\mathbb{R}^3$ son:

$\{0\}$

Una recta que pasa por el origen.

Un plano que pasa por el origen.

$\mathbb{R}^3$

Pudiéramos demostrar que estas son las únicas posibilidades, pero lo dejaremos para más adelante, pues será más fácil. Ahora nos interesa sobre todo la interpretación geométrica de planos y rectas como subespacios.

2. Sean E un espacio vectorial sobre K y $v \in E$ un vector no nulo. Consideremos el conjunto
$E_1 = \{\lambda v / \lambda \in K\}$.
Este subconjunto de E es un subespacio vectorial, según la caracterización de subespacios dada en el teorema 1.1.

Sean $x \in E_1$, $y \in E_1$, $\alpha, \beta \in K$. Hay que probar que
$\alpha x + \beta y \in E_1$.

$x \in E_1 \Longleftrightarrow \exists \lambda_1 \in K / x = \lambda_1 v$
$y \in E_1 \Longleftrightarrow \exists \lambda_2 \in K / y = \lambda_2 v$

Por tanto:

$$\begin{aligned}\alpha x + \beta y &= \alpha\ (\lambda_1 v)\ + \beta\ (\lambda_2 v)\\ &= (\alpha\lambda_1)\, v + (\beta\lambda_2)\, v\\ &= (\alpha\lambda_1 + \beta\lambda_2)\, v\end{aligned}$$

Y como $\alpha, \beta, \lambda_1, \lambda_2 \in K$, entonces $\alpha\lambda_1 + \beta\lambda_2 \in K$ y, por tanto,
$\alpha x + \beta y \in E_1$.

Un subespacio de la forma de E_1 se denomina, por analogía con el caso de $\mathbb{R}^3$, la *recta* generada por el vector $v \in E$ y se denota por Kv.

3. Sea E un espacio vectorial sobre K y sean v, $w \in E$ dos vectores no nulos de E. Consideremos el conjunto
$E_2 = \{\lambda v + \mu w / \lambda,\ \mu \in K\}$.
Este subconjunto de elementos de E es un subespacio vectorial. Para comprobarlo, basta demostrar que
$x, y \in E_2$ y $\alpha, \beta \in K \Rightarrow \alpha x + \beta y \in E_2$.
Tenemos que
$x \in E_2 \Longleftrightarrow \exists \lambda_1, \mu_1 \in K /\ x = \lambda_1 v + \mu_1 w$,
$y \in E_2 \Longleftrightarrow \exists \lambda_2, \mu_2 \in K / y = \lambda_2 v + \mu_2 w$.
Por tanto, para todo par de escalares $\alpha, \beta \in K$:
$$\begin{aligned}\alpha x + \beta y &= \alpha\ (\lambda_1 v + \mu_1 w)\ + \beta\ (\lambda_2 v + \mu_2 w)\\ &= \alpha\lambda_1 v + \alpha\mu_1 w + \beta\lambda_2 v + \beta\mu_2 w\\ &= (\alpha\lambda_1 + \beta\lambda_2)\, v + (\alpha\mu_1 + \beta\mu_2)\, w\end{aligned}$$
Como $\alpha, \beta, \lambda_1, \lambda_2, \mu_1, \mu_2 \in K$, entonces
$\alpha\lambda_1 + \beta\lambda_2 \in K$ y $\alpha\mu_1 + \beta\mu_2 \in K$

y, por tanto,

$\alpha x+\beta y\in E_2$.

Si los vectores v y w no son proporcionales, es decir, no existe un escalar λ tal que $v=\lambda w$, los subespacios del tipo E_2 se denominan, por analogía con el caso de $\mathbb{R}^3$, el *plano* generado por los vectores v y w.

Si uno de los vectores es un múltiplo del otro, por ejemplo $v=\lambda w$, entonces el subespacio que hemos definido se reduce a la recta generada por el vector w, es decir, a Kw.

1.5 Isomorfismo de espacios vectoriales

Cada vez que en Matemática se define una estructura, es necesario precisar en qué caso dos de ellas son esencialmente diferentes y en cual, dos de ellas son semejantes o equivalentes.

DEFINICIÓN 1.3

Sean E y F dos espacios vectoriales sobre un mismo campo escalar K. Se dice que E y F son isomorfos o *equivalentes* si existe una aplicación $f: E \longrightarrow F$ biyectiva tal que para todo $x, y\in E$; $\alpha,\beta\in K$, se cumple:

$$f(\alpha x+\beta y)=\alpha f(x)+\beta f(y).$$

La aplicación f se denomina un *isomorfismo* entre los espacios vectoriales E y F.

Ejemplo

El espacio $\mathbf{R}_n[x]$ es isomorfo a $\mathbf{R}^n$.

Para construir un isomorfismo recordemos que, según la definición de polinomio en una indeterminada, cada polinomio $p(x)$ está determinado por sus coeficientes:

$$a_0, a_1, a_2, \ldots, a_n, \ldots$$

donde a partir de un cierto valor de n, todos son nulos.

Como estamos trabajando con polinomios de grado menor que n, los coeficientes $a_n, a_{n+1}\ldots$ son todos nulos, y por ello, consideramos solamente los coeficientes $a_0, a_1, \ldots, a_{n-1}$; es decir, todo polinomio $p(x)$ de $\mathbf{R}_n[x]$ se escribe en forma única de la manera siguiente:

$$p(x)=a_0+a_1x+a_2x^2+\ldots+a_{n-1}x^{n-1},$$

donde $a_0, a_1, \ldots, a_{n-1}$ son únicos para cada polinomio. Establezcamos ahora la aplicación

$$f\colon \mathbf{R}_n[x]\longrightarrow \mathbf{R}^n$$

definida por

$$f(p(x))=(a_0, a_1, a_2, \ldots, a_{n-1}),$$

donde

$$p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}.$$

Vamos a demostrar que f es un isomorfismo.

1) f es inyectiva.
Sean $p(x) = a_0 + a_1 x + \ldots + a_{n-1} x^{n-1}$, $q(x) = b_0 + b_1 x + \ldots + b_{n-1} x^{n-1}$ dos polinomios de $\mathbb{R}_n[x]$ y supongamos:

$$f(p(x)) = f(q(x)).$$

Ahora bien,

$$f(p(x)) = (a_0, a_1, \ldots, a_{n-1})$$
$$f(q(x)) = (b_0, b_1, \ldots, b_{n-1})$$

Luego

$$f(p(x)) = f(q(x)) \Rightarrow \begin{cases} a_0 = b_0 \\ a_1 = b_1 \\ \vdots \\ a_{n-1} = b_{n-1} \end{cases}$$

y esto significa que

$$p(x) = q(x).$$

Por tanto, f es inyectiva.

2) f es sobreyectiva.
Sea $u = (c_0, c_1, \ldots, c_{n-1})$ un elemento cualquiera de $\mathbb{R}^n$ es evidente que si tomamos el polinomio $r(x) = c_0 + c_1 x + \ldots + c_{n-1} x^{n-1}$ de $\mathbb{R}_n[x]$, entonces

$$f(r(x)) = u.$$

Por tanto, f es sobreyectiva.

3) f cumple

$$f(\alpha p(x) + \beta q(x)) = \alpha f(p(x)) + \beta f(q(x))$$

para todo par de polinomios $p(x)$, $q(x) \in \mathbb{R}_n[x]$ y todo par de números $\alpha, \beta \in \mathbb{R}$, y como

$$p(x) = a_0 + a_1 x + \ldots + a_{n-1} x^{n-1},$$
$$q(x) = b_0 + b_1 x + \ldots + b_{n-1} x^{n-1},$$

entonces

$$\alpha p(x) = \alpha a_0 + \alpha a_1 x + \ldots + \alpha a_{n-1} x^{n-1},$$
$$\beta q(x) = \beta b_0 + \beta b_1 x + \ldots + \beta b_{n-1} x^{n-1}.$$

Luego

$$\alpha p(x) + \beta q(x) = (\alpha a_0 + \beta b_0) + (\alpha a_1 + \beta b_1) x + \ldots + (\alpha a_{n-1} + \beta b_{n-1}) x^{n-1}.$$

Por tanto,

$$\begin{aligned} f(\alpha p(x) + \beta q(x)) &= (\alpha a_0 + \beta b_0, \alpha a_1 + \beta b_1, \ldots, \alpha a_{n-1} + \beta b_{n-1}) \\ &= (\alpha a_0, \alpha a_1, \ldots, \alpha a_{n-1}) + (\beta b_0, \beta b_1, \ldots, \beta b_{n-1}) \\ &= \alpha (a_0, a_1, \ldots, a_{n-1}) + \beta (b_0, b_1, \ldots, b_{n-1}) \\ &= \alpha f(p(x)) + \beta f(q(x)). \end{aligned}$$

Así, tenemos que $\mathbb{R}_n[x]$ es isomorfo a $\mathbb{R}^n$.

Así, tenemos que $\mathbb{R}_n[x]$ es isomorfo a $\mathbb{R}^n$.

Esta demostración se puede realizar con un conjunto cualquiera de escalares K, y así demostrar que $K_n[x]$ es isomorfo a K^n.

A partir de este ejemplo podemos darnos cuenta del significado de la definición de isomorfismo. Al exigir que la aplicación f sea biyectiva, estamos exigiendo en primer lugar que los conjuntos sean equivalentes, que tengan el mismo cardinal. Pero esta primera condición no afirma nada sobre la estructura algebraica en E y en F y hemos dicho que estas estructuras son equivalentes. Este sentido de equivalencia de las estructuras hay que buscarlo en la condición

$$f(\alpha x+\beta y)=\alpha f(x)+\beta f(y),$$

lo cual se expresa en las condiciones siguientes:

(1) $f(x+y)=f(x)+f(y)$
(2) $f(\lambda x)=\lambda f(x)$,
$x, y \in E;\ \lambda \in K$.

Estas dos condiciones ponen en evidencia que las operaciones algebraicas tienen como resultados en los dos conjuntos elementos que se corresponden.

Si tomamos dos elementos x, y en E y buscamos su suma $x+y$, esto corresponde en F a buscar las imágenes de los elementos y sumar estas imágenes. Lo mismo sucede con el producto por un escalar, si tenemos un vector x y su imagen $f(x)$, para multiplicar $f(x)$ por un escalar λ, basta multiplicar x por λ y buscar la imagen de λx.

En resumen, si f es un isomorfismo, la suma de dos elementos de E se transforma en la suma de los elementos correspondientes en F, y el producto de un elemento de E por un escalar se transforma en el producto por el mismo escalar de la imagen del elemento.

1.6 Combinación lineal de un sistema finito de vectores de un espacio vectorial

En algunos de los problemas tratados en el epígrafe anterior, aparecieron expresiones del tipo

$$\alpha x+\beta y,$$

donde x, y eran elementos de un espacio vectorial E, y α, β escalares. Este tipo de expresión es una combinación de las dos operaciones que existen en un espacio vectorial; una suma de productos por un escalar.

Vamos a dar una definición precisa de qué entendemos por una *combinación lineal*. Por necesidades de trabajo en los espacios vectoriales, tendremos que hacer este tipo de combinaciones con varios vectores y para ello definiremos el concepto *sistema de vectores*.

DEFINICIÓN 1.4

Sea E un espacio vectorial sobre K. Se denomina *sistema de vectores* de E, todo subconjunto ordenado de vectores de E.

En este caso establecemos una diferencia entre un *subconjunto de vectores*, que no tienen que estar dados en un cierto orden, y *un sistema de vectores* del cual debemos dar no solo los elementos que lo forman, sino también el orden de estos. Además, en un sistema se pueden repetir vectores pero en un conjunto los elementos tienen que ser diferentes entre sí.

Por ejemplo, si en $\mathbb{R}^3$ tomamos los vectores $e_1=(1,0,0)$, $e_2=(0,1,0)$ y $e_3=(0, 0, 1)$, estos tres vectores forman un subconjunto de vectores de $\mathbb{R}^3$, y cuando hablamos de ellos como conjunto $\{e_1, e_2, e_3\}$ no nos interesa el orden, es decir, da lo mismo decir $\{e_3, e_1, e_2\}$. Si tenemos interés en que se tomen los vectores en un cierto orden, hablaremos del *sistema* de vectores formado por (e_1, e_2, e_3), y de este modo el orden ya está fijado. Por tanto, el sistema (e_1, e_2, e_3) y el sistema (e_3, e_1, e_2) son dos sistemas de vectores diferentes, porque los vectores se toman en distinto orden.

A primera vista esta terminología puede parecer no necesaria y un tanto preciosista, pero en casos determinados comprobaremos la necesidad de tomar un sistema de vectores y no un conjunto.

DEFINICIÓN 1.5

Sean E un espacio vectorial sobre K, $(a_1,a_2,\ldots,a_n)$ un sistema finito de vectores de E y $\lambda_1,\lambda_2,\ldots,\lambda_n$ una familia de escalares de K. Se denomina *combinación lineal* del sistema de vectores $(a_1,a_2,\ldots,a_n)$ con los escalares $\lambda_1,\lambda_2,\ldots,\lambda_n$, a la expresión

$$\lambda_1a_1+\lambda_2a_2+\ldots+\lambda_na_n.$$

Los escalares $\lambda_1,\lambda_2,\ldots,\lambda_n$ se denominan los *coeficientes* de la combinación lineal.

Ejemplos

1. Si en $\mathbb{R}^3$ consideramos los vectores $e_1=(1,0,0)$, $e_2=(0,1,0)$, $e_3(0,0,1)$, las siguientes expresiones son combinaciones lineales del sistema (e_1, e_2, e_3) de vectores de $\mathbb{R}^3$:
   a) $2e_1+e_2+3e_3$
   b) $0e_1+e_2+0e_3$
   c) $e_1\ -e_2+e_3$
   d) $\pi e_1+\sqrt{2}\,e_2-0e_3$
2. En el espacio vectorial $\mathbb{R}_n\,[x]$ consideremos el sistema de vectores (polinomios)
   $(1,x,x^2,\ldots,x^{n-1})$.

   Son combinaciones lineales de este sistema:
   a) $1+3x+2x^2+0x^3+\ldots+0x^{n-1}$
   b) $a_0+a_1x+a_2x^2+\ldots+a_{n-1}x^{n-1},\ a_0, a_1, a_2,\ldots,a_{n-1}\in\mathbb{R}$

3. En $\mathbb{R}^4$ tenemos la siguiente combinación lineal:

   $2(1,2,3,0) - 1(0,-1,1,1) - 1(2,5,5,-1)$

   de los vectores $(1,2,3,0)$, $(0,-1,1,1)$, $(2,5,5,-1)$ con los escalares $2, -1, -1$.
4. En $\mathbb{C}$ como $\mathbb{R}$-espacio vamos a considerar los vectores $e_1=1$, $e_2=i$. Una combinación lineal de ellos es una expresión de la forma:

   $ae_1+be_2=a1+bi=a+bi$,

   donde a y b son números reales.

Como podemos ver en los ejemplos, una combinación lineal indica operaciones de suma de vectores y producto por un escalar, y el resultado de estas operaciones es un vector del espacio. Si tomamos algunos de los ejemplos y realizamos las operaciones, obtenemos los resultados siguientes:

1. a) $(2,1,3)$

   b) $(0,1,0)=e_2$
2. $1+3x+2x^2$
3. $(0,0,0,0)$
4. El resultado de efectuar la combinación lineal es un número complejo cuya parte real es el número a, y cuya parte imaginaria es el número b.

Si el resultado de la combinación lineal

$$\alpha_1 a_1+\alpha_2 a_2+\ldots+\alpha_n a_n$$

es un vector x, escribimos:

$$x=\alpha_1 a_1+\alpha_2 a_2+\ldots+\alpha_n a_n$$

y decimos que el vector x es el resultado de la combinación lineal dada o que x es combinación lineal de $(a_1,a_2,\ldots,a_n)$.

Es necesario observar bien la diferencia que establecemos entre la combinación lineal

$$\alpha_1 a_1+\alpha_2 a_2+\ldots+\alpha_n a_n$$

y el *resultado* de efectuar la combinación lineal. Esto puede parecer un preciosismo de lenguaje, pero cuando avancemos en el estudio de los espacios vectoriales, el estudiante se percatará de que la combinación lineal es un instrumento de trabajo y que el resultado de la combinación se emplea con otros objetivos.

Si el sistema que tomamos tiene un solo vector (a), una combinación lineal es una expresión de la forma

$$\lambda a, \; \lambda\in K,$$

y el resultado de la combinación lineal es un vector proporcional al vector a. En este caso, por analogía con la forma de los subespacios de $\mathbb{R}^3$, decimos que el resultado de la combinación lineal se encuentra en la *recta* determinada por a.

Si el sistema tiene dos vectores (u, v), una combinación lineal es de la forma

$$\lambda u + \mu v, \quad \lambda, \mu \in K,$$

y por analogía con el caso $\mathbf{R}^3$, si los vectores u, v no son proporcionales, decimos que el resultado de la combinación lineal se encuentra en el *plano* determinado por los vectores u y v.

Para todo sistema de vectores hay una combinación lineal que recibe un nombre especial, y es aquella en la cual todos los coeficientes son iguales a cero:

$$0a_1 + 0a_2 + \ldots + 0a_n.$$

Esta combinación lineal se denomina *combinación trivial* y su resultado es siempre el vector cero. Sin embargo, en los ejemplos dados podemos ver que la combinación trivial no es la única que tiene como resultado el vector cero.

En el concepto de combinación lineal se basan los problemas fundamentales que trataremos en el resto del capítulo.

Aunque nuestro objetivo es tratar solamente problemas con conjuntos finitos de vectores, vamos a explicar qué sucede si queremos hacer una combinación lineal de un sistema infinito de vectores.

Sea $(a_i)_{i \in I}$ una familia infinita de vectores de un espacio vectorial E sobre un campo K. Si tratamos de extender directamente la definición de combinación lineal, nos enfrentamos con una dificultad: obtenemos una expresión,

$$\sum_{i \in I} \alpha_i a_i,$$

donde $\alpha_i \in K$ para $i \in I$. Esto es una suma de infinitos elementos de E que no podemos efectuar. Para poder extender entonces el concepto combinación lineal, buscaremos la forma de obtener sumas finitas.

Sea $(a_i)_{i \in I}$ un sistema de vectores de E. Una combinación lineal del sistema es una expresión del tipo

$$\sum_{i \in I} \alpha_i a_i,$$

donde $\alpha_i \in K$ para $i \in I$, y todos los valores de los coeficientes son iguales a cero, excepto un número finito de ellos. De esta forma garantizamos que aunque hay (teóricamente) infinitos términos en la suma, casi todos son nulos, por lo que hay que sumar solamente un número finito de ellos.

La expresión: "todos los coeficientes son cero excepto un número finito de ellos" se encontrará muchas veces en problemas de esta índole; también diremos brevemente: "casi todos son cero", y si el conjunto de índices I es el conjunto de los números naturales, $\mathbf{N}$, diremos que "a partir de un valor de n, todos son cero". Esta última expresión ya la encontramos cuando definimos los polinomios y servirá para dar un ejemplo de combinación lineal infinita.

Debemos observar que esta definición de combinación lineal de un sistema de vectores (sin precisar si es infinito o no el conjunto de índices) contiene el caso en que el sistema es finito, pues si I es finito, afirmar que los coeficientes son nulos excepto un número finito de ellos, no impone ninguna condición sobre los coeficientes, ya que en este caso el conjunto de los coeficientes es finito.

Ejemplo

1. Consideremos el espacio $K[x]$ de los polinomios en una indeterminada con coeficientes en K.
   En $K[x]$ tomamos la familia infinita.
   $(1, x, x^2, \ldots, x^n, \ldots)$.
   Si tomamos la combinación lineal
   $1+2x+x^2+0x^3+\ldots+0x^n+\ldots$, el resultado de esta combinación es el polinomio $1+2x+x^2$.
2. Consideremos un polinomio cualquiera $p(x)=a_0+a_1x+\ldots+a_nx^n$.
   Recordemos que una de las definiciones de polinomio es la dada por la sucesión de sus coeficientes
   $(a_0, a_1, a_2, \ldots, a_n, \ldots,)$,
   y que un polinomio es una sucesión de números tal que "a partir de un valor de n, todos son cero". Esta expresión es exactamente la que empleamos para la combinación lineal.
   Si tomamos la familia $(x^n)_{n\in\mathbb{N}}$ y como coeficientes de la combinación lineal la sucesión de coeficientes de $p(x)$, podemos escribir la combinación lineal

$$\sum_k a_kx^k = a_0 + a_1x + a_2x^2 + \ldots + a_nx^n + \ldots$$

y el resultado de la combinación lineal es exactamente el polinomio $p(x)$.
En este ejemplo queremos señalar que aunque la definición parezca más compleja, no hay ninguna dificultad nueva, pues ya hemos trabajado con combinaciones lineales de sistemas infinitos. El trabajo con polinomios no es más que esto: trabajar con combinaciones lineales del sistema $(1, x, x^2, \ldots, x^n, \ldots)$.

1.7 El subespacio generado por un conjunto de vectores

Si en un espacio vectorial E tenemos un conjunto finito de vectores $\{a_1, a_2, \ldots, a_k\}$, podemos plantearnos el problema siguiente:

¿Cuál es el menor subespacio de E que contiene al conjunto $\{a_1, a_2, \ldots, a_k\}$?

Vamos a demostrar la existencia de este subespacio que denominaremos *el subespacio generado por el conjunto* $\{a_1, a_2, \ldots, a_k\}$.

Si pensamos un poco en las condiciones que tiene que cumplir un subespacio vectorial que contenga al conjunto $A = \{a_1, a_2, \ldots, a_k\}$, podemos obtener las conclusiones siguientes:

1ra. Si un subespacio vectorial contiene a A, contiene a los múltiplos de los vectores de A.

2da. Si se cumple lo anterior, el subespacio vectorial debe contener, además, las sumas de los elementos de A y de sus múltiplos.

En resumen, si un subespacio vectorial contiene al conjunto A, debe contener a todas las combinaciones lineales del sistema formado por A, es decir, a los vectores de la forma

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k, \quad \lambda_1, \ldots, \lambda_k \in K.$$

Vamos a demostrar que el conjunto de todos los vectores que se pueden expresar como combinación lineal del sistema A, es un subespacio vectorial de E, y es, además, el menor subespacio que contiene al conjunto A.

LEMA 1.1

Sean E un espacio vectorial sobre K y $A = \{a_1, a_2, \ldots, a_k\}$ un conjunto de vectores de E. Entonces el conjunto

$$F = \{v \in E / v = \lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k;\ \lambda_i \in K\}$$

es un subespacio vectorial de E, y si un subespacio F' contiene al conjunto A, entonces $F \subset F'$.

Demostración

Probemos que F es un subespacio, a partir de $v, u \in F \Rightarrow \alpha v + \beta u \in F$ para todo $\alpha, \beta \in K$.

$v \in F \Longleftrightarrow \exists \lambda_1, \lambda_2, \ldots, \lambda_k \in K$ tales que

$$v = \lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k,$$

$u \in F \Longleftrightarrow \exists \mu_1, \mu_2, \ldots, \mu_k \in K$ tales que

$$u = \mu_1 a_1 + \mu_2 a_2 + \ldots + \mu_k a_k.$$

Entonces, para todo $\alpha, \beta \in K$ tenemos:

$$\begin{aligned} \alpha v + \beta u &= \alpha(\lambda_1 a_1 + \ldots + \lambda_k a_k) + \beta(\mu_1 a_1 + \ldots + \mu_k a_k) \\ &= \alpha\lambda_1 a_1 + \ldots + \alpha\lambda_k a_k + \beta\mu_1 a_1 + \ldots + \beta\mu_k a_K \\ \alpha v + \beta u &= (\alpha\lambda_1 + \beta\mu_1) a_1 + (\alpha\lambda_2 + \beta\mu_2) a_2 + \ldots + (\alpha\lambda_k + \beta\mu_k) a_k. \end{aligned}$$

Esta expresión es una combinación lineal de los vectores de A y, por tanto, $\alpha v + \beta u \in F$ y F es un subespacio.

A partir del razonamiento anterior, resulta de inmediato que F es el menor subespacio de E que contiene a A, pues si un subespacio contiene a un conjunto de vectores, tiene que contener a sus combinaciones lineales.

Por tanto, todo subespacio F' que contenga al conjunto A, tiene que contener a F.

Ejemplos

1. Consideremos en $\mathbf{R}^3$ el subespacio generado por los vectores $e_1=(1,0,0)$ y $e_2=(0,1,0)$. Este subespacio es el de los vectores que se pueden expresar como combinación lineal de ellos, es decir, en la forma:
   $v=\lambda_1 e_1+\lambda_2 e_2$
   $=(\lambda_1,0,0)+(0,\lambda_2,0)=(\lambda_1,\lambda_2,0),\ \lambda_1,\ \lambda_2\in \mathbf{R}$.
   Por tanto, si un vector está en el espacio generado por $\{e_1,e_2\}$, tiene como tercera componente el cero y se encuentra en el plano xy.
2. Consideremos en $\mathbf{R}^3$ el subespacio generado por el conjunto de vectores $\{a_1,a_2\}$, donde
   $a_1=(1,0,1),\ a_2=(0,1,1)$.
   Este subespacio está formado por los vectores que se pueden expresar en la forma
   $v=\lambda_1 a_1+\lambda_2 a_2=(\lambda_1,0,\lambda_1)+(0,\lambda_2,\lambda_2)$
   $=(\lambda_1,\lambda_2,\lambda_1+\lambda_2),\lambda_1,\ \lambda_2\in \mathbf{R}$
   Si analizamos esta expresión, vemos que la podemos expresar de la forma siguiente:
   Un vector de $\mathbf{R}^3$ de componentes (x,y,z) está en el subespacio generado por $\{a_1,\ a_2\}$ si y solo si el vector cumple
   $z=x+y$.
   Esta es la ecuación del plano determinado por los vectores a_1 y a_2.
3. En $\mathbf{R}^4$ consideremos el subespacio S, formado por las soluciones del sistema de ecuaciones
   $$\begin{aligned} x-2y+z&=0\\ -2x+3y-t&=0\\ y-2z+t&=0 \end{aligned}$$
   La solución de este sistema de ecuaciones está formada por los vectores de $\mathbf{R}^4$ de la forma:
   $(\alpha,\beta,-\alpha+2\beta,\ -2\alpha+3\beta),\ \alpha,\beta\in \mathbf{R}$
   A partir de esta expresión de los elementos de S, podemos afirmar que si $v\in S$, entonces
   $v=\alpha(1,0,-1,-2)+\beta(0,1,2,3),\ \alpha,\beta\in \mathbf{R}$.
   Por tanto, todo vector de S se puede escribir como combinación lineal de $(1,0,-1,-2)$ y $(0,1,2,3)$ y S es el subespacio generado por estos dos vectores.

Si el conjunto de vectores fuese un conjunto infinito $(a_i)_{i\in I}$, no se altera lo que hemos planteado, solamente hay que sustituir las combinaciones lineales finitas por combinaciones infinitas.

Para denotar el subespacio de E generado por un conjunto de vectores $\{a_1,a_2,\ldots,a_n\}$ emplearemos las notaciones:
$L(a_1,a_2,\ldots,a_n)$ o $\langle a_1,a_2,\ldots,a_n\rangle$.

1.8. Subespacios de K^n y sistemas de ecuaciones

Hemos visto que dada una matriz A de orden $m \times n$, el conjunto de las soluciones del sistema de ecuaciones $AX=0$ es un subespacio vectorial de K^n. Vimos también que dado un conjunto finito de vectores $\{a_1, a_2, \dots, a_k\}$ de K^n, existe un menor subespacio de K^n que contiene a los vectores $a_1, a_2, \dots, a_k$, que llamamos subespacio generado por $\{a_1 a_2, \dots, a_k\}$ y que es exactamente el subespacio formado por todas las combinaciones lineales de los vectores $a_1 a_2, \dots, a_k$. Más adelante veremos que, en definitiva, para todo subespacio vectorial S de K^n, existe siempre un conjunto finito de vectores de S, $\{b_1, b_2, \dots, b_k\}$, tal que

$$S = L(b_1, b_2, \dots, b_k) = \langle b_1, b_2, \dots, b_k \rangle$$

A partir de estas consideraciones, en los problemas que plantearemos, los subespacios vectoriales de K^n están dados en dos formas fundamentales:

Como solución de un sistema de ecuaciones.

Como el subespacio generado por un conjunto finito de vectores.

Sin embargo, dado un subespacio vectorial S de K^n, siempre podemos encontrar un sistema de ecuaciones homogéneo cuya solución sea S. Para ello, según lo que hemos planteado, tenemos que partir de un subespacio S de K^n, dado como el subespacio generado por un conjunto finito de vectores,

$$S = L(a_1, a_2, \dots, a_k),$$

y encontrar un sistema de ecuaciones que tenga por solución a S. Veamos mediante un ejemplo, cómo se trabaja en la práctica.

Ejemplo

Consideremos el subespacio de K^5 generado por los vectores:

$a_1 = (1,0,1,0,1)$, $a_2 = (0,1,0,1,0)$,
$a_3 = (-1,0,1,0,-1)$, $a_4 = (0,2,2,2,0)$.

Sea $S = L\ (a_1, a_2, a_3, a_4)$. Si $x = (x_1, x_2, x_3, x_4, x_5) \in K^5$ es tal que $x \in S$, entonces existen escalares $\alpha_1, \alpha_2, \alpha_3, \alpha_4$ tales que

$$x = \alpha_1 a_1 + \alpha_2 a_2 + \alpha_3 a_3 + \alpha_4 a_4, \quad \alpha_1, \alpha_2, \alpha_3, \alpha_4 \in K.$$

Desarrollemos estas combinaciones lineales

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \alpha_1 \begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix} + \alpha_2 \begin{pmatrix} 0 \\ 1 \\ 0 \\ 1 \\ 0 \end{pmatrix} + \alpha_3 \begin{pmatrix} -1 \\ 0 \\ 1 \\ 0 \\ -1 \end{pmatrix} + \alpha_4 \begin{pmatrix} 0 \\ 2 \\ 2 \\ 2 \\ 0 \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} \alpha_1 - \alpha_3 \\ \alpha_2 + 2\alpha_4 \\ \alpha_1 + \alpha_3 + 2\alpha_4 \\ \alpha_2 + 2\alpha_4 \\ \alpha_1 - \alpha_3 \end{pmatrix}$$

Obtenemos el sistema de ecuaciones, con las incógnitas a_1, a_2, a_3, a_4:

$$\begin{aligned} a_1 - a_3 &= x_1 \\ a_2 + 2a_4 &= x_2 \\ a_1 + a_3 + 2a_4 &= x_3 \\ a_2 + 2a_4 &= x_4 \\ a_1 - a_3 &= x_5 \end{aligned}$$

El problema se interpreta de la manera siguiente:

El vector $x \in K^5$ pertenece al subespacio S si y solo si existen escalares $\alpha_1, a_2, a_3, a_4 \in K$ tales que

$$x = a_1 a_1 + a_2 a_2 + a_3 a_3 + a_4 a_4,$$

y esto equivale a afirmar que el sistema de ecuaciones obtenido tiene que ser compatible.

Para obtener el sistema homogéneo que buscamos, aplicamos el método de Gauss a este sistema. Así llegamos al sistema equivalente:

$$\begin{cases} a_1 - a_3 = x_1 \\ a_2 + 2a_4 = x_2 \\ 2a_3 + 2a_4 = x_3 - x_1 \\ 0 = x_4 - x_2 \\ 0 = x_5 - x_1 \end{cases}$$

y la condición de compatibilidad del sistema está dada por:

$$\begin{cases} x_2 - x_4 = 0 \\ x_1 - x_5 = 0 \end{cases}$$

Tenemos entonces las equivalencias siguientes:

$x \in S \iff \{$Existen $a_1, a_2, a_3, a_4 \in K$ tales que $x = a_1 a_1 + a_2 a_2 + a_3 a_3 + a_4 a_4\}$

$\iff \{$El sistema

$$\begin{cases} a_1 - a_3 = x_1 \\ a_2 + 2a_4 = x_2 \\ a_1 + a_3 + 2a_4 = x_3 \\ a_2 + 2a_4 = x_4 \\ a_1 - a_3 = x_5 \end{cases}$$

es compatible.$\}$

$\iff \{$Las componentes $x = (x_1, x_2, x_3, x_4, x_5)$ satisfacen al sistema homogéneo

$$\left. \begin{cases} x_2 - x_4 = 0 \\ x_1 - x_5 = 0 \end{cases} \right\}.$$

En este ejemplo partimos de un subespacio de K^5, dado como el subespacio generado por un conjunto finito de vectores, y llegamos al sistema de ecuaciones homogéneas, del cual S es la solución.

Vamos a demostrar en general este resultado. Para ello emplearemos varios conceptos que pueden encontrarse en el libro de Kurosh de Álgebra Superior.

Necesitaremos específicamente dos teoremas:

Teorema del rango

Sea A una matriz con coeficientes en K. El orden superior de los menores de A diferentes de cero, es igual al rango de la matriz.

Teorema de Kronecker-Capelli

Un sistema de ecuaciones lineales $AX=B$ es compatible si y solo si el rango de la matriz del sistema es igual al rango de la matriz ampliada.

Enunciemos el resultado fundamental.

TEOREMA 1.2

Sea S un subespacio vectorial de K^n tal que existen vectores $a_1, a_2, \dots, a_r$ de K^n tales que $S=L\ (a_1, a_2, \dots, a_r)$. Entonces existe un sistema de ecuaciones homogéneo tal que S es el subespacio solución del sistema.

Demostración

Sea $x \in K^n, x=(x_1, x_2, \dots, x_n)$. Entonces x pertenece al subespacio S si y solo si existen escalares $\alpha_1, \alpha_2, \dots, \alpha_r \in K$ tales que

$$x = \alpha_1\ a_1 + \alpha_2 a_2 + \dots + \alpha_r\ a_r.$$

Vamos a proceder como en el ejemplo, construyendo un sistema de ecuaciones no homogéneo que tiene que ser compatible. Para ello desarrollamos la expresión dada considerando los componentes de cada vector: $a_1, a_2, \dots, a_r$.

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \alpha_1 \begin{pmatrix} a_{11} \\ a_{21} \\ \cdot \\ \cdot \\ \cdot \\ a_{n1} \end{pmatrix} + \alpha_2 \begin{pmatrix} a_{12} \\ a_{22} \\ \cdot \\ \cdot \\ \cdot \\ a_{n2} \end{pmatrix} + \dots + \alpha_r \begin{pmatrix} a_{1r} \\ a_{2r} \\ \cdot \\ \cdot \\ \cdot \\ a_{nr} \end{pmatrix}$$

$$\begin{pmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} = \begin{pmatrix} a_{11}\alpha_1 + a_{12}\ \alpha_2 + \dots + a_{1r}\ \alpha_r \\ a_{21}\alpha_1 + a_{22}\ \alpha_2 + \dots + a_{2r}\ \alpha_r \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ \cdot \quad \cdot \quad \cdot \\ a_{n1}\alpha_1 + a_{n2}\ \alpha_2 + \dots + a_{nr}\alpha_r \end{pmatrix}$$

$$\begin{aligned} a_{11}\ \alpha_1 + a_{12}\ \alpha_2 + \dots + a_{1r}\ \alpha_r &= x_1 \\ a_{21}\ \alpha_1 + a_{22}\ \alpha_2 + \dots + a_{2r}\ \alpha_r &= x_2 \\ \dots\dots\dots\dots\dots\dots&\dots\dots \\ a_{n1}\ \alpha_1 + a_{n2}\ \alpha_2 + \dots + a_{nr}\alpha_r &= x_r \end{aligned}$$

Escribiremos este sistema así:

$Aa = x.$

La condición de que $x \in S$ equivale a que este sistema es compatible. Por el teorema de Kronecker-Capelli, el sistema es compatible si y solo si el rango de la matriz del sistema

$$A = \begin{pmatrix} a_{11} & a_{12} \cdots & a_{1r} \\ a_{21} & a_{22} \cdots & a_{2r} \\ \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} \cdots & a_{nr} \end{pmatrix}$$

es igual al rango de la matriz ampliada:

$$\overline{A} = \begin{pmatrix} a_{11} & a_{12} \cdots & a_{1r} & x_1 \\ a_{21} & a_{22} \cdots & a_{2r} & x_2 \\ \cdots & \cdots & \cdots & \cdots \\ a_{n1} & a_{n2} \cdots & a_{nr} & x_n \end{pmatrix}$$

Supongamos que el rango de A es k. Esto quiere decir que existe un menor de orden k, M_k tal que $M_k \neq 0$, y que cualquier menor de orden $k+1$ que contiene a M_k es nulo.

Para facilitar la notación tomemos:

$$M_k = \begin{vmatrix} a_{11} & a_{12} \cdots & a_{1k} \\ a_{21} & a_{22} \cdots & a_{2k} \\ \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} \cdots & a_{kk} \end{vmatrix}$$

Entonces, en A el menor M_k es no nulo, y como queremos que

rango A = rango $\overline{A}$,

todo menor de orden $k+1$ que pueda construirse a partir de M_k tiene que ser nulo.

Para formar menores de orden $k+1$, bordearemos a M_k con los elementos de otra fila y otra columna. No es necesario hacerlo con columnas de la matriz A, pues ya sabemos que cualquier menor de orden $k+1$ así formado será nulo (recordemos que rango $A = k$). Por tanto, solo nos interesan los menores de orden $k+1$ que se pueden formar añadiendo la columna que contiene los términos independientes y una fila cualquiera, es decir, menores de la forma

$$M^s = \begin{vmatrix} a_{11} & a_{12} \cdots & a_{1k} & x_1 \\ a_{21} & a_{22} \cdots & a_{2k} & x_2 \\ \cdots & \cdots & \cdots & \cdots \\ a_{k1} & a_{k2} \cdots & a_{kk} & x_k \\ a_{s1} & a_{s2} \cdots & a_{sk} & x_s \end{vmatrix} \qquad \text{con } k \leqslant s \leqslant n.$$

De ese modo tenemos $(n-k)$ menores de orden $k+1$ que tienen que satisfacer

$M^s=0, \ k \leqslant s \leqslant n.$

Según la definición de determinante de una matriz, M^s es una expresión lineal en las variables $x_1, x_2, \ldots, x_k, x_s$, y así obtenemos $(n-k)$ ecuaciones lineales que forman un sistema de ecuaciones lineales homogéneo:

$$
\begin{aligned}
&M^{k+1}=0\\
&M^{k+2}=0\\
&\cdots\cdots\cdots\cdots\\
&\ M^{n}=0
\end{aligned}
$$

que tiene $(n-k)$ ecuaciones y n incógnitas: $x_1, x_2, \ldots, x_n$.

De este modo tenemos las equivalencias siguientes:

$$
\begin{aligned}
x \in S &\Longleftrightarrow \{\text{existen } \alpha_1, \alpha_2, \ldots, \alpha_r \in K \text{ tales que } x=\alpha_1\, a_1+\alpha_2\, a_2+\ldots+\alpha_r\, a_r\}\\
&\Longleftrightarrow \{\text{el sistema } A\alpha=x \text{ es compatible}\}\\
&\Longleftrightarrow \{\text{Rango } A=\text{rango } \bar{A}\}\\
&\Longleftrightarrow \{M^s=0 \text{ para } k \leqslant s \leqslant n\}.
\end{aligned}
$$

$M^s=0, \ k \leqslant s \leqslant n$ es un sistema de ecuaciones lineales homogéneo de $(n-k)$ ecuaciones y n incógnitas, cuya solución está formada por los vectores $x \in S$.

Hemos encontrado el sistema de ecuaciones lineales homogéneo, del cual el subespacio S es la solución.

Como conclusión tenemos:

Todo subespacio vectorial de K^n se puede expresar como el espacio solución de un sistema de ecuaciones lineales homogéneo.

El método práctico para buscar el sistema de ecuaciones es el siguiente:

1ro. Se expresa un vector genérico $v=(x_1, x_2, \ldots, x_n)$ como combinación lineal de los vectores que generan el subespacio.

2do. A partir de esa combinación lineal, se llega a un sistema de ecuaciones lineales no homogéneo, que es compatible solamente si v pertenece al espacio.

3ro. Se le aplica el método de Gauss al sistema obtenido y se establecen condiciones para que el sistema sea compatible.

4to. Al establecer las condiciones, se obtiene un sistema de ecuaciones lineales homogéneo, cuya solución es el espacio buscado.

Ejemplo

Sea F el subespacio de $\mathbb{R}^4$ generado por los vectores:

$a_1=(1,0,-1,0), \ a_2=(0, \ 1,-1, \ 0), \ a_3=(0,1,-2,1), \ a_4=(0,2-3,0).$

Busquemos el sistema de ecuaciones del cual F es el subespacio solución.

Planteamos el sistema, cuya matriz ampliada es de la forma

$$\begin{pmatrix} 1 & 0 & 0 & 0 & x \\ 0 & 1 & 1 & 2 & y \\ -1 & -1 & -2 & -3 & z \\ 0 & 0 & 1 & 0 & t \end{pmatrix}$$

y por el método de Gauss lo reducimos a la forma escalonada equivalente:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & x \\ 0 & 1 & 1 & 2 & y \\ 0 & 0 & 1 & 0 & t \\ 0 & 0 & 0 & -1 & x+y+z+t \end{pmatrix}$$

Aquí podemos ver que cualesquiera sean los valores que tomen x, y, z, t, el sistema es siempre compatible. Esto significa que todo vector de $\mathbb{R}^4$ se puede escribir como una combinación lineal de los vectores (a_1, a_2, a_3, a_4) y, por tanto, $F = \mathbb{R}^4$.

Por lo general, es más fácil identificar un subespacio de K^n a partir del sistema de ecuaciones lineales homogéneo del cual el subespacio es la solución. Esto se debe a que si tenemos un sistema de ecuaciones $AX=0$, al hallar la solución del sistema, los vectores del subespacio solución se obtienen a partir de relaciones entre las componentes de este. Así, en el ejemplo presentado al inicio de este epígrafe el subespacio de K^5 generado por los vectores (a_1, a_2, a_3, a_4) satisfacía la relación

$$(x_1, x_2, x_3, x_4, x_5) \in S \iff x_2 - x_4 = 0,\ x_1 - x_5 = 0.$$

De aquí concluimos que

$$S = \{(\alpha, \beta, \gamma, \beta, \alpha) / \alpha, \beta, \gamma \in K\}.$$

Por lo general, es mucho más fácil analizar el subespacio a partir del sistema de ecuaciones del cual él es la solución. Por ejemplo, en $\mathbb{R}^3$, si un subespacio es la solución de un sistema de ecuaciones que tiene dos ecuaciones no proporcionales, el subespacio es la *recta* determinada por las dos ecuaciones, y si el sistema tiene una sola ecuación, el subespacio es el *plano* que satisface la ecuación.

Que sea más fácil trabajar con el subespacio a partir de un sistema de ecuaciones, no quiere decir que un subespacio dado a partir de un sistema de vectores que lo genera no esté completamente determinado.

1.9 Operaciones con espacios y subespacios vectoriales

Como un espacio vectorial es un conjunto, que además posee otras propiedades, podemos estudiar algunas de las operaciones que se realizan con conjuntos: el producto, la intersección y la unión de conjuntos, y su relación con la estructura de espacio vectorial. Sobre la base de estas operaciones podemos construir nuevos ejemplos a partir de los ya conocidos.

4 mes/ 15 D

Producto de espacios

A partir de los ejemplos de $\mathbb{R}^n$ y $\mathbb{R}$, podemos obtener una idea bastante exacta de lo que vamos a hacer; por esto invitamos al estudiante a meditar un poco sobre la forma de definir la estructura vectorial en $\mathbb{R}^n$, y cómo en definitiva esta estructura está definida a partir de la estructura existente en $\mathbb{R}$. En particular, en $\mathbb{R}^n$ definimos las operaciones componente a componente. Esta es la idea fundamental para dar la construcción siguiente:

Sean E y F dos espacios vectoriales. Consideremos el producto cartesiano $E \times F = \{(x, y)/x \in E,\ y \in F\}$, conjuntamente con las operaciones:

Suma

Para todo $(x,\ y) \in E \times F$, $(u, v) \in E \times F$,
$(x, y) + (u, v) = (x+u, y+v) \in E \times F$.

Producto por un escalar

Para todo $(x, y) \in E \times F$, $\lambda \in K$,

$\lambda(x, y) = (\lambda x, \lambda y) \in E \times F$.

Esto es, para adicionar elementos de $E \times F$ sumamos componente a componente, y para multiplicar elementos de $E \times F$ por un escalar multiplicamos cada una de los componentes por el escalar dado.

No es difícil comprobar, aunque en ocasiones es trabajoso, que el conjunto $E \times F$ con la suma y el producto por un escalar que hemos definido, cumple con los axiomas que definen un espacio vectorial. Por ejemplo, el elemento neutro de la suma es el $(0,0)$, formado por el cero de E y el cero de F; cada elemento $(x, y) \in E \times F$ tiene un opuesto que es $(-x, -y)$. Las demás propiedades se cumplen porque son válidas para cada una de las operaciones entre las componentes. Proponemos al estudiante, como ejercicio, comprobar que se cumple una de las propiedades.

DEFINICIÓN 1.6

Dados dos espacios vectoriales E y F sobre K, el conjunto $E \times F$ con las operaciones suma y producto por un escalar es un espacio vectorial sobre K, se denomina el *espacio producto* de E y F.

Ejemplo

A partir de $\mathbb{R}$ podemos formar $\mathbb{R}^2 = \mathbb{R} \times \mathbb{R}$, y la estructura que damos a $\mathbb{R}^2$ como $\mathbb{R}$-espacio es la estructura producto de $\mathbb{R}$ por $\mathbb{R}$.

Nota. Es posible definir *otras* estructuras vectoriales sobre el conjunto $E \times F$, pero ya no serán la estructura producto. Así, $\mathbb{R}^2$ puede ser dotado de otra estructura que conocemos como $\mathbb{C}$, y en este caso es otra estructura vectorial en el conjunto $\mathbb{R} \times \mathbb{R}$, o sea, otro espacio vectorial.

Es evidente que lo que podamos hacer para dos espacios lo podemos extender para el producto de un conjunto finito de espacios vectoriales

Don. Yuniel Octavio Garcés Suárez

$E_1, E_2, \ldots, E_m$ y definir el espacio producto

$$E_1 \times E_2 \times \ldots \times E_m$$

con las operaciones definidas componente a componente.

¿Podemos extender esto al producto de una familia infinita de espacios? Dejamos al estudiante responder esta pregunta.

Intersección de subespacios

Sean E un espacio vectorial, E_1 y E_2 dos subespacios de E. Podemos considerar el subconjunto de E dado por $E_1 \cap E_2$ y preguntarnos si este es un subespacio de E.

La respuesta es afirmativa y fácil de comprobar a partir de la caracterización de un subespacio de un espacio vectorial y de la definición de intersección de conjuntos. Veamos.

$E_1 \cap E_2 \neq \varphi$ pues $0 \in E_1$ y $0 \in E_2$.

Sean $x, y \in E_1 \cap E_2$; $\alpha, \beta \in K$. Hay que probar que

$\alpha x + \beta y \in E_1 \cap E_2$.
$x \in E_1 \cap E_2 \Longleftrightarrow x \in E_1$ y $x \in E_2$
$y \in E_1 \cap E_2 \Longleftrightarrow y \in E_1$ y $y \in E_2$

Como E_1 y E_2 son subespacios, entonces

$\alpha x + \beta y \in E_1$ y $\alpha x + \beta y \in E_2$.

Por tanto,

$\alpha x + \beta y \in E_1 \cap E_2$,

lo cual demuestra que $E_1 \cap E_2$ es un subespacio de E.

Esto puede plantearse para la intersección de una familia cualquiera de subespacios. En la demostración seguiremos los mismos pasos que en la realizada para dos subespacios. Recordemos la definición de intersección:

Si $\{A_i\}_{i \in I}$ *es una familia de subconjuntos de* A, *entonces*

$x \in \bigcap_{i \in I} A_i \Longleftrightarrow x \in A_i$ *para todo* $i \in I$.

PROPOSICIÓN 1.2

Sean E un espacio vectorial sobre K, $\{E_i\}_{i \in I}$ una familia de subespacios de E. Entonces $\bigcap_{i \in I} E_i$ es un subespacio de E

Demostración

Sea $E' = \bigcap_{i \in I} E_i$. Empleando la caracterización de los subespacios, hay que comprobar que E' es no vacío y que si $x, y \in E'$, $\alpha, \beta \in K$, entonces $\alpha x + \beta y \in E'$.

Es evidente que E' es no vacío pues el cero de E está en todos y cada uno de los E_i, ya que son subespacios, y por tanto, está en la intersección E'.

Sea ahora $x, y \in E'$. Esto quiere decir que

$x \in E_i$ para todo $i \in I$,
$y \in E_i$ para todo $i \in I$.

Como cada E_i es un subespacio de E, entonces para cada $i \in I$, $\alpha x + \beta y \in E_i$. Pero si para cada $i \in I$, $\alpha x + \beta y \in E_i$, entonces por la definición de intersección,

$$\alpha x + \beta y \in E'$$

y E' es un subespacio de E.

Ejemplo

En el espacio $\mathbb{R}^5$ consideremos los subespacios S_1, S_2.

S_1 es el subespacio solución del sistema:

$$\begin{cases} 2x - y + \dfrac{1}{3} z - t = 0 \\ x + \dfrac{2}{3} z - u = 0 \\ 9x - 3y + 6z - 3t - 3u = 0 \end{cases}$$

S_2 es el subespacio solución del sistema:

$$\begin{cases} x + y - z + 7t - 3u = 0 \\ x + 2z + 3t - u = 0 \\ 2x - y + 7z + 2t = 0 \end{cases}$$

Resolviendo estos dos sistemas obtenemos:

$S_1 = \{(\alpha, 2\alpha - \beta, 0, \beta, \alpha) / \alpha, \beta \in \mathbb{R}\}$
$S_2 = \{(\gamma - 2\delta, 2\gamma + 3\delta, \delta, 0, \gamma) / \gamma, \delta \in \mathbb{R}\}$

Comparando la expresión de S_1, con la de S_2, resulta:

$$S_1 \cap S_2 = \{(\alpha, 2\alpha, 0, 0, \alpha) / \alpha \in \mathbb{R}\}.$$

Pero podemos obtener una vía más rápida para hallar $S_1 \cap S_2$ a partir del razonamiento siguiente:

Si un vector $v \in \mathbb{R}^5$ pertenece a $S_1 \cap S_2$, esto quiere decir que sus componentes satisfacen el sistema de ecuaciones correspondientes a S_1, y lo mismo sucede en relación con S_2. Por tanto, si $v \in S_1 \cap S_2$, sus componentes tienen que satisfacer a la vez los sistemas de ecuaciones que determinan a S_1 y S_2. Luego, los componentes de v tienen que satisfacer un sistema de ecuaciones formado por las ecuaciones que determinan a S_1 y las que determinan a S_2.

En nuestro ejemplo, $S_1 \cap S_2$ es el subespacio solución del sistema:

$$\begin{cases} 2x-y+\dfrac{1}{3}z-t=0 \\ x+\dfrac{2}{3}z-u=0 \\ 9x-3y+6z-3t-3u=0 \\ x+y-z+7t-3u=0 \\ x+2z+3t-u=0 \\ 2x-y+7z+2t=0 \end{cases}$$

Para obtener la expresión de los elementos de $S_1 \cap S_2$ resolvemos este sistema.

Una aplicación de la intersección de subespacios permite definir el subespacio generado por un conjunto cualquiera de vectores.

DEFINICIÓN 1.7

Sean E un espacio vectorial sobre K, A un subconjunto de E. Se denomina *subespacio generado por A* y se denota por $\langle A\rangle$, al menor subespacio de E que contiene a A.

Vamos a demostrar que para todo subconjunto de un espacio vectorial, siempre existe el subespacio generado por el subconjunto.

TEOREMA 1.3

Sean E un espacio vectorial sobre K, A un subconjunto de E. Entonces siempre existe el subespacio generado por A.

Demostración

Consideremos la familia de subespacios de $E, \{E_i\}_{i\in I}$ tales que $E_i \supset A$ para todo $i \in I$. Esta familia es no vacía pues como $E \supset A$, el espacio completo, E, siempre está en la familia. Probemos que, entonces, el subespacio

$$E' = \bigcap_{i\in I} E_i$$

es el menor subespacio de E que contiene a A.

En primer lugar, $E' \supset A$ pues para todo $i \in I$, $E_i \supset A$. En segundo lugar, E' es un subespacio pues es la intersección de una familia de subespacios. Además E' es el menor de todos los subespacios que contienen a A pues para todo $i \in I$, $E' \subset E_i$, y si existe otro subespacio F de E tal que $F \supset A$ y $F \subset E'$, entonces como F es un subespacio de E que contiene a A, existe un E_i de la familia tal que $F = E_i$, y tenemos:

$$F \subset E' \text{ y } E' \subset E_i = F.$$

Por tanto, $F = E'$.

Con esto demostramos la existencia de $\langle A\rangle$ para cualquier subconjunto A de un espacio vectorial E.

Sin embargo, el teorema 1.3 no dice cómo son los elementos de $\langle A\rangle$. Para saberlo debemos remitirnos al lema 1.2 y enunciarlo como sigue:

LEMA 1.1

Sean E un espacio vectorial sobre K, A un subconjunto no vacío de E. Entonces

$$\langle A\rangle=\{x\in E/x=\lambda_1a_1+\lambda_2a_2+\ldots+\lambda_na_n;\ a_1,a_2,\ldots,a_n\in A;\lambda_1,\lambda_2,\ldots,\lambda_n\in K\}.$$

No vamos a hacer la demostración de este lema pues es muy similar a la anterior; esta consiste en probar que el conjunto formado por las combinaciones lineales de elementos de A es un subespacio de E y contiene a A, y después observar que todo otro espacio de E que contenga a A contiene a este subespacio. Le proponemos al estudiante que haga esta demostración. Para el caso en que el subconjunto que tomamos de E sea el vacío, el subespacio generado por él es $\{0\}$:

$$\langle\phi\rangle=\{0\}.$$

Suma de subespacios

Al inicio de este epígrafe planteamos estudiar las operaciones: producto, intersección y unión. Falta por analizar la unión. En este caso no tenemos un resultado similar al de los casos anteriores pues, en general, la unión de dos subespacios de un espacio vectorial *no es* un subespacio. Veamos un ejemplo en $\mathbb{R}^3$.

Ejemplo

En $\mathbb{R}^3$ consideremos dos vectores no proporcionales:

$$u=(1,\ 0,\ 1)$$
$$v=(0,-1,0)$$

y sean:

$$\mathbb{R}u=\{\lambda u/\lambda\in\mathbb{R}\}=\{(\lambda,0,\lambda)/\lambda\in\mathbb{R}\}$$
$$\mathbb{R}v=\{\mu v/\mu\in\mathbb{R}\}=\{(0,-\mu,\ 0)/\mu\in\mathbb{R}\}$$

Sea $F=\mathbb{R}u\cup\mathbb{R}v$. Si F es un subespacio de $\mathbb{R}^3$, entonces cumple la condición

$$x,y\in F,\ \alpha,\beta\in\mathbb{R}\Rightarrow\alpha x+\beta y\in F.$$

En particular, como $u\in F$ y $v\in F$, para todo $\alpha,\beta\in\mathbb{R}$ tendría que cumplirse $\alpha u+\beta v\in F$; pero el vector

$(1,-1,1)=u+v$ no pertenece a F.

A partir de este ejemplo podemos concluir que, en general, la unión de dos subespacios vectoriales no es un subespacio vectorial. El problema básico es que si F y G son dos subespacios de E, $F\cup G$ contiene a F y a G, pero no tiene por qué contener a las combinaciones de los elementos de F con los de G. Es por esto que damos la definición siguiente:

DEFINICIÓN 1.8

Sean E un espacio vectorial sobre K, F y G dos subespacios de E. Se denomina *suma* de los subespacios F y G y se denota por $F+G$, al subespacio de E generado por $F\cup G$. Es decir,

$$F+G=\langle F\cup G\rangle.$$

Denominar *suma* de los subespacios al subespacio generado por la unión, está justificado por la proposición siguiente:

PROPOSICIÓN 1.3

Sean E un espacio vectorial sobre K, F y G dos subespacios de E. Entonces

$$F+G=\{x+y/x\in F,\ y\in G\}.$$

Demostración

Hay que probar que el conjunto

$$H=\{x+y/x\in F,\ y\in G\}$$

es el subespacio generado por $F\cup G$, es decir, el menor subespacio de E que contiene a $F\cup G$.
Demostremos, en primer lugar, que H es un subespacio.

Para todo $h, h'\in H$;

$$h=x+y,\ h'=x'+y' \text{ con } x, x'\in F,\ y, y'\in G.$$

Y para todo par de escalares $\alpha, \beta\in K$,

$$\begin{aligned}\alpha h+\beta h' &= \alpha(x+y)+\beta(x'+y')\\ &= \alpha x+\alpha y+\beta x'+\beta y'\\ &= (\alpha x+\beta x')+(\alpha y+\beta y')\\ \alpha h+\beta h' &= x''+y''\end{aligned}$$

Como F y G son subespacios de E, entonces $x''\in F$ y $y''\in G$; por tanto,

$$\alpha h+\beta h'\in H.$$

Con esto demostramos que H es un subespacio.

En segundo lugar, H contiene a $F\cup G$ pues contiene a todos los elementos de la forma $x+0$, $x\in E$, y a los de la forma $0+y$, $y\in G$. Por tanto,

$$F\subset H,\ G\subset H\Rightarrow F\cup G\subset H.$$

Solamente falta probar que H es el menor subespacio que contiene a $F\cup G$; pero si observamos que cualquier otro subespacio que contenga a F y a G tiene que contener a los elementos de la forma $x+y, x\in F, y\in G$, y por tanto a H, concluimos que H es el menor subespacio que contiene a $F\cup G$. Luego

$$F+G=\{x+y/x\in F, y\in G\}.$$

Ejemplos

1. Calculemos la suma de los espacios del ejemplo anterior:
$Ku=\{(\lambda,0,\lambda)/\lambda\in \mathbb{R}\}$
$Kv=\{(0,-\mu,0,/\mu\in \mathbb{R}\}$
Según la proposición 1,3, tenemos:
$Ku+Kv=\{(\lambda\ 0,\lambda)+(0,-\mu,0)\,|\,\lambda\in \mathbb{R},\mu\in \mathbb{R}\}$
$=\{(\lambda,-\mu,\lambda)/\lambda,\mu\in \mathbb{R}\}$
$=\{(\lambda,\mu,\lambda)/\lambda,\ \mu\in \mathbb{R}\}$.
Este subespacio es el plano determinado por las rectas Ku y Kv.

2. Consideremos en $\mathbb{R}^3$ los subespacios:
$V=\{(x,y,z)/z=0\}$
$W=\{(x,y,z)/x+y-z=0,\ x+3y+z=0\}$
V es el plano xy y W es la recta $\{(2a,-a,a)/a\in \mathbb{R}\}$.
Todo elemento de $V+W$ se expresa como la suma de un elemento de V y un elemento de W.
Expresemos V y W en la misma forma:
$V=\{(\beta,\gamma,0)/\beta,\gamma\in \mathbb{R}\}$
$W=\{(2a,-a,a)/a\in \mathbb{R}\}$
Entonces,
$V+W=\{(\beta+2a,\gamma-a,a)/a,\beta,\gamma\in \mathbb{R}\}$
Podemos identificar este espacio $V+W$. Si partimos de la representación geométrica de V y W y de la definición de $V+W$, podemos tener una idea de quién es este subespacio. Estamos sumando un elemento cualquiera de un plano con un elemento cualquiera de una recta no contenida en el plano; esto sugiere la posibilidad de encontrar un elemento cualquiera de $\mathbb{R}^3$. Veamos si esto es posible.
Sea $(x,y,z)\in \mathbb{R}^3$ un elemento cualquiera de $\mathbb{R}^3$. Investiguemos si existen escalares $a,\beta,\gamma\in \mathbb{R}$ tales que
$(x,y,z)=(\beta,\gamma,0)+(2a,-a,a)$
$=(\beta+2a,\gamma-a,a)$.
De aquí obtenemos:
$x=\beta+2a,\ \ y=\gamma-a,\ \ z=a$
Por tanto,
$a=z$,
$\beta=x-2z,\ \ \gamma=y+z$.
Es decir, dado un vector cualquiera de $\mathbb{R}^3$, (x,y,z), los números a,β,γ se pueden calcular a partir de estas expresiones. Por tanto, el subespacio $V+W$ es el espacio $\mathbb{R}^3$.
Observemos, además, que los números a, β, γ son únicos para cada vector (x,y,z) de $\mathbb{R}^3$. En otras palabras: la expresión de los vectores de $V+W$ como la suma de un vector de V y un vector de W, es única.

3. En $\mathbb{R}^3$ consideremos los subespacios:
$V=\{(x,y,z)/z=0\}$
$V'=\{(x,y,z)/x+y-z=0\}$

Empleando la otra forma de expresarlos, tenemos:

$V=\{(\alpha,\beta,0)/\alpha,\beta\in \mathbb{R}\}$

$V'=\{(\gamma,\delta,\ \gamma+\delta)/\gamma,\delta\in \mathbb{R}\}$

De aquí,

$V+V'=\{(\alpha+\gamma,\ \beta+\delta,\ \gamma+\delta)/\alpha,\beta,\gamma,\delta\in \mathbb{R}\}$.

También en este caso podemos pensar que la suma es el espacio $\mathbb{R}^3$. Sea (x,y,z) un elemento cualquiera de $\mathbb{R}^3$. Veamos si existen escalares $\alpha,\beta,\gamma,\delta\in \mathbb{R}$ tales que

$$\begin{aligned}(x,y,z) &=(\alpha,\beta,0)+(\gamma,\delta,\gamma+\delta)\\ &=(\alpha+\gamma,\ \beta+\delta,\ \gamma+\delta).\end{aligned}$$

De aquí obtenemos el sistema:

$$\begin{cases} x=\alpha+\gamma\\ y=\beta+\delta\\ z=\gamma+\delta\end{cases}$$

que es un sistema compatible indeterminado. Tomando δ como variable independiente, obtenemos la solución:

$\alpha=x-z+\delta$

$\beta=y-\delta$

$\gamma=z-\delta$

Observemos que en este caso los valores de $\alpha,\ \beta,\gamma,\delta\in \mathbb{R}$ para cada vector $(x,y,z)\in \mathbb{R}^3$ *no* son únicos, pues para cada valor que demos a δ encontraremos valores para α,β,γ.

Como hemos visto en estos ejemplos, existe la posibilidad de que en una suma de subespacios $V+W$, la expresión de los elementos del espacio suma como un elemento de V más un elemento de W, no sea única. Veamos un teorema que caracteriza esta condición.

TEOREMA 1.4

Sean E un espacio vectorial sobre K, V y W dos subespacios de E. Entonces las dos condiciones siguientes son equivalentes:

(1) $V\cap W=\{0\}$

(2) Para todo elemento x de $V+W$, existen $v\in V$ y $w\in W$ únicos, tales que $x=v+w$.

Demostración

1) Probemos que (1) implica (2).

Supongamos que para $x\in V+W$ existen $v,v'\in V$ y $w,w'\in W$ tales que $v\neq v'$, $w\neq w'$ y

$x=v+w$,

$x=v'+w'$.

De aquí obtenemos:

$v+w=v'+w'$,

y reuniendo los elementos de V en un mismo miembro de la igualdad y los de W en el otro:
$v-v'=w'-w$.
Entonces, $v-v'\in V$ por ser V un subespacio y $w'-w\in W$ por ser W un subespacio. Además, $v-v'\neq 0$ y como $v-v'=w'-w$, esto significa que $v-v'\in W$; por tanto, $v-v'\in V\cap W$. Pero $V\cap W=\{0\}$, de donde $v-v'=0$ y $v=v'$, $w=w'$, lo que contradice que la expresión de x no es única.

2) Probemos ahora que (2) implica (1).
Si la expresión de todo elemento de $V+W$ como suma de un elemento de V y otro de W es única, en particular tenemos para $0\in V+W$:
$0=0+0$.
Sea $x\in V\cap W$. Entonces, $x\in V$ y $x\in W$,
y como V y W son subespacios, tenemos $-x\in V$ y $-x\in W$. Por tanto, podemos escribir:
$0=x+(-x)$.
Pero como la expresión del 0 como suma de elementos de V y W es única, y ya tenemos que
$0=0+0$,
entonces $x=0$ y, por tanto, $V\cap W=\{0\}$.

La condición (1) del teorema nos lleva a la definición siguiente:

DEFINICIÓN 1.9

Sean E un espacio vectorial sobre K, F y G dos subespacios de E. Se dice que la *suma* de F y G es *directa*, y se denota por $F\oplus G$ si y solo si $F\cap G=\{0\}$.

Lo anterior equivale a decir que todo elemento del espacio suma se expresa en forma única como la suma de un elemento de F y un elemento de G.

Ejemplos

1) Si consideramos en $\mathbb{R}^3$ los subespacios:
$Ku=\{(\lambda,0,\lambda)/\lambda\in\mathbb{R}\}$
$Kv=\{(0,-\mu,0)/\mu\in\mathbb{R}\}$
entonces $Ku\oplus Kv$.

2. En $\mathbb{R}^3$ consideremos los subespacios:
$V=\{(x,y,z)/z=0\}$
$W=\{(x,y,z)/x+y-z=0,\ x+3y+z=0\}$
Como $V\cap W=\{0\}$, tenemos $V\oplus W$.

3. En $\mathbb{R}^3$ tenemos los subespacios:
$V=\{(x,y,z)/z=0\}$
$V'=\{(x,y,z)/x+y-z=0\}$
La suma de V y V' no puede ser directa pues, como vimos anteriormente, la expresión de un elemento de la suma como un elemento de V más

otro de V' no es única; pero además, si calculamos la intersección, tenemos:

$V \cap V' = \{(\alpha, -\alpha, 0) / \alpha \in \mathbb{R}\} \neq \{0\}$.

En el caso en que la suma de los subespacios F y G sea el espacio total, E, y además la suma sea directa, diremos que E es la suma directa de los subespacios F y G, y escribiremos:

$$E = F \oplus G.$$

DEFINICIÓN 1.10

Sean E un espacio vectorial sobre K, F y G dos subespacios de E. Si $E = F \oplus G$, se dice que F y G son *subespacios suplementarios* con respecto a E.

Ejemplos

1. En $\mathbb{R}^3$:
   $V = \{(x, y, z) / z = 0\}$
   $W = \{(x, y, z) / x + y - z = 0,\ x + 3y + z = 0\}$ son suplementarios.
2. En $\mathbb{R}^3$:
   $V = \{(x, y, z) / z = 0\}$
   $W' = \{(x, y, z) / x = 0,\ y = 0\}$
   son también suplementarios.
   Para esto hay que comprobar que todo elemento de $\mathbb{R}^3$ se puede expresar en forma única como suma de un elemento de V y otro de W'. Calculemos primero $V + W'$.
   $V = \{(\alpha, \beta, 0) / \alpha,\ \beta \in \mathbb{R}\}$
   $W' = \{(0, 0, \gamma) / \gamma \in \mathbb{R}\}$.
   Por tanto, un elemento de $V + W'$ se expresa como
   $(\alpha, \beta, 0) + (0, 0, \gamma) = (\alpha, \beta, \gamma) \quad \alpha, \beta, \gamma \in \mathbb{R}$.
   De esta forma es evidente que
   $\mathbb{R}^3 = V \oplus W'$.

A partir de estos ejemplos podemos ver que un subespacio F de un espacio vectorial E puede tener varios subespacios suplementarios diferentes. Más adelante demostraremos que en un espacio vectorial E, todo subespacio siempre tiene suplementarios.

Nota. El concepto *subespacios suplementarios* depende también del espacio total E en que estemos trabajando. Así por ejemplo, los espacios vectoriales reales $\mathbb{R}[x]$ polinomios en una indeterminada, $C(\mathbb{R}, \mathbb{R})$ funciones continuas, $F(\mathbb{R}, \mathbb{R})$ funciones de $\mathbb{R}$ en $\mathbb{R}$, cumplen la relación

$\mathbb{R}[x] \subset C(\mathbb{R}, \mathbb{R}) \subset F(\mathbb{R}, \mathbb{R})$.

No es lo mismo considerar un suplementario de $\mathbb{R}[x]$ con respecto a $C(\mathbb{R}, \mathbb{R})$ que con respecto a $F(\mathbb{R}, \mathbb{R})$.

Si analizamos las posibilidades para la suma, tenemos:

1ra. Dados dos subespacios F y G, siempre se puede calcular $F + G$, donde $F + G = \langle F \cup G \rangle = \{x + y / x \in F,\ y \in G\}$.

2da. Si $F \cap G = \{0\}$, se dice que la suma de F y G es directa y se denota por $F \oplus G$. En este caso, la expresión de los elementos de $F \oplus G$ como suma de un elemento de F y uno de G, es única.

3ra. $E = F + G$. Esto quiere decir que todo elemento de E se puede escribir como la suma de un elemento de F y un elemento de G.

4ta. $E = F \oplus G$. En este caso todo elemento de E se puede expresar en forma única como la suma de un elemento de F y un elemento de G, lo que es equivalente a las dos condiciones siguientes:

(1) Todo elemento de E se escribe como suma de un elemento de F y un elemento de G.

(2) $F \cap G = \{0\}$

Es importante diferenciar las cuatro posibilidades.

La suma de dos subespacios puede ser directa o no. La suma de dos subespacios puede ser el espacio completo o no. Veamos algunos ejemplos más.

Ejemplos

1. En $\mathbb{R}^4$ sean:
   $F = \{(x, y, 0, 0) / x,\ y \in \mathbb{R}\}$
   $G = \{(0, y, z, 0) / y,\ z \in \mathbb{R}\}$
   Evidentemente, $F \cap G \neq \{0\}$; por tanto, su suma no es directa. Además, $F + G \neq \mathbb{R}^4$, pues si observamos que tanto los elementos de F como los de G tienen la cuarta componente nula, su suma no puede ser todo $\mathbb{R}^4$.

2. En $\mathbb{R}^4$ tomemos $u = (1, 0, -1, 0)$ y $v = (0, 1, 0, 1)$. En este caso para las rectas generadas por estos vectores, se cumple:
   $\mathbb{R}u \cap \mathbb{R}v = \{0\}$,
   y por tanto, $\mathbb{R}u \oplus \mathbb{R}v$, aunque también $\mathbb{R}u \oplus \mathbb{R}v \neq \mathbb{R}^4$.

3. En $\mathbb{R}^3$ consideremos los planos:
   $F = \{(\alpha, \beta, 0) /\ \alpha,\ \beta \in \mathbb{R}\}$
   $G = \{(0, \gamma, \delta) / \gamma\ , \delta \in \mathbb{R}\}$
   Es evidente que $F \cap G \neq \{0\}$ ya que es la recta generada por el vector $(0, 1, 0)$, y es fácil comprobar que $F + G = \mathbb{R}^3$.

4. Sea F el espacio de las funciones de $\mathbb{R}$ en $\mathbb{R}$. Vamos a considerar dos subespacios:
   $P = \{f \in \mathrm{F} / f(x) = f(-x)\ \forall x \in \mathbb{R}\}$
   $I = \{f \in \mathrm{F} / f(x) = -f\ (-x)\ \forall x \in \mathbb{R}\}$
   P es el subespacio de las funciones pares o simétricas e I el de las funciones impares o antisimétricas.
   Le proponemos al estudiante que compruebe que efectivamente, P e I son subespacios de F.
   Vamos a demostrar que $\mathrm{F} = P \oplus I$. Para ello tenemos que probar que:
   1) La suma de P e I es directa.
   2) La suma de P e I es todo el espacio F.

Para probar 1) lo más sencillo, por lo general, es comprobar si $P \cap I = \{0\}$. Supongamos que una función $f \in F$ está en $P \cap I$.

$f \in P \cap I \Longleftrightarrow f \in P$ y $f \in I$
$f \in P \Longleftrightarrow \forall x \in \mathbb{R},\ f(x) = f(-x)$
$f \in I \Longleftrightarrow \forall x \in \mathbb{R},\ f(x) = -f(-x)$

Por tanto, $f \in P \cap I \Longleftrightarrow \{-f(x) = f(-x)$ y $f(x) = f(-x)\ \forall x \in \mathbb{R}\}$

$$\begin{aligned} &\Rightarrow f(x) = -f(x) \quad \forall x \in \mathbb{R} \\ &\Rightarrow f(x) = 0 \quad \forall x \in \mathbb{R} \\ &\Longleftrightarrow f = 0 \end{aligned}$$

Por tanto, $P \cap I = \{0\}$.

Con esto hemos probado que la suma de P e I es directa. Falta probar que toda función se puede escribir como la suma de una función par y una función impar, o sea, que la suma de P e I es todo F. Sea $f \in F$ una función cualquiera. Definimos:

$$f_p(x) = \frac{f(x) + f(-x)}{2}$$

$$f_i(x) = \frac{f(x) - f(-x)}{2}$$

$f_p(x)$ es par y $f_i(x)$ es impar.

Se cumple:

$$\begin{aligned} f_p(x) + f_i(x) &= \frac{f(x) + f(-x)}{2} + \frac{f(x) - f(-x)}{2} \\ &= \frac{f(x) + f(-x) + f(x) - f(-x)}{2} \\ &= \frac{2f(x)}{2} \\ &= f(x), \end{aligned}$$

o sea,
$f(x) = f_p(x) + f_i(x)$.
Esto prueba que la suma de los subespacios P e I es el espacio F, y como ya probamos que $P \cap I = \{0\}$ concluimos que $F = P \oplus I$.

Las definiciones de suma y suma directa de dos subespacios pueden extenderse naturalmente para una familia finita de subespacios.

DEFINICIÓN 1.11

Sean E un espacio vectorial sobre K, $E_1, E_2, \ldots, E_n$ subespacios de E. El espacio $E_1 + E_2 + \ldots + E_n$ se define como:

$$\begin{aligned} E_1+E_2+\ldots+E_n &= \langle E_1\cup E_2\cup\ldots\cup E_n\rangle \\ &= \{x_1+x_2+\ldots+x_n/x_1\in E_1, x_2\in E_2,\ldots,x_n\in E_n\} \end{aligned}$$

Si, además, la expresión de cada vector del espacio suma como suma de elementos de $E_1, E_2, \ldots, E_n$ es única, se dice que la suma de estos subespacios es directa y se denota por

$$E_1\oplus E_2\oplus\ldots\oplus E_n$$

Ejemplos

1. En $\mathbb{R}^3$ consideremos las rectas generadas por los vectores $e_1=(1,0,0)$, $e_2=(0,1,0)$, $e_3=(0,0,1)$. Entonces
   $\mathbb{R}^3=\mathbb{R}e_1\oplus\mathbb{R}\,e_2\oplus\mathbb{R}\,e_3$.
2. En $K_n[x]$, consideremos los subespacios
   $K_p=\{ax^p/a\in K\}$, $p=0,1,2,\ldots,n-1$ constituidos, para cada valor de p, por los polinomios homogéneos de grado p.
   K_p es un subespacio de $K_n[x]$ para cada valor de $p(p<n)$. Además, se cumple:
   $K_n[x]=K_0\oplus K_1\oplus\ldots\oplus K_{n-1}$

1.10 Dependencia e independencia lineal

En el ejemplo 3 de combinación lineal (epígrafe 1.6) vimos que la combinación lineal $2(1,2,3,0)-1(0,-1,1,1)-1(2,5,5,-1)$, tiene como resultado el vector nulo. Para este caso vamos a denotar $a_1=(1,2,3,0)$, $a_2=(0,-1,1,1)$, $a_3=(2,5,5,-1)$, y por tanto,

$$2a_1+(-1)a_2+(-1)a_3=0.$$

En esta expresión despejamos el vector a_2 y obtenemos: $a_2=2a_1-a_3$, es decir, el vector a_2 se expresa como combinación lineal de los vectores a_1 y a_3. Podemos hacer lo mismo con a_1 y a_3. ¿Será siempre posible hacer esto? Veamos.

Si tomamos los vectores del ejemplo 1 de este epígrafe, e_1, e_2, e_3, y tratamos de expresar uno de ellos como combinación lineal de los otros comprobaremos que esto no es posible. Solamente haremos el cálculo de uno de los tres casos posibles; dejamos los demás como ejercicio para el estudiante. Probemos con e_2.

Si e_2 se escribe como combinación lineal de los vectores e_1 y e_3, esto quiere decir que existen números $\alpha,\beta\in\mathbb{R}$ tales que

$$e_2=\alpha e_1+\beta e_3.$$

Efectuemos la combinación lineal:

$$\begin{aligned} (0,1,0) &= \alpha(1,0,0)+\beta(0,0,1) \\ (0,1,0) &= (\alpha,0,0)+(0,0,\beta) \\ (0,1,0) &= (\alpha,0,\beta). \end{aligned}$$

Esto implica las ecuaciones

$\alpha=0,\ 1=0,\ \beta=0,$

lo cual es imposible.

Proponemos que el estudiante analice los otros dos casos y verá que en todos llega a una incompatibilidad.

Por tanto, dado un conjunto de vectores, no siempre es posible expresar uno de ellos como combinación lineal de los demás vectores. Por ello definimos el concepto siguiente:

DEFINICIÓN 1.12

Sean E un espacio vectorial sobre un campo K, $(a_1,a_2,\ldots,a_n)$ un sistema de vectores. Se dice que el sistema $(a_1,a_2,\ldots,a_n)$ es *linealmente dependiente* si uno de los vectores del sistema se puede expresar como combinación lineal del resto de los vectores del sistema.
Si ninguno de los vectores del sistema se puede expresar como combinación lineal del resto, decimos que el sistema de vectores es *linealmente independiente*.

De acuerdo con esta definición, dado un sistema de vectores hay dos posibilidades mutuamente excluyentes:

1ra. El sistema es linealmente dependiente.
2da. El sistema es linealmente independiente.

En el caso de sistemas infinitos de vectores, la definición es perfectamente válida.

Un sistema $(a_i)_{i\in I}$ es linealmente dependiente si uno de los vectores se puede expresar como combinación lineal de los demás.

Como veremos para el caso finito, la definición que hemos dado proporciona una idea muy clara sobre el caso de un sistema linealmente dependiente, pero como instrumento de trabajo para demostrar si un sistema es linealmente dependiente o no, esta definición no es eficiente, por lo que daremos una caracterización que resuelve las dificultades que suelen presentarse.

Ejemplos

1. El sistema de $\mathbf{R}^4$ formado por los vectores $((1,2,3,0),\ (0,-1,1,1),\ (2,5,5,-1))$ es linealmente dependiente.
2. El sistema de $\mathbf{R}^3$ formado por los vectores $e_1=(1,0,0)$, $e_2=(0,1,0)$, $e_3=(0,0,1)$ es linealmente independiente.
3. En el espacio $K_n[x]$ el sistema $(1,x,x^2,\ldots,x^{n-1})$ es linealmente independiente, pues cada uno de los vectores del sistema es un polinomio homogéneo y todos tienen los grados diferentes entre sí. Como es imposible escribir un polinomio homogéneo de grado k como suma de polinomios homogéneos de otros grados, entonces ninguno de los elementos del sistema se puede escribir como combinación lineal de los demás.

097 100 USD ⇒ 97 CUC

4. Si en un espacio vectorial tenemos un sistema formado por dos vectores (a_1,a_2), el sistema es linealmente dependiente si uno de ellos, por ejemplo a_2, se expresa como combinación lineal del otro, es decir, si existe un escalar $\lambda \in K$ tal que $a_2=\lambda a_1$. En otras palabras, a_1 y a_2 son proporcionales o colineales.

   Si el sistema es linealmente independiente, entonces los vectores a_1 y a_2 no son proporcionales.

5. Un sistema de vectores $(a_1,a_2,\ldots,a_n,0)$, que contiene al vector nulo, es linealmente dependiente, pues el vector nulo siempre se puede escribir como combinación lineal de los demás:

   $$0=0a_1+0a_2+\ldots+0a_n$$

PROPOSICIÓN 1.4

Sean E un espacio vectorial sobre K, $(a_1,a_2,\ldots,a_n)$ un sistema de vectores de E. Si existe un subsistema del sistema $(a_1,a_2,\ldots,a_k)$ que sea linealmente dependiente, entonces el sistema $(a_1,a_2,\ldots,a_n)$ es linealmente dependiente.

Demostración

Supongamos que el sistema $(a_1,a_2,\ldots,a_n)$ tiene un subsistema $(a_1,a_2,\ldots,a_k)$ $(k<n)$ que es linealmente dependiente. Esto quiere decir que uno de los vectores se puede expresar como combinación lineal de los demás; para simplificar, supongamos que sea el a_1. Entonces existen escalares $\lambda_2,\ldots,\lambda_k$, tales que

$$a_1=\lambda_2a_2+\lambda_3a_3+\ldots+\lambda_ka_k.$$

A partir de esta combinación lineal, podemos escribir

$$a_1=\lambda_2a_2+\lambda_3a_3+\ldots+\lambda_k\,a_k+0a_{k+1}+\ldots+0a_n$$

Esto significa que en el sistema $(a_1,a_2,\ldots,a_n)$ el vector a_1 se escribe como combinación lineal del resto de los vectores del sistema. Por tanto, $(a_1,a_2,\ldots a_n)$ es linealmente dependiente.

Nota. Aunque en la demostración hacemos algunas suposiciones sobre la posición que ocupa en el sistema el vector que es dependiente de los demás, esto no altera el resultado final, y sí complicaría la notación de la demostración considerar como dependiente un vector arbitrario.

Veamos mediante algunos ejemplos cómo se plantea el problema de conocer si un sistema de vectores es o no linealmente independiente.

Ejemplos

1. En $\mathbf{R}^4$ consideremos los vectores:

   $a_1=(0,3,2,1)$, $a_2=(1,2,3,0)$, $a_3=(-1,1,-1,1)$, $a_4=(1,-1,1,-1)$.

   Vamos a tratar de escribir a_1 como combinación lineal de a_2, a_3 y a_4. Para ello hay que encontrar escalares α, β, γ tales que

   $a_1=\alpha a_2+\beta a_3+\gamma a_4$.

   Trabajemos con los valores en columnas.

$$\begin{pmatrix}0\\3\\2\\1\end{pmatrix}=\alpha\begin{pmatrix}1\\2\\3\\0\end{pmatrix}+\beta\begin{pmatrix}-1\\1\\-1\\1\end{pmatrix}+\gamma\begin{pmatrix}1\\-1\\1\\-1\end{pmatrix}$$

$$\begin{pmatrix}0\\3\\2\\1\end{pmatrix}=\begin{pmatrix}\alpha\\2\alpha\\3\alpha\\0\end{pmatrix}+\begin{pmatrix}-\beta\\\beta\\-\beta\\\beta\end{pmatrix}+\begin{pmatrix}\gamma\\-\gamma\\\gamma\\-\gamma\end{pmatrix}$$

$$\begin{pmatrix}0\\3\\2\\1\end{pmatrix}=\begin{pmatrix}\alpha-\beta+\gamma\\2\alpha+\beta-\gamma\\3\alpha-\beta+\gamma\\\beta-\gamma\end{pmatrix}$$

De aquí tenemos que α, β y γ son las soluciones del sistema de ecuaciones:

$$\begin{cases}\alpha-\beta+\gamma=0\\2\alpha+\beta-\gamma=3\\3\alpha-\beta+\gamma=2\\\beta-\gamma=1\end{cases}$$

Aplicando el método de Gauss para resolverlo, el sistema se reduce al siguiente sistema equivalente:

$$\begin{cases}\alpha-\beta+\gamma=0\\\beta-\gamma=1\end{cases}$$

Por tanto, el sistema tiene solución ya que existen $\alpha,\beta,\gamma\in\mathbb{R}$ que satisfacen las condiciones pedidas. Valores particulares son los siguientes:
$\alpha=1,\ \beta=2,\ \gamma=1,$
y por tanto,
$a_1=a_2+2a_3+a_4.$
En este caso podíamos haber prescindido de los cálculos, pues una primera inspección de los vectores que forman el sistema nos hubiese indicado que $a_4=-a_3$ y, por tanto, el subsistema formado por los vectores (a_3,a_4) es linealmente dependiente, y por la proposición 1.4, el sistema mayor es también linealmente dependiente.

2. En $\mathbb{R}^3$ consideremos el sistema formado por los vectores $a_1=(1,1,1)$, $a_2=(1,-1,1)$, $a_3=(1,1,0)$.
Investiguemos si a_1 es combinación lineal de a_2 y a_3.
$a_1=\alpha a_2+\beta a_3.$

$$\begin{pmatrix}1\\1\\1\end{pmatrix}=\alpha\begin{pmatrix}1\\-1\\1\end{pmatrix}+\beta\begin{pmatrix}1\\1\\0\end{pmatrix}$$

$$\begin{pmatrix}1\\1\\1\end{pmatrix}=\begin{pmatrix}\alpha+\beta\\-\alpha+\beta\\\alpha\end{pmatrix}$$

Obtenemos el sistema de ecuaciones:

$$\begin{cases}\alpha+\beta=1\\-\alpha+\beta=1\\\alpha\quad=1\end{cases}$$

Mediante el método de Gauss este sistema se reduce al sistema escalonado equivalente:

$$\begin{cases}\alpha+\beta=1\\-\beta=0\\0=2\end{cases}$$

Por tanto, el sistema es incompatible y no existen escalares α, β que satisfagan las condiciones dadas lo que quiere decir que a_1 no se puede escribir como combinación lineal de a_2 y a_3. Sin embargo, eso no significa que el sistema formado por los vectores (a_1, a_2, a_3) sea linealmente independiente, pues hay que considerar otras dos posibilidades: a_2 es combinación lineal de a_1 y a_3, y a_3 es combinación lineal de a_1 y a_2. Analicemos la primera posibilidad.

$a_2=\alpha\ a_1+\beta a_3$.

Obtenemos el sistema:

$$\begin{cases}\alpha+\beta=\ \ 1\\\alpha+\beta=-1\\\alpha\quad=\ \ 1\end{cases}$$

Este sistema es incompatible. Por tanto, a_2 no se puede expresar como combinación lineal de a_1 y a_3.

Veamos la segunda posibilidad.

$a_3=\alpha a_1+\beta a_2$

Obtenemos el sistema:

$$\begin{cases}\alpha+\beta=1\\\alpha-\beta=1\\\alpha\quad=0\end{cases}$$

Este sistema también es incompatible. Por tanto, a_3 no se puede expresar como combinación lineal de a_1 y a_3.
Ahora sí podemos afirmar que el sistema formado por los vectores $a_1=(1,1,1)$, $a_2(1,-1,1)$, $a_3=(1,1,0)$ es linealmente independiente.

De estos ejemplos podemos sacar dos conclusiones:

1ra. Si en un espacio $\mathbb{R}^n$ (K^n) un vector se puede expresar como combinación lineal de otros, los coeficientes de la combinación lineal son las soluciones de un sistema de ecuaciones lineales. En este caso hay dos posibilidades; el sistema es compatible y el vector es combinación lineal de los otros o el sistema es incompatible y, por tanto, el vector dado no se puede expresar como combinación lineal de los otros vectores.

2da. En cuanto al método de trabajo empleado, es bastante difícil trabajar con la definición dada, pues si queremos probar que un sistema es linealmente dependiente, hay que escribir al menos uno de los vectores como combinación lineal de los restantes, para lo cual es preciso probarlos uno a uno. Por otra parte, para demostrar que un sistema es linealmente independiente, hay que probar que ninguno de los vectores se puede escribir como combinación lineal de los demás. Por tanto, investigar, de acuerdo con la definición dada, si un sistema es linealmente dependiente o linealmente independiente puede involucrar un gran volumen de cálculo. Por eso vamos a dar una caracterización equivalente.

TEOREMA 1.5

Sean E un espacio vectorial sobre K, $(a_1,a_2,\dots,a_n)$ un sistema de vectores. Entonces el sistema de vectores $(a_1,a_2,\dots,a_n)$ es linealmente independiente si y solo si la *única* combinación lineal del sistema que tiene como resultado el vector nulo, es la combinación lineal trivial, es decir, si se cumple:

$$\lambda_1 a_1+\lambda_2 a_2+\dots+\lambda_n a_n=0 \Rightarrow \lambda_1=\lambda_2=\dots=\lambda_n=0.$$

Demostración

Supongamos que el sistema $(a_1,a_2,\dots,a_n)$ es linealmente independiente. Tenemos que demostrar que si una combinación lineal tiene como resultado cero, es la trivial. Sea, por tanto,

$$\lambda_1 a_1+\lambda_2 a_2+\dots+\lambda_n a_n=0 \text{ con } \lambda_1,\lambda_2,\dots,\lambda_n\in K,$$

y vamos a suponer que al menos uno de los coeficientes es no nulo; para simplificar la notación tomemos λ_n. Entonces podemos escribir:

$$\lambda_n a_n=-\lambda_1 a_1-\lambda_2 a_2-\dots-\lambda_{n-1}a_{n-1}.$$

Como hemos supuesto que $\lambda_n\neq 0$, podemos dividir por λ_n, y obtenemos:

$$a_n=\left(-\frac{\lambda_1}{\lambda_n}\right)a_1+\left(-\frac{\lambda_2}{\lambda_n}\right)a_2+\dots+\left(-\frac{\lambda_{n-1}}{\lambda_n}\right)a_{n-1}.$$

Luego a_n se expresa como combinación lineal del resto de los vectores del sistema. Esto significa que el sistema $(a_1,a_2,\dots,a_n)$ es linealmente dependiente, lo cual contradice lo supuesto. Por tanto, ninguno de los coeficientes de la combinación lineal es no nulo.

Para completar la demostración, consideremos ahora que la única combinación lineal que tiene como resultado cero es la trivial y demostremos que el sistema $(a_1,a_2,\dots,a_n)$ es linealmente independiente.

Supongamos que uno de los vectores del sistema se puede expresar como combinación lineal de los demás; para simplificar la notación, tomemos el vector a_1. Entonces escribimos:

$$a_1=\lambda_2a_2+\lambda_3a_3+\dots+\lambda_na_n.$$

De aquí tenemos:

$$-a_1+\lambda_2a_2+\lambda_3a_3+\dots+\lambda_na_n=0,$$

lo cual contradice lo supuesto, pues se trata de una combinación lineal que tiene como resultado cero, y no es la trivial porque al menos el coeficiente de a_1 es no nulo.

Por tanto, *ninguno* de los vectores del sistema se puede expresar como combinación lineal de los demás y el sistema es linealmente independiente.

Nota. El estudiante debe observar lo siguiente:

1) La demostración no pierde generalidad cuando tomamos los vectores a_1 y a_n para simplificar la notación. Sugerimos que intente escribir una demostración sin hacer estas suposiciones.
2) La demostración es completamente válida, aun en el caso de una familia infinita de vectores.

Ejemplos

1. En $\mathbb{R}^3$ tomemos el sistema

   $(a_1=(1,1,1), a_2=(1,-1,1), a_3=(1,1,0)$.

   del ejemplo anterior e investiguemos si este es linealmente independiente. Para ello tenemos que comprobar si la única combinación lineal que tiene como resultado cero es la trivial. Tomemos una combinación lineal cualquiera que tenga como resultado cero.

$$\lambda_1a_1+\lambda_2a_2+\lambda_3a_3=0$$

$$\begin{pmatrix}\lambda_1\\ \lambda_1\\ \lambda_1\end{pmatrix}+\begin{pmatrix}\lambda_2\\ -\lambda_2\\ \lambda_2\end{pmatrix}+\begin{pmatrix}\lambda_3\\ \lambda_3\\ 0\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

$$\begin{pmatrix}\lambda_1+\lambda_2+\lambda_3\\ \lambda_1-\lambda_2+\lambda_3\\ \lambda_1+\lambda_2\end{pmatrix}=\begin{pmatrix}0\\ 0\\ 0\end{pmatrix}$$

Obtenemos el sistema de ecuaciones lineales homogéneo:

$$\begin{cases} \lambda_1+\lambda_2+\lambda_3=0 \\ \lambda_1-\lambda_2+\lambda_3=0 \\ \lambda_1+\lambda_2 \quad\;\; =0 \end{cases}$$

Al aplicarle el método de Gauss, este sistema se transforma en el sistema equivalente:

$$\begin{cases} \lambda_1+\lambda_2+\lambda_3=0 \\ 0-2\lambda_2+0=0 \\ \qquad -\lambda_3=0 \end{cases}$$

Por tanto, la única solución del sistema es la trivial, lo cual significa que $\lambda_1 a_1+\lambda_2 a_2+\lambda_3 a_3=0 \Rightarrow \lambda_1=\lambda_2=\lambda_3=0$.
Luego, el sistema (a_1,a_2,a_3) es linealmente independiente.

2. En $\mathbb{R}^4$ consideremos el sistema
$(a_1=(0,3,2,1),\ a_2=(1,2,3,0),\ a_3=(-1,1,-1,1),\ a_4=(1,-1,1,-1))$.
Queremos investigar si el sistema es linealmente independiente. Por tanto, planteamos:
$\lambda_1 a_1+\lambda_2 a_2+\lambda_3 a_3+\lambda_4 a_4=0$.
Operando obtenemos el sistema de ecuaciones:

$$\begin{cases} \qquad \lambda_2-\ \lambda_3+\lambda_4=0 \\ 3\lambda_1+2\lambda_2+\lambda_3-\lambda_4=0 \\ 2\lambda_1+3\lambda_2-\lambda_3+\lambda_4=0 \\ \quad\ \lambda_1+\quad \lambda_3-\lambda_4=0 \end{cases}$$

Los coeficientes $\lambda_1,\lambda_2,\lambda_3,\lambda_4$ que buscamos son soluciones de este sistema de ecuaciones. Aplicando Gauss el sistema se reduce al sistema equivalente:

$$\begin{cases} \lambda_1+\lambda_3-\lambda_4=0 \\ \lambda_2-\lambda_3+\lambda_4=0 \end{cases}$$

Como vemos este sistema de cuatro incógnitas y de rango 2, tiene infinitas soluciones no nulas. Por tanto, existen valores de $\lambda_1,\lambda_2,\lambda_3,\lambda_4$ no todos iguales a cero, que dan como resultado:
$\lambda_1 a_1+\lambda_2 a_2+\lambda_3 a_3+\lambda_4 a_4=0$.
Luego el sistema de vectores (a_1,a_2,a_3,a_4) es linealmente dependiente. Si tomamos una solución particular del sistema:
$\lambda_1=3, \lambda_2=-3, \lambda_3=-1, \lambda_4=2$, tenemos:
$3a_1-3a_2-a_3+2a_4=0$.

Podemos concluir que esta propiedad elimina la dificultad práctica de la definición de dependencia lineal, pues ya no hay que ir tomando uno a uno los vectores del sistema para comprobar si es combinación lineal de las demás o no. En el caso de los espacios vectoriales K^n, el problema siempre se reduce a hallar la solución de un sistema de ecuaciones lineales homogéneo. En este caso hay dos posibilidades:

1ra. El sistema tiene una solución única, la trivial; por tanto, el sistema de vectores que se está investigando es linealmente independiente.

2da. El sistema tiene infinitas soluciones; por tanto, tiene soluciones no triviales y el sistema de vectores que se está considerando es linealmente dependiente.

Algunos autores emplean como definición esta caracterización de la dependencia lineal, porque tiene valor práctico; pero en esta forma no se aprecia el problema fundamental, y es que uno de los vectores se puede expresar como combinación lineal de los demás. Es importante tener bien claro el concepto definido, y cómo demostrar prácticamente que un sistema de vectores es linealmente independiente o no.

Una consecuencia más de esta caracterización consiste en que ahora tiene sentido hablar de la dependencia o independencia lineal de un sistema formado por un solo vector.

Sea (v) un sistema formado por un único vector. Si le aplicamos la caracterización de independencia lineal, tenemos que investigar las posibilidades para que, dado un escalar λ, tengamos:

$\lambda v=0.$

En este caso hay dos posibilidades:

1ra. Si $v\neq 0$, entonces $\lambda=0$ y, por tanto, el sistema (v) es linealmente independiente.

2da. Si $v=0$, entonces para cualquier valor de λ resulta $\lambda v=0$ y, por tanto, el sistema (0) es linealmente dependiente.

Ya habíamos visto el resultado siguiente: *Si un sistema de vectores contiene al vector cero, el sistema es linealmente dependiente.*

Ahora podemos analizarlo de otra manera:

Si el sistema contiene al vector cero, entonces contiene a un subsistema, formado por el vector cero, que es linealmente dependiente, y como contiene a un subsistema linealmente dependiente, el sistema es linealmente dependiente por la proposición 1.4.

Ejemplo

Vamos a probar que en $\mathbb{R}^2$ un sistema de vectores linealmente independiente tiene a lo sumo dos vectores.

Tomemos un sistema de tres vectores de $\mathbb{R}^2$:

$a=(a_1,a_2),\ b=(b_1,b_2),\ c=(c_1,c_2),$

y demostremos que son linealmente dependientes.

$$xa+yb+zc=0$$
$$x(a_1,a_2)+y(b_1,b_2)+z(c_1,c_2)=(0,0)$$
$$(xa_1+yb_1+zc_1,\ xa_2+yb_2+zc_2)=(0,0)$$

Esto equivale al sistema de ecuaciones

$$\begin{cases} a_1x+b_1y+c_1z=0 \\ a_2x+b_2y+c_2z=0 \end{cases}$$

que por ser homogéneo y tener menos ecuaciones que incógnitas, admite infinitas soluciones no triviales.

Luego, siempre existen números x, y, z, no nulos todos, tales que $xa+yb+zc=0$ y, por tanto, el sistema (a,b,c) es linealmente dependiente.

Aquí se nos plantea un problema que vamos a dejar abierto por el momento:

En un espacio vectorial, ¿existe un límite al número de vectores que puede tener un sistema linealmente independiente?

Veamos un ejemplo de cómo se comprueba esta propiedad con un sistema infinito de vectores.

Ejemplo

En el espacio $K[x]$ consideremos el sistema $(1,x,x^2,\ldots,x^n\ldots)$. Vamos a demostrar que este sistema es linealmente independiente. Para ello tomamos una combinación lineal del sistema y la igualamos a cero.

$$\lambda_0+\lambda_1x+\lambda_2x^2+\ldots+\lambda_nx^n+\ldots=0$$

Recordemos que casi todos los coeficientes λ_k son iguales a cero 0, lo que es lo mismo en este caso, que a partir de un cierto n todos son iguales a cero.

$$\lambda_{n+1}=\lambda_{n+2}=\ldots=0$$

Por tanto, podemos escribir:

$$\lambda_0+\lambda_1x+\lambda_2x^2+\ldots+\lambda_nx^n=0.$$

Pero en esta igualdad el miembro de la izquierda es un polinomio y un polinomio es nulo si y solo si todos y cada uno de sus coeficientes son iguales a cero; por tanto,

$$\lambda_0=\lambda_1=\lambda_2=\ldots=\lambda_n=0.$$

Luego, todos los coeficientes λ_k de la combinación lineal son iguales a cero. De aquí concluimos que el sistema $(1,x,x^2,..,x^n\ldots)$ es linealmente independiente.

Como una consecuencia del teorema 1.5, obtenemos el siguiente criterio para determinar cuándo n vectores de K^n son linealmente independientes.

PROPOSICIÓN 1.5

Sean $(a_1, a_2, \ldots, a_n)$ un sistema de n vectores del espacio K^n. El sistema es linealmente independiente si y solo si la matriz A cuyas columnas están formadas por las componentes de los vectores $a_1, a_2, \ldots, a_n$ tiene determinante no nulo.

Demostración

Si el sistema $(a_1, a_2, \ldots, a_n)$ es linealmente independiente entonces la única combinación lineal que tiene como resultado cero es la trivial. Si tomamos

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n = 0,$$

obtenemos un sistema de ecuaciones lineales con las incógnitas $\lambda_1, \lambda_2, \ldots, \lambda_n$ cuya matriz es A. Como este sistema admite solamente la solución trivial, entonces rango $A = n$ y, por tanto,

$$\det A \neq 0.$$

Para demostrar que la condición es suficiente, formemos una combinación lineal que sea igual a cero.

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_n a_n = 0.$$

De aquí obtenemos un sistema de ecuaciones lineales homogéneo, que tiene como matriz a la matriz A. Como $\det A \neq 0$, rango $A = n$ y el sistema tiene solución única. Además, como el sistema es homogéneo y admite la solución trivial, esta es la única solución posible. Por tanto, el sistema $(a_1, a_2, \ldots, a_n)$ es linealmente independiente.

1.11 Sistema generador de un espacio vectorial

Al plantear la definición de combinación lineal surgen dos problemas fundamentales, uno de los cuales ya estudiamos: la dependencia lineal de un sistema de vectores; el otro problema consiste en saber cuándo un vector del espacio se puede expresar como una combinación lineal de un sistema de vectores dado.

Ejemplos

1. En $K_n[x]$, si consideramos el sistema de vectores $(1, x, x^2, \ldots, x^{n-1})$, entonces todo polinomio $p(x)$ se puede expresar como combinación lineal del sistema:

   $p(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}$, donde los coeficientes de la combinación lineal son los coeficientes del polinomio.

2. Si en $\mathbb{R}^4$ tomamos los vectores

   $a_1 = (1,2,3,0)$, $a_2 = (0,-1,1,1)$, $a_3 = (2,5,5,-1)$,

   podemos demostrar que no todos los vectores del espacio $\mathbb{R}^4$ se pueden expresar como combinación lineal de estos tres vectores. En particular, vamos a tomar el vector $(0,0,0,1)$ y comprobar que no se puede expresar como combinación lineal del sistema (a_1, a_2, a_3). Para ello planteamos la ecuación:

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = x \begin{pmatrix} 1 \\ 2 \\ 3 \\ 0 \end{pmatrix} + y \begin{pmatrix} 0 \\ -1 \\ 1 \\ 1 \end{pmatrix} + z \begin{pmatrix} 2 \\ 5 \\ 5 \\ -1 \end{pmatrix}$$

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x \\ 2x \\ 3x \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ -y \\ y \\ y \end{pmatrix} + \begin{pmatrix} 2z \\ 5z \\ 5z \\ -z \end{pmatrix}$$

$$\begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} x \quad +2z \\ 2x - y + 5z \\ 3x + y + 5z \\ y - z \end{pmatrix}$$

Por tanto, si el vector $(0,0,0,1)$ se puede expresar como combinación lineal de los vectores a_1, a_2, a_3, entonces los coeficientes de la combinación lineal satisfacen el sistema:

$$\begin{cases} x \quad +2z = 0 \\ 2x - y + 5z = 0 \\ 3x + y + 5z = 0 \\ y - z = 1 \end{cases}$$

Este sistema es incompatible, por tanto, el vector $(0,0,0,1)$ no se puede escribir como combinación lineal de los vectores a_1, a_2, a_3.

De las posibilidades que hemos planteado resulta que un sistema de vectores puede ser tal, que todo elemento del espacio se escriba como una combinación lineal del sistema. Es por esto que damos la definición siguiente:

DEFINICIÓN 1.13

Sean E un espacio vectorial sobre K, $(a_1, a_2, \ldots, a_n)$ un sistema de vectores. Se dice que el sistema $(a_1, a_2, \ldots, a_n)$ es un *generador* del espacio E, o que el espacio E está generado por el sistema $(a_1, a_2, \ldots, a_n)$, si todo vector de E se puede expresar como combinación lineal del sistema $(a_1, a_2, \ldots, a_n)$.

Ejemplos

1. En $K_n[x]$, el sistema $(1, x, x^2, \ldots, x^{n-1})$ es un generador del espacio.
2. En K como K-espacio, el sistema de un solo elemento (1) es un generador.
3. Si consideramos $\mathbb{C}$ como $\mathbb{R}$-espacio, el sistema $(1, i)$ es un generador, pues todo número complejo se escribe en la forma $a \cdot 1 + bi;\ a, b \in \mathbb{R}$.
4. El sistema $(e_1 = (1,0,0),\ e_2 = (0,1,0),\ e_3 = (0,0,1))$ es un generador de $\mathbb{R}^3$.

5. Sea E el espacio solución del sistema de ecuaciones homogéneo:

$$\begin{cases} x+2y-\ \ z=0 \\ 2x+5y+2z=0 \\ x+4y+7z=0 \\ x+3y+3z=0 \end{cases}$$

1 2
2 5
1 4
1 3

Este sistema se reduce por el método de Gauss al sistema equivalente:

$$\begin{cases} x+2y-\ \ z=0 \\ \qquad y+4z=0 \end{cases}$$

x − 8z − z = 0
y = −4z
x − 9z = 0
x = 9z

cuyas soluciones se pueden expresar en la forma:
(9z, −4z, z)
$E=\{(9a, -4a, a)/a\in \mathbb{R}\}$.
Esto significa que todo vector solución del sistema, es decir, todo elemento del espacio E, es un múltiplo del vector
$v=(9, -4, 1)$.
Por tanto, el espacio E está generado por el sistema $G=(v)$.

6. Sea S el espacio solución del sistema de ecuaciones:

$$\begin{cases} x+2y-3z+2t=0 \\ 2x+5y-8z+6t=0 \\ 3x+4y-5z+2t=0 \end{cases}$$

Este sistema se reduce por el método de Gauss al sistema equivalente:

$$\begin{cases} x+2y-\ 3z+2t=0 \\ \qquad y-2z+2t=0 \end{cases}$$

Escribimos las soluciones de este sistema en la forma:
$S=\{(-a+2\beta,\ 2a-2\beta, a, \beta,)/a,\beta\in \mathbb{R}\}$.
De aquí resulta que un vector $u\in S$ se puede expresar así:
$u=a(-1,2,1,0)+\beta(2,-2,0,1)$
Por tanto, el sistema de vectores de S formado por los vectores $(-1,2,1,0)$ y $(2,-2,0,1)$ es un generador de S.

7. En $K[x]$ el sistema infinito $(1, x, x^2, \ldots, x^n, \ldots)$ es un generador del espacio.

Algunos espacios vectoriales tienen sistemas generadores con un número finito de vectores y otros no admiten sistemas generadores finitos. Es evidente que si consideramos sistema infinito de vectores, todo espacio vectorial admite un generador; por ejemplo, el espacio completo es un generador de sí mismo.
$K[x]$ no admite un generador finito, al igual que ninguno de los espacios de funciones que dimos como ejemplos de espacios vectoriales.
Veamos por qué $K[x]$ no admite un generador finito.
Supongamos, por reducción al absurdo, que en $K[x]$ hay un sistema finito de polinomios $G=(p_1, p_2, \ldots, p_r)$ que genera todo $K[x]$.

En este sistema G, por ser finito, existe un polinomio que tiene grado mayor o igual que todos los demás; supongamos que este grado es n. Por tanto, G está formado por polinomios de grado menor o igual que n. Pero sabemos, por las propiedades de los polinomios, que
grad $(a_1p_1+a_2p_2+\ldots+a_kp_k)\leqslant$ Max (grad p_1, grad $p_2,\ldots$, grad p_k).
Por tanto, el grado de un polinomio que sea combinación lineal de los elementos de G es a lo sumo n, y de aquí que G no pueda generar a todo $K[x]$. Luego $K[x]$ no admite generadores finitos.

De acuerdo con los objetivos que nos hemos propuesto, solamente vamos a trabajar con espacios vectoriales finitamente generados, es decir, espacios que admiten un sistema generador con un número finito de elementos, y que también se denominan *de tipo finito*. En algunos momentos mencionaremos los casos infinitos, pero vamos a concentrarnos en el caso finito.

Algunas propiedades de los sistemas generadores

PROPOSICIÓN 1.6

Sean E un espacio vectorial sobre K, G un sistema generador. Entonces todo otro sistema de vectores que contenga a G es también un generador.

Demostración

Vamos a hacer la demostración para el caso en que E sea finitamente generado.

Sea $G=(e_1,e_2,\ldots,e_n)$ un generador de E. Entonces, para todo vector $v\in E$ existen escalares $\lambda_1,\lambda_2,\ldots,\lambda_n\in K$ tales que

$$v=\lambda_1e_1+\lambda_2\ e_2+\ldots+\lambda_ne_n.$$

Supongamos que tenemos otro sistema,

$$G'=(e_1,e_2,\ldots,e_n,a_1,\ a_2,\ \ldots,a_m),$$

que incluye al generador G. Entonces G' es también un generador pues para todo vector $v\in E$ podemos escribir:

$$v=\lambda_1e_1+\lambda_2e_2+\ldots+\lambda_ne_n+0a_1+0a_2+\ldots+0a_m$$

Nota. La demostración es esencialmente la misma para el caso infinito y solamente hay que tener en cuenta como se hacen las combinaciones lineales de sistemas infinitos.

PROPOSICIÓN 1.7

Sean E un espacio vectorial, $G=(a_1,a_2,\ldots,a_n)$ un sistema generador de E. Si uno de los vectores de G, a_i, se puede expresar como combinación lineal del resto de los vectores de G, entonces el sistema formado por $G\setminus\{a_i\}$ es también un generador de E.

Demostración

Tenemos que demostrar que todo vector $v\in E$ se puede expresar como combinación lineal de $(a_1,a_2,\ldots,a_{i-1},a_{i+1},\ldots,a_n)$.

Como el sistema G es un generador de E, dado un vector $v \in E$ podemos escribir:

$$v = \lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_i a_i + \ldots + \lambda_n a_n. \quad (*)$$

Pero, por hipótesis, el vector a_i se puede expresar como combinación lineal del resto de los vectores; por tanto, existen escalares $\alpha_1, \alpha_2, \ldots, \alpha_{i-1}, \alpha_{i+1}, \ldots, \alpha_n \in K$ tales que

$$a_i = \alpha_1 \; a_1 + \alpha_2 a_2 + \ldots + \alpha_{i-1} a_{i-1} + \alpha_{i+1} \; a_{i+1} + \alpha_n a_n.$$

Sustituyendo esta expresión de a_i en (*) tenemos:

$$v = \lambda_1 a_1 + \ldots + \lambda_{i-1} a_{i-1} + \lambda_i(\alpha_1 a_1 + \ldots + \alpha_{i-1} a_{i-1} + \alpha_{i+1} \; a_{i+1} + \ldots + \alpha_n a_n) + \\ + \ldots + \lambda_n a_n$$

Operando y agrupando:

$$v = (\lambda_1 + \lambda_i \alpha_1) \; a_1 + (\lambda_2 + \lambda_i \alpha_2) \; a_2 + \ldots + (\lambda_{i-1} + \lambda_i \; \alpha_{i-1}) a_{i-1} + \\ + (\lambda_{i+1} + \lambda_i \alpha_{i+1}) \; a_{i+1} + \ldots + (\lambda_n + \lambda_i \; \alpha_n) a_n$$

$$v = \beta_1 a_1 + \beta_2 a_2 + \ldots + \beta_{i-1} a_{i-1} + \beta_{i+1} a_{i+1} + \ldots + \beta_n a_n; \; \beta_1, \beta_2, \ldots, \; \beta_{i-1}, \; \beta_{i+1}, \ldots, \beta_n \in K.$$

Vemos que un vector v de E se puede expresar como combinación lineal de $(a_1, a_2, \ldots, a_{i-1}, a_{i+1}, \ldots, a_n) = G \setminus \{a_i\}$. Por tanto, $G \setminus \{a_i\}$ es un generador de E.

A partir de esta proposición podemos analizar el problema siguiente:

Dado un sistema generador de un espacio vectorial, ¿cuántos vectores se pueden eliminar del sistema generador sin que pierda la propiedad de ser generador?, ¿existe un número mínimo de vectores con los cuales se puede generar un espacio?

Vamos a analizar este problema para el caso del espacio $\mathbb{R}^3$.

Analicemos primeramente el problema del número mínimo de vectores que puede tener un sistema generador, y además, el problema pendiente del número máximo de vectores que puede tener un sistema linealmente independiente.

$\mathbb{R}^3$ es un espacio finitamente generado, pues el sistema formado por los vectores:

$$e_1 = (1,0,0), \; e_2 = (0,1,0), \; e_3 = (0,0,1)$$

es un generador porque todo elemento $v = (x, y, z) \in \mathbb{R}^3$ se puede escribir en la forma:

$$v = x e_1 + y e_2 + z e_3.$$

Observemos que el sistema (e_1, e_2, e_3) es, además de generador, linealmente independiente. ¿Existe un sistema con menos vectores que sea un generador? La respuesta es negativa, es decir, un conjunto de menos de tres vectores no es generador. Analicemos los casos.

Primer caso:

Si el sistema tiene un solo vector, no puede generar a todo el espacio; lo que genera es una recta.

Segundo caso:

El sistema tiene dos vectores (u_1, u_2), linealmente independientes (porque si no, volvemos a estar en el caso anterior); este sistema genera un plano, y no puede generar el espacio completo pues al menos el vector $u_1 x u_2$, producto vectorial de u_1 y u_2, es un vector de $\mathbb{R}^3$ que no se encuentra en el plano determinado por u_1 y u_2, y por tanto, no se puede expresar como combinación lineal de u_1 y u_2.

En resumen, un generador de $\mathbb{R}^3$ tiene que tener tres o más vectores.

Por otra parte, todo cojunto de más de tres vectores de $\mathbb{R}^3$ es linealmente dependiente. En efecto, consideremos cuatro vectores u_1, u_2, u_3, u_4 de $\mathbb{R}^3$ y demostremos que son linealmente dependientes. Para ello hay que demostrar que existen números reales x, y, z, t, no todos iguales a cero, tales que

$$xu_1 + yu_2 + zu_3 + tu_4 = 0. \quad (*)$$

Si tomamos:

$$u_1 = (a_{11}, a_{21}, a_{31}),$$
$$u_2 = (a_{12}, a_{22}, a_{32}),$$
$$u_3 = (a_{13}, a_{23}, a_{33}),$$
$$u_4 = (a_{14}, a_{24}, a_{34}),$$

escribiendo (*) componente a componente, obtenemos el sistema de ecuaciones lineales homogéneo:

$$\begin{cases} a_{11}x + a_{12}y + a_{13}z + a_{14}t = 0 \\ a_{21}x + a_{22}y + a_{23}z + a_{24}t = 0 \\ a_{31}x + a_{32}y + a_{33}z + a_{34}t = 0 \end{cases}$$

Este es un sistema compatible de rango menor o igual que 3 y con cuatro incógnitas, por tanto, admite infinitas soluciones.

En conclusión, existen valores para x, y, z y t, no todos iguales a cero, tales que

$$xu_1 + yu_2 + zu_3 + tu_4 = 0.$$

Luego un sistema que tenga cuatro vectores en $\mathbb{R}^3$ es linealmente dependiente. De aquí que un sistema linealmente independiente en $\mathbb{R}^3$ tiene a lo sumo tres vectores.

Analicemos ahora los resultados encontrados:

Un sistema generador de $\mathbb{R}^3$ tiene al menos tres vectores.

Un sistema linealmente independiente en $\mathbb{R}^3$ tiene a lo sumo tres vectores.

El número 3 actúa como frontera" entre la cantidad de vectores que puede haber en un sistema linealmente independiente y la cantidad de vectores que puede tener un sistema generador en el espacio $\mathbb{R}^3$. Entonces podemos afirmar:

En el espacio vectorial real $\mathbb{R}^3$, el número de vectores que tiene un sistema linealmente independiente es siempre menor o igual que el número de vectores que tiene un sistema generador del espacio.

A partir de estas ideas vamos a introducir, en el próximo epígrafe, dos nuevos conceptos.

1.12 Base y dimensión de un espacio vectorial

Según la segunda propiedad de los sistemas generadores (proposición 1.7), en un sistema generador de un espacio vectorial pueden existir vectores superfluos, es decir, vectores tales, que si los eliminamos del sistema, el sistema que queda sigue siendo un sistema generador. Pero, ¿cuántos vectores se pueden eliminar sin que se pierda la propiedad de ser un sistema generador? La respuesta a esta pregunta la daremos en este epígrafe.

La proposición 1.7 plantea que en un sistema generador G se puede eliminar un vector que sea combinación lineal de los demás; esto lo podemos hacer hasta obtener un sistema $G' \subset G$ que sea también generador y en el cual ningún vector se pueda expresar como combinación lineal de los demás. Es decir, G' es un generador, que además es linealmente independiente.

Aunque enunciaremos este resultado para sistemas finitos, lo cierto es que existen generadores linealmente independientes, y a estos generadores les daremos un nombre especial, pues satisfacen las dos propiedades fundamentales de los sistemas de vectores: la de independencia lineal y la de ser generador.

DEFINICIÓN 1.14

Sea E un espacio vectorial. Se denomina una *base* de E a un sistema de vectores que es a la vez linealmente independiente y un generador del espacio E.

Ejemplos

1. En $\mathbb{R}^3$ el sistema (e_1, e_2, e_3) es una base.
2. En $\mathbb{R}^3$ el sistema formado por $a_1=(1,1,1)$, $a_2=(1,1,0)$, $a_3=(1,0,0)$ es una base.
3. En $K_n[x]$ el sistema $(1, x, x^2, \ldots, x^{n-1})$ es una base.
4. En $K[x]$ el sistema infinito $(1, x, x^2, \ldots, x^n, \ldots)$ es una base.
5. En el espacio $\mathbb{C}$ como $\mathbb{R}$-espacio, podemos considerar la base $(1, i)$.

Nuestro interés fundamental se va a dedicar a los espacios que admiten una base finita y que denominaremos *espacios vectoriales de dimensión finita,* nombre que veremos plenamente justificado en este epígrafe, pues vamos a asociar a todo espacio vectorial finitamente generado, un número natural: la *dimensión* del espacio. Ya este número apareció en el caso de $\mathbb{R}^3$ como la frontera entre el número máximo de vectores en un sistema linealmente independiente y el número mínimo de vectores en un sistema generador. Este número está relacionado con el concepto *base*.

Lo esencial del empleo de una base está dado por el siguiente teorema, que será uno de los fundamentos del trabajo con los espacios vectoriales en el resto de este texto.

TEOREMA 1.6

Sean E un espacio vectorial sobre K, $(e_1,e_2,\dots,e_n)$ una familia de vectores en E. Entonces $(e_1,e_2,\dots,e_n)$ es una base de E si y solo si para todo vector $v\in E$ existen escalares *únicos* $\alpha_1,\alpha_2,\dots,\alpha_n\in K$ tales que

$$v=\alpha_1e_1+\alpha_2e_2+\dots+\alpha_ne_n.$$

Demostración

1) Supongamos que $(e_1,e_2,\dots,e_n)$ es una base. Entonces es un generador y todo vector $v\in E$ se expresa como combinación lineal de los elementos de la base, es decir, existen $\alpha_1,\alpha_2,\dots,\alpha_n\in K$ tales que
$$v=\alpha_1e_1+\alpha_2e_2+\dots+\alpha_n\ e_n.$$
Tenemos que demostrar que los escalares $\alpha_1,\alpha_2,\dots,\alpha_n$ son únicos.
Supongamos que existe otro conjunto de escalares, $\beta_1,\beta_2,\dots,\beta_n\in K$ tales que
$$v=\beta_1e_1+\beta_2e_2+\dots+\beta_ne_n.$$
Entonces
$$\alpha_1e_1+\alpha_2e_2+\dots+\alpha_ne_n=\beta_1e_1+\beta_2e_2+\dots+\beta_ne_n$$
Operando en esta identidad obtenemos:
$$(\alpha_1-\beta_1)\,e_1+(\alpha_2-\beta_2)\,e_2+\dots+(\alpha_n-\beta_n)\ e_n=0.$$
Pero como el sistema $(e_1,e_2,\dots,e_n)$, por ser base, es linealmente independiente, la única combinación lineal que tiene como resultado cero es la trivial, de donde
$$\alpha_1-\beta_1=0,\ \alpha_2-\beta_2=0,\dots,\ \alpha_n-\beta_n=0.$$
y
$$\alpha_1=\beta_1,\ \alpha_2=\beta_2,\dots,\ \alpha_n=\beta_n.$$
Por tanto, la forma de expresar un vector como combinación lineal de la base es única.

2) Supongamos ahora que el sistema de vectores $(e_1,e_2,\dots,e_n)$ cumple que todo vector $v\in E$ se puede expresar en forma única como combinación lineal del sistema. Tenemos que demostrar que $(e_1,\ e_2,\dots,e_n)$ es una base, es decir, que es un generador linealmente independiente.
Que $(e_1,e_2,\dots,e_n)$ es un generador es evidente por hipótesis, pues hemos afirmado que todo vector se expresa como combinación lineal del siste-

ma. Para demostrar que es linealmente independiente, tomamos una combinación lineal que tenga como resultado cero:

$\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_n e_n = 0$

Pero esto quiere decir que el vector cero se expresa como combinación lineal de $(e_1, e_2, \ldots, e_n)$ con los coeficientes $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$, y sabemos que el vector cero siempre se puede escribir:

$0 = 0e_1 + 0e_2 + \ldots + 0e_n$.

Como por hipótesis, la forma de expresar un elemento del espacio como combinación lineal del sistema $(e_1, e_2, \ldots, e_n)$ es única, entonces

$\lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$.

Esto demuestra que el sistema $(e_1, e_2, \ldots, e_n)$ es también linealmente independiente y, por tanto, es una base.

Este teorema es de gran utilidad práctica, pues dada una base en un espacio vectorial E, a todo vector $v \in E$ están asociados n números: $a_1, a_2, \ldots, a_n \in K$, que son los coeficientes de la expresión de v como combinación lineal de la base. Estos n números se denominan las *coordenadas* del vector en la base $(e_1, e_2, \ldots, e_n)$.

Ejemplos

1. En el espacio $\mathbb{R}^3$ si consideramos la base $(e_1 = (1,0,0),\ e_2 = (0,1,0),\ e_3 = (0,0,1))$, vemos inmediatamente que dado un elemento cualquiera $v = (a, b, c) \in \mathbb{R}^3$, podemos expresarlo como $v = ae_1 + be_2 + ce_3$ y los números a, b, c son precisamente las coordenadas del vector v en la base (e_1, e_2, e_3).
   En Física y en Geometría analítica se acostumbra emplear la notación $\vec{i}$, $\vec{j}$, $\vec{k}$ para e_1, e_2, e_3, respectivamente, y así tenemos que un vector $v \in \mathbb{R}^3$ se expresa como $v = a\vec{i} + b\vec{j} + c\vec{k}$.

2. En $K_n[x]$ consideremos la base $(1, x, x^2, \ldots, x^{n-1})$. Entonces, dado un polinomio $p(x) = a_{n-1}x^{n-1} + \ldots + a_1 x + a_0$ de $K_n[x]$, sus coordenadas en la base dada son $a_0, a_1, a_2, \ldots, a_{n-1}$, que son los coeficientes del polinomio.

3. En el plano π generado por los vectores
   $a_1 = (1,0,1)$ y $a_2 = (-1,1,0)$
   del espacio $\mathbb{R}^3$, que es el plano que satisface la ecuación cartesiana $x + y - z = 0$, tenemos que (a_1, a_2) es una base de dicho plano pues a_1 y a_2 son linealmente independientes (a_1 no es múltiplo de a_2 y a_2 no es múltiplo de a_1) y generan el subespacio π.

   El vector $v = (-2,3,1)$ pertenece al plano, y según el teorema 1.6 se puede expresar como combinación lineal única de a_1 y a_2. Busquemos esta combinación.

$$\begin{aligned} v &= ta_1 + sa_2 \\ (-2,3,1) &= (t,0,t) + (-s,s,0) \\ &= (t-s, s, t). \end{aligned}$$

Por tanto, obtenemos el sistema:

$$\begin{cases} t-s=-2 \\ s=\ \ 3 \\ t=\ \ 1 \end{cases}$$

que tiene la solución única $t=1$, $s=3$, y podemos escribir:

$$v=1a_1+3a_2$$

Esto significa que las coordenadas del vector $v=(-2,3,1)$ del espacio vectorial formado por los vectores del plano π cuya ecuación es $x+y-z=0$, en la base $(a_1=(1,0,1), a_2=(-1,1,0))$, son los números 1 y 3.

Observemos en este último ejemplo, que no importa como definimos los vectores de un espacio; en este caso son los elementos de $\mathbb{R}^3$ que satisfacen la ecuación $x+y-z=0$ pues una vez que tenemos una base del espacio, en el ejemplo (a_1,a_2), todo elemento del espacio está determinado por tantos números como vectores hay en la base; aquí son dos, y estos números son *únicos*. Esto va a simplificar extraordinariamente el trabajo con los espacios vectoriales, pues podemos trabajar con las coordenadas de los vectores, como veremos a partir del resultado siguiente:

PROPOSICIÓN 1.8

Sean E un espacio vectorial, $(e_1,e_2,\ldots,e_n)$ una base de E, y u, v dos vectores de E tales que

$$u=\alpha_1e_1+\alpha_2e_2+\ldots+\alpha_ne_n,$$
$$v=\beta_1e_1+\beta_2e_2+\ldots+\beta_ne_n.$$

entonces las coordenadas del vector $u+v$ son los números $\alpha_1+\beta_1$, $\alpha_2+\beta_2,\ldots,\alpha_n+\beta_n$, y las coordenadas del vector λu, $\lambda\in K$, son $\lambda\alpha_1,\lambda\alpha_2,\ldots,\lambda\alpha_n$.

Demostración

Para buscar las coordenadas de $u+v$, debido a su unicidad con respecto a la base dada, basta encontrar una expresión de $u+v$ como combinación lineal de $(e_1,e_2,\ldots,e_n)$.

$$\begin{aligned} u+v &=(\alpha_1e_1+\alpha_2e_2+\ldots+\alpha_ne_n)+(\beta_1e_1+\beta_2e_2+\ldots+\beta_ne_n) \\ &=\alpha_1e_1+\beta_1e_1+\alpha_2e_2+\beta_2e_2+\ldots+\beta_ne_n+\alpha_ne_n \\ &=(\alpha_1+\beta_1)\,e_1+(\alpha_2+\beta_2)\,e_2+\ldots+(\alpha_n+\beta_n)\,e_n. \end{aligned}$$

Por tanto, las coordenadas de $u+v$ en la base $(e_1, e_2,\ldots,e_n)$ son los números

$$\alpha_1+\beta_1, \alpha_2+\beta_2,\ldots, \alpha_n+\beta_n.$$

De igual forma, para λu.

$$\lambda u=\lambda(\alpha_1e_1+\alpha_2e_2+\ldots+\alpha_ne_n)=\lambda\alpha_1e_1+\lambda\alpha_2e_2+\ldots+\lambda\alpha_ne_n.$$

Por tanto, las coordenadas de λu son:

$\lambda a_1, \lambda a_2, \ldots, \lambda a_n$.

Esta proposición, que se demuestra mediante unos pocos cálculos, abre grandes posibilidades en el empleo de las bases y las coordenadas de un vector.

Ejemplo

Consideremos el espacio E como el conjunto de las soluciones del sistema:

$$\begin{cases} x+y-z=0 \\ x+2z+t-u=0 \\ x-3z-t+u=0 \\ 2x+y-4z-t+u=0 \end{cases}$$

Resolviendo este sistema obtenemos que toda solución, o sea, todo elemento de E, se expresa en la forma:

$$\begin{aligned} v &= (\alpha-\beta, \beta, \alpha, -\alpha+2\beta, \alpha+3\beta), \ \alpha,\beta \in \mathbb{R} \\ &= \alpha(1,0,1,-1,1) + \beta(-1,1,0,2,3). \end{aligned}$$

De aquí podemos inferir que los vectores $v_1=(1,0,1,-1,1)$ y $v_2=(-1,1,0,2,3)$ forman un generador del espacio solución y, además, comprobamos que v_1 no es múltiplo de v_2, ni v_2 lo es de v_1, por tanto, el sistema (v_1, v_2) es linealmente independiente en el espacio E. Luego una base del espacio E está formada por los vectores v_1 y v_2.

De acuerdo con la proposición 1.8 y el teorema 1.6, esto significa que para describir cualquier elemento del espacio E solamente se necesitan dos números: sus coordenadas en la base (v_1, v_2), y que se puede operar con los elementos de E solamente a partir de estos dos números. Así por ejemplo, el elemento

$$a=(5,-2,3,-7,-3)=3v_1-2v_2.$$

Por tanto, decimos que este es el vector de coordenadas $(3,-2)$ en la base (v_1, v_2).

Si tenemos otro elemento:

$$b=(-5,3,-2,8,7)=-2v_1+3v_2,$$

decimos que este es el vector de coordenadas $(-2,3)$ en la base (v_1,v_2).

Si queremos hacer una combinación lineal de a y b, según la proposición 1.8, en lugar de tomar las cinco componentes, podemos tomar las coordenadas en la base (v_1,v_2). Así,

$$\begin{aligned} 2a+3b &= 2(3,-2)+3(-2,3) \\ &= (0,5), \end{aligned}$$

es decir, el resultado es el vector de coordenadas $(0,5)$ en la base (v_1,v_2).

También podemos escribir las cinco componentes del vector:

$$2a+3b=0v_1+5v_2=(-5,5,0,10,15),$$

pero esto no es necesario pues para cualquier otra operación a realizar en E, es suficiente utilizar las coordenadas del vector.

Las coordenadas de un vector v de un espacio vectorial E sobre el cuerpo K, con respecto a una base $B=(a_1, a_2,...,a_n)$ es un conjunto ordenado de n escalares, $(\alpha_1, \alpha_2,..., \alpha_n)$, que denotaremos por $[v]_B$ y denominaremos *vector de coordenadas* del vector v con respecto a la base B de E.

La proposición 1.8 y el teorema 1.6 garantizan el trabajo con el vector de coordenadas en lugar del vector.

Por lo general, consideraremos el vector de coordenadas como una matriz de una columna o como un vector columna.

Vamos a analizar otro ejemplo que puede aclarar un poco más esta idea.

Ejemplo

Sea $MS(2,K)$ el conjunto de las matrices simétricas de orden 2 con coeficientes en K, es decir, las matrices de la forma

$$S=\begin{pmatrix} a & b \\ b & c \end{pmatrix} \text{ con } a, b, c, \in \mathrm{K}.$$

Este conjunto con las operaciones definidas para matrices, es un espacio vectorial. S es un subespacio de $M_2(K)$, lo cual dejamos para que el estudiante lo demuestre como ejercicio.

En el espacio $MS(2,K)$ las matrices

$$E_1=\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}, E_2=\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, E_3=\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

forman una base. Es fácil comprobar que (E_1,E_2,E_3) genera a $MS(2,K)$, pues una matriz simétrica cualquiera se expresa así:

$$S=\begin{pmatrix} a & b \\ b & c \end{pmatrix}=a\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}+b\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}+c\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$S=aE_1+bE_2+cE_3.$$

Queda por demostrar que (E_1,E_2,E_3) es linealmente independiente. Para ello tomamos una combinación lineal de la familia y la igualamos a cero.

$$xE_1+ yE_2+zE_3=0$$

$$xE_1+yE_2+zE_3=x\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}+y\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}+z\begin{pmatrix} 0 & 0 \\ 0 & 1 \end{pmatrix}$$

$$=\begin{pmatrix} x & y \\ y & z \end{pmatrix}$$

Entonces, a partir de la condición $xE_1+yE_2+zE_3=0$, obtenemos:

$$\begin{pmatrix} x & y \\ y & z \end{pmatrix}=0,$$

de donde resulta que

$$x=0,\ y=0,\ z=0.$$

Luego el sistema (E_1,E_2,E_3) es linealmente independiente y, por tanto, una base de $MS(2,K)$. Entonces, de acuerdo con el teorema 1.6, todo elemento de $MS(2,K)$ está determinado por tres números que son sus coordenadas en la base (E_1,E_2,E_3), y todas las operaciones entre matrices simétricas de orden 2 se pueden realizar a partir de estos tres números.

Este resultado era completamente esperado, pues en definitiva, en una matriz simétrica de orden 2 hay, en general, tres elementos distintos.

Veamos a qué conclusiones podemos llegar a partir de estos dos ejemplos.

En el primer ejemplo observamos que un subconjunto de vectores de $\mathbf{R}^5$, que es la solución de un sistema de ecuaciones, se puede expresar como pares de números reales: sus coordenadas en una cierta base; pero el conjunto de pares de números reales es el espacio $\mathbf{R}^2$.

En el segundo ejemplo vimos que en un cierto espacio de matrices, todos los elementos se expresan por un trío de elementos de K, esto es, K^3. Además, las operaciones a partir de las coordenadas coinciden con las operaciones en los espacios del tipo K^n. Esto conduce al resultado siguiente:

TEOREMA 1.7

Sea E un espacio vectorial sobre K. Si E tiene una base $(a_1,a_2,\ldots,a_n)$ formada por n vectores, entonces E es isomorfo a K^n.

Demostración

Recordemos que debemos definir una aplicación $f: E \longrightarrow K^n$ que sea biyectiva y que para todo $a, b \in E$, $\lambda \in K$, cumple:

$$f(a+b)=f(a)+f(b)$$
$$f(\lambda a)=\lambda f(a).$$

Como en E hay una base, por el teorema 1.6 a todo vector v de E le corresponden n elementos de K: sus coordenadas en la base B, o sea,

$$v=\alpha_1 a_1+\alpha_2 a_2+\ldots+\alpha_n a_n,$$

y estos n escalares son únicos para cada vector.

Definimos f como la aplicación que hace corresponder a cada vector de E sus coordenadas en la base B. Esto es,

$$f(v)=[v]_B.$$

El teorema 1.6 garantiza que la aplicación f está bien definida.

Esta aplicación es inyectiva, pues si $a, b \in E$ son tales que $f(a) = f(b)$, esto quiere decir que tienen las mismas coordenadas, $\alpha_1, \alpha_2, \ldots, \alpha_n$ y, por tanto,

$$a = \alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n = b.$$

f es sobreyectiva, pues para todo elemento de K^n, $(\alpha_1, \alpha_2, \ldots, \alpha_n)$, existe siempre el vector

$$\alpha_1 a_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n \text{ de } E \text{ y}$$
$$f(\alpha_1 + \alpha_2 a_2 + \ldots + \alpha_n a_n) = (\alpha_1, \alpha_2, \ldots, \alpha_n).$$

Acabamos de demostrar que f es biyectiva; falta ahora comprobar que la aplicación f cumple las otras dos condiciones.

Sean $a, b \in E$ con coordenadas $(\alpha_1, \alpha_2, \ldots, \alpha_n)$ y $(\beta_1, \beta_2, \ldots, \beta_n)$, respectivamente, en la base B. Entonces, por la proposición 1.8, las coordenadas correspondientes al vector $a+b$ y al vector λa en la base B son, respectivamente $(\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n)$ y $(\lambda \alpha_1, \lambda \alpha_2, \ldots, \lambda \alpha_n)$, y tenemos:

$$\begin{aligned} f(a+b) &= (\alpha_1 + \beta_1, \alpha_2 + \beta_2, \ldots, \alpha_n + \beta_n) \\ &= (\alpha_1, \ldots, \alpha_n) + (\beta_1, \ldots, \beta_n) \\ &= f(a) + f(b). \end{aligned}$$
$$f(\lambda a) = (\lambda \alpha_1, \lambda \alpha_2, \ldots, \lambda \alpha_n) = \lambda(\alpha_1, \alpha_2, \ldots, \alpha_n) = \lambda f(a).$$

Luego, la aplicación f es un isomorfismo entre E y K^n.

La construcción de la aplicación f es una consecuencia natural del teorema 1.6. Pero hay un problema: ¿El exponente n de la potencia K^n es *único* para el espacio E? Según la demostración, n es el número de elementos de una base de E; ¿puede E tener otra base con un número de elementos diferente de n? Como veremos, esto no es posible, ya que el número n tal que E es isomorfo a K^n es una *constante* del espacio E. Esto es una consecuencia del teorema siguiente que resuelve el problema sobre la relación entre el número de vectores que puede tener un sistema de vectores linealmente independiente y el número de vectores que puede tener un sistema generador.

TEOREMA 1.8

Sea E un espacio vectorial sobre K, finitamente generado. Sean $(g_1, g_2, \ldots, g_m)$ un sistema generador de E y $(a_1, a_2, \ldots, a_n)$ un sistema linealmente independiente de E. Entonces $n \leqslant m$.

Demostración

Como $(g_1, g_2, \ldots, g_m)$ es un generador de E, todo vector del espacio se puede expresar como combinación lineal de los vectores del sistema. En particular, vamos a expresar los elementos $a_1, a_2, \ldots, a_n$ como combinación de los elementos del generador.

$$a_1 = \alpha_{11} g_1 + \ldots + \alpha_{m1} g_m,\ a_2 = \alpha_{12} g_1 + \ldots + \alpha_{m2} g_m, \ldots, a_n = \alpha_{1n} g_1 + \ldots + \alpha_{mn} g_m.$$

En la notación de sumatoria,

$$a_i = \sum_{j=1}^{m} a_{ji}\, g_j,\quad i=1,2,\ldots,n$$

Vamos a demostrarlo por reducción al absurdo. Supongamos que $n>m$. Una combinación del sistema $(a_1,\ a_2,\ldots,\ a_n)$ es:

$$\begin{aligned} x_1a_1+x_2a_2+\ldots+x_na_n &= x_1(a_{11}g_1+\ldots+a_{m1}\,g_m)+\\ &+x_2(a_{12}g_1+\ldots+a_{m2}\,g_m)+\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &+x_n(a_{1n}g_1+\ldots+a_{mn}g_m). \end{aligned}$$

Utilizando la notación de sumatoria:

$$\begin{aligned} \sum_{i=1}^{n} x_ia_i &= \sum_{i=1}^{n} x_i\left(\sum_{j=1}^{m} a_{ji}\, g_j\right)\\ &= \sum_{i=1}^{n}\sum_{j=1}^{m} a_{ji}\, x_i\, g_j\\ &= \sum_{j=1}^{m}\left(\sum_{i=1}^{n} a_{ji}\, x_i\right) g_j \end{aligned}$$

La última expresión es una combinación lineal del sistema $(g_1,\ g_2,\ldots,g_m)$, en la cual vamos a investigar si todos los coeficientes pueden ser cero de modo que

$$\sum_{i=1}^{n} x_ia_i = \sum_{j=1}^{m} 0g_j = 0.$$

Analicemos las posibilidades con los coeficientes $x_1,x_2,\ldots,x_n$.
Tenemos que encontrar valores de $x_1,x_2,\ldots,x_n$ que cumplan:

$$\begin{aligned} &a_{11}x_1+a_{12}x_2+\ldots+a_{1n}x_n=0\\ &a_{21}x_1+a_{22}x_2+\ldots+a_{2n}x_n=0\\ &\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\\ &a_{m1}x_1+a_{m2}x_2+\ldots+a_{mn}x=0 \end{aligned}$$

Este es un sistema de ecuaciones lineales homogéneo que tiene más incógnitas que ecuaciones (supusimos $n>m$); por tanto, tiene infinitas soluciones no nulas. De aquí resulta que hay valores de $x_1,x_2,\ldots,x_n$, no todos nulos, tales que

$$\sum_{i=1}^{n} a_{ji}\, x_i=0,\ j=1,2,\ldots,m.$$

Por tanto, tenemos valores de $x_1, x_2, \ldots, x_n$, no todos nulos, tales que

$$x_1 a_1 + x_2 a_2 + \ldots + x_n a_n = \sum_{j=1}^{m} \left(\sum_{i=1}^{n} a_{ji}\, x_i \right) g_j$$
$$= \sum_{j=1}^{m} 0 g_j = 0$$

y el sistema de vectores $(a_1, a_2, \ldots, a_n)$ no es linealmente independiente, lo cual contradice la hipótesis.

Luego, $n \leqslant m$.

En resumen:

En un espacio vectorial finitamente generado, el número de vectores que puede tener un sistema linealmente independiente es menor o igual que el número de vectores que puede tener un sistema generador.

Además, en un espacio finitamente generado, todo sistema de vectores linealmente independiente tiene necesariamente un número finito de vectores.

Otra conclusión inmediata del teorema 1.8 está dada por el siguiente corolario, que también indica que hay una frontera entre el número de elementos que pueden tener los dos tipos de sistemas, en los espacios finitamente generados.

COROLARIO

Sea E un espacio vectorial sobre K, finitamente generado. Todas las bases de E tienen el mismo número de elementos.

Demostración

Supongamos que tenemos dos bases con números diferentes de elementos:

$(a_1, a_2, \ldots, a_n)$, $(b_1, b_2, \ldots, b_m)$

Como $(a_1, a_2, \ldots, a_n)$ es base, es linealmente independiente y como $(b_1, b_2, \ldots, b_m)$ es base, es generador. Entonces, por el teorema 1.8 tenemos:

$n \leqslant m$.

Pero por otra parte, como $(b_1, b_2, \ldots, b_m)$ es base, es linealmente independiente, y como $(a_1, a_2, \ldots, a_n)$ es base, es generador. Entonces, del teorema 1.8 resulta

$m \leqslant n$.

Por tanto, $n = m$, que era lo que queríamos demostrar.

Si un espacio vectorial está generado finitamente, todas sus bases tienen el mismo número de elementos y ese número es la frontera entre el número de elementos en un sistema linealmente independiente y el número de elementos en un sistema generador. Si ese número es n, podemos afirmar:

Un sistema linealmente independiente tiene a lo sumo n elementos y un sistema generador tiene al menos n elementos.

DEFINICIÓN 1.15

Sea E un espacio vectorial sobre K, finitamente generado. Se denomina *dimensión de E sobre K*, o simplemente dimensión de E, al número de elementos de una cualquiera de sus bases.

Ejemplos

1. La dimensión de $\mathbb{R}^3$ es 3.
2. La dimensión de K^n es n.
3. $K_n[x]$ tiene dimensión n.
4. $\mathbb{C}$ como $\mathbb{R}$-espacio tiene dimensión 2.
5. El espacio de las soluciones del sistema

$$\begin{aligned}&x+y-z=0\\&y+2z+t-u=0\\&x-3z-t+u=0\\&2x+y-4z-t+u=0\end{aligned}$$

tiene dimensión 2.

6. El espacio $MS(2,K)$ tiene dimensión 3.

Para completar la definición, establecemos

$\dim\{0\}=0.$

En lo sucesivo trabajaremos con espacios de dimensión finita.

Como vemos, y así se indica en el ejemplo de $\mathbb{R}^3$, hay un número que actúa como frontera entre el número de elementos de un sistema linealmente independiente y el número de elementos de un sistema generador. Esto podemos concretarlo en la siguiente caracterización de una base.

TEOREMA 1.9

Sean E un espacio vectorial, B una familia de elementos de E. Las siguientes propiedades son equivalentes:

(1) B es una base.

(2) B es un sistema linealmente independiente maximal.

(3) B es un sistema generador minimal.

Demostración

Vamos a hacer la demostración para el caso en que E sea de dimensión finita, pero el resultado es válido en general.

1) Probemos que (1) implica (2).
Supongamos que B es una base. Tenemos que demostrar que B es linealmente independiente maximal.
Como B es una base, es linealmente independiente. Solamente falta por probar que es maximal. Para ello hay que probar que no existe ninguna familia L de vectores de E que sea linealmente independiente y que contenga estrictamente a B.
Sea $B=(b_1,b_2,\ldots,b_n)$, dim $E=n$, y supongamos que existe $L=(b_1,b_2,\ldots,b_n,b_{n+1},\ldots,b_m)$ linealmente independiente y que $L\supset B$. Esto contradice el hecho de que B sea una base, pues B por ser base es generador, y por el teorema 1.8, $m\leqslant n$.
Por tanto, L no puede contener a B estrictamente.

2) Probemos que (2) implica (3).
Suponiendo que B es linealmente independiente maximal, tenemos que probar que es generador minimal. Esto quiere decir que no hay otro sistema de vectores de E, que sea un generador, estrictamente contenido en B.
Supongamos que existe un tal generador, sea este $G=(b_1,b_2,\ldots,b_k)$, tal que $G\subset B$, $G\neq B$.
Como B es un sistema linealmente independiente maximal y G un generador, por el teorema 1.8 tenemos: $n\leqslant k$.
Por tanto, G no puede estar contenido en B.
Falta probar que B es generador. Supongamos que no lo es, es decir, que existe un vector $x\in E$ que no se pueda expresar como combinación lineal de los elementos de B. Por tanto, el sistema $B'=B\cup\{x\}$ es un sistema linealmente independiente y $B'\supset B, B'\neq B$ lo cual contradice lo supuesto, es decir, que B es linealmente independiente maximal.

3) Probemos que (3) implica (1).
Supongamos que B es generador minimal. Hay que probar que B es una base. Para ello solamente falta demostrar que es linealmente independiente.
Supongamos que no lo sea. Esto significa que uno de los vectores de B se puede expresar como combinación lineal de los demás. Supongamos que este es b_i; pero la proposición 1.7 afirma que podemos eliminar el vector b_i, y $B\setminus\{b_i\}$ sigue siendo un generador, lo que contradice el carácter minimal de B, pues $B\setminus\{b_i\}\neq B$.
Por tanto, B es linealmente independiente y es una base.

Aunque todavía no hemos demostrado que en todo espacio vectorial siempre existe una base, de las dos caracterizaciones de las bases dadas en el teorema 1.9 resultan dos métodos para construirla, uno a partir de un sistema generador del espacio y otro a partir de un sistema linealmente independiente. Vamos a enunciar primero el método relativo al generador y a analizar la forma de trabajo que podemos emplear.

TEOREMA 1.10

Sean E un espacio vectorial, $E\neq\{0\}$, G un generador finito del espacio. Entonces existe una base de E, sea esta B, tal que

$$B\subset G.$$

Demostración

Tenemos que obtener un sistema de vectores que esté contenido en G y que sea linealmente independiente y generador.

G es un generador. Si G es linealmente independiente, entonces es una base. Luego, $B=G$.

Si G no es linealmente independiente, uno de los elementos de G, llamémosle a_{i_1}, se puede expresar como combinación lineal del resto de los elementos de G, y por la proposición 1.7, el sistema

$$G_1=G\setminus\{a_{i_1}\}$$

es también un generador de E.

Si G_1 es linealmente independiente, esta es la base que buscamos, y así termina el proceso. Si G_1 es linealmente dependiente, uno de sus elementos, llamémoslo a_{i_2}, se puede expresar como combinación lineal de los demás, y nuevamente por la proposición 1.7, el sistema

$$G_2=G_1\setminus\{a_{i_2}\}$$

es también un generador de E.

Si G_2 es linealmente independiente, terminó el proceso y G_2 es la base que buscamos; de lo contrario, continuamos el proceso hasta llegar a obtener un sistema

$$G_k=G_{k-1}\setminus\{a_{i_k}\}=G\setminus\{a_{i_1},a_{i_2},\ldots,\ a_{i_k}\}$$

que es un generador y del cual no se puede eliminar ningún vector sin perder la cualidad de generar el espacio. Esto quiere decir que ninguno de los vectores de G_k se puede expresar como combinación lineal de los demás; por tanto, G_k es un sistema linealmente independiente, y como también es un generador, G_k es una base de E.

Hemos encontrado una base $B=G_k$ tal que

$$B\subset G.$$

El teorema 1.10 demuestra que en un espacio vectorial, no nulo, que admite un generador finito, siempre existe una base, y además, da un método práctico para encontrar una base de un espacio vectorial, conocido un generador del espacio.

En este teorema se parte de un sistema generador finito y se encuentra una base que está contenida en ese sistema, es decir, se parte de la propiedad de que el sistema de vectores es generador y se busca la propiedad de la independencia lineal.

Podemos enunciar otro teorema que, en definitiva, garantiza el mismo resultado: la existencia de una base, pero que, a la inversa, parte de un sistema linealmente independiente para encontrar la propiedad de generador.

TEOREMA 1.11

Sean E un espacio vectorial de dimensión finita y $L = (a_1, a_2, \ldots, a_k)$ un sistema de vectores linealmente independiente. Entonces existe una base B de E tal que

$$L \subset \mathrm{B}.$$

Demostración

Como E es de dimensión finita n, el número máximo de vectores de un sistema linealmente independiente es n.

A partir del sistema L, vamos a construir un sistema de vectores que sea, además de linealmente independiente, un generador.

Por hipótesis, $L = (a_1, a_2, \ldots, a_k)$ es linealmente independiente. Si L es generador, ya tenemos la base que buscamos. Si no lo es, esto quiere decir que existe al menos un vector a_{k+1} de E que no está en el subespacio generado por L y, por tanto, no se puede expresar como combinación lineal de los elementos de L. Entonces, el sistema

$$L_1 = L \cup \{a_{k+1}\} = (a_1, a_2, \ldots, a_{k+1})$$

es un sistema linealmente independiente.

Supongamos que el sistema L_1 no es linealmente independiente. Entonces existen escalares $\lambda_1, \lambda_2, \ldots, \lambda_k, \lambda_{k+1}$, no todos nulos, tales que

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k + \lambda_{k+1} a_{k+1} = 0.$$

Como a_{k+1} no es combinación lineal del sistema L, entonces $\lambda_{k+1} = 0$, porque de lo contrario, podemos expresar a_{k+1} como combinación lineal de L. Como $\lambda_{k+1} = 0$ queda

$$\lambda_1 a_1 + \lambda_2 a_2 + \ldots + \lambda_k a_k = 0$$

y no todos los coeficientes de la combinación lineal son cero. Pero esto contradice la hipótesis de que L es linealmente independiente. Por tanto, L_1 es un sistema linealmente independiente.

Si L_1 es un generador de E, ya tenemos la base deseada; si no lo es, repetimos el proceso. Existe un vector a_{k+2} de E que no se puede expresar como combinación lineal de los vectores de L_1, y el sistema

$$L_2 = L_1 \cup \{a_{k+2}\} = (a_1, a_2, \ldots, a_k, a_{k+1}, a_{k+2})$$

es linealmente independiente.

Si L_2 es un generador, es la base buscada; si no, repetimos el proceso hasta llegar a un sistema

$$\begin{aligned} B &= L_m \\ &= L_{m-1} \cup \{a_{k+m}\} \\ &= (a_1, a_2, \ldots, a_k, a_{k+1}, \ldots, a_{k+m}), \quad k+m=n \end{aligned}$$

que, además de ser linealmente independiente, será un generador y, por tanto, una base de E.

Podemos garantizar que llegaremos a B en un número finito de pasos porque el espacio es de dimensión finita y hay, por tanto, un límite al número de vectores que puede tener un sistema linealmente independiente. En este caso, $B=L_m$ es linealmente independiente y tiene n vectores, que es la dimensión del espacio. Por tanto, B es un conjunto linealmente independiente maximal, y por el teorema 1.9, es una base.

El teorema 1.11 se conoce como *teorema de completamiento de base.*

Sobre la base de los teoremas 1.10 y 1.11 podemos enunciar el siguiente resultado para espacios vectoriales de dimensión finita.

TEOREMA 1.12

Sea E un espacio vectorial de dimensión finita, $E\neq\{0\}$. Entonces siempre existen bases para E.

Demostración

Como $E\neq\{0\}$, existe en E un vector a, no nulo, y el sistema $L=(a)$ es un sistema linealmente independiente. Y por el teorema 1.11, existe una base B que contiene a L.

Nota. El teorema 1.12 es válido aun si el espacio es de dimensión infinita, porque el teorema de completamiento de base se cumple también en el caso infinito, pero los instrumentos matemáticos para la demostración no están dentro de los objetivos que nos hemos propuesto en este curso. La demostración general puede verse en alguno de los textos de la bibliografía (por ejemplo, en Chambadal).

Para completar el enunciado del teorema es usual incluir el caso en que $E=\{0\}$, diciendo que el conjunto vacío, ϕ, es un generador y una base de $\{0\}$ y, por tanto, dim $\{0\}=0$.

Entonces podemos afirmar:

Todo espacio vectorial admite siempre una base.

TEOREMA 1.13

Dos espacios de dimensión finita son isomorfos si y solo si tienen la misma dimensión.

Demostración

Sean E y F dos espacios de dimensión finita.

1) Supongamos que E y F tienen la misma dimensión. Vamos a demostrar que E y F son isomorfos.
Si E y F tienen la misma dimensión n, toda base de E y toda base de F tienen n elementos. Además, sabemos que todo espacio vectorial tiene base; sean $(e_1,e_2,\dots,e_n)$ una base de E y $(e'_1,e'_2,\dots,e'_n)$ una base de F. Definimos una aplicación $f\colon E \longrightarrow F$ por:

$$f(x) = \sum_{i=1}^{n} x_i\, e'_i \ \forall x\in E \text{ donde } x = \sum_{i=1}^{n} x_i e_i.$$

1) Comprobemos que f es un isomorfismo.
1ro. f es inyectiva.

Sean $x, y \in E$; $x = \sum_{i=1}^{n} x_i e_i$, $y = \sum_{i=1}^{n} y_i e_i$.

$$f(x) = f(y) \Leftrightarrow \sum_{i=1}^{n} x_i e'_i = \sum_{i=1}^{n} y_i e'_i \Leftrightarrow x_i = y_i, \ i = 1, 2, \ldots, n,$$

Por tanto, por la unicidad de las coordenadas, $x = y$.

2do. f es sobreyectiva.
$\forall v \in F, v$ se expresa en forma única como combinación lineal de la base de F:

$$v = \sum_{i=1}^{n} v_i e'_i.$$

Entonces el vector $x = \sum_{i=1}^{n} v_i e_i \in E$ cumple:

$f(x) = v$.

3ro. $f(x+y) = f(x) + f(y)$.

Sean $x = \sum_{i=1}^{n} x_i e_i$, $y = \sum_{i=1}^{n} y_i e_i$ elementos de E. Entonces

$$x + y = \sum_{i=1}^{n} (x_i + y_i) e_i.$$

Por tanto,

$$f(x+y) = \sum_{i=1}^{n} (x_i + y_i) e'_i = \sum_{i=1}^{n} x_i e'_i + y_i e'_i = \sum_{i=1}^{n} x_i e'_i + \sum_{i=1}^{n} y_i e'_i =$$

$= f(x) + f(y)$.

4to. $f(\lambda x) = \lambda f(x)$.

Sean $x = \sum_{i=1}^{n} x_i e_i$, $\lambda \in K$. Entonces

$$\lambda x = \sum_{i=1}^{n} \lambda x_i e_i.$$

Por tanto,

$$f(\lambda x) = \sum_{i=1}^{n} (\lambda x_i) e'_i = \sum_{i=1}^{n} \lambda (x_i e'_i) = \lambda \sum_{i=1}^{n} x_i e'_i = \lambda f(x).$$

Por consiguiente, existe un isomorfismo entre E y F.

2) Supongamos ahora que E y F son isomorfos, es decir, que existe una aplicación biyectiva $f\colon E \longrightarrow F$ tal que
$f(x+y) = f(x) + f(y)$ y $f(\lambda x) = \lambda f(x)\, x, y \in E,\ \lambda \in K.$
Como E es de dimensión finita n, existe una base $(e_1, e_2, \ldots, e_n)$ de E. Vamos a demostrar que $(f(e_1), f(e_2), \ldots f(e_n))$ es una base de F.

1ro. $(f(e_1), f(e_2), \ldots f(e_n))$ es linealmente independiente en F.
Tomemos una combinación lineal e igualémosla a cero.

$$\lambda_1 f(e_1) + \lambda_2 f(e_2) + \ldots + \lambda_n f(e_n) = \sum_{i=1}^{n} \lambda_i f(e_i) = 0$$

Por las propiedades de f,

$$\sum_{i=1}^{n} \lambda_i f(e_i) = \sum_{i=1}^{n} f(\lambda_i e_i) = f\left(\sum_{i=1}^{n} \lambda_i e_i\right) = 0.$$

Además, $f(0x) = 0f(x)$, $0 \in K$, $x \in E \Rightarrow f(0) = 0$.
Y por ser f inyectiva, el cero es el único que se transforma en el cero.

$$f\left(\sum_{i=1}^{n} \lambda_i e_i\right) = 0 \Rightarrow \sum_{i=1}^{n} \lambda_i e_i = 0 \Rightarrow \lambda_1 = \lambda_2 = \ldots = \lambda_n = 0$$

porque $(e_1, e_2, \ldots, e_n)$ es linealmente independiente por ser una base.

2do. $(f(e_1), f(e_2), \ldots, f(e_n))$ genera a F.
Sea $v \in F$. Por ser f sobreyectiva, existe $x \in E$ tal que $v = f(x)$, y para $x \in E$ existen escalares $x_1, x_2, \ldots, x_n \in K$ tales que $x = \sum_{i=1}^{n} x_i e_i$ por ser $(e_1, e_2, \ldots, e_n)$ una base de E. Como f es un isomorfismo, tenemos:

$$v = f(x) = f\left(\sum_{i=1}^{n} x_i e_i\right) = \sum_{i=1}^{n} f(x_i e_i) = \sum_{i=1}^{n} x_i f(e_i).$$

Esto quiere decir que $(f(e_1), f(e_2), \ldots f(e_n))$ genera a F.
Entonces, como $(f(e_1), f(e_2, \ldots f(e_n))$ es linealmente independiente y genera a F es una base de F. Por tanto,
$\dim F = n = \dim E$.

Hasta aquí, los resultados fundamentales sobre las bases en los espacios de dimensión finita son los siguientes:

1) Todo espacio vectorial admite base.
2) Todas las bases de un espacio vectorial tienen el mismo número de elementos.

3) Una base es un sistema linealmente independiente maximal.
4) Una base es un sistema generador minimal.
5) Dada una base, cada elemento del espacio se expresa en forma única como combinación lineal de sus elementos.
6) De todo sistema generador finito se puede extraer una base.
7) Todo sistema linealmente independiente puede completarse para obtener una base.

1.13 Base y dimensión de la suma y la suma directa de subespacios

Vamos a emplear los teoremas que demostramos sobre las bases en el epígrafe anterior, para resolver algunos problemas relacionados con las bases y la dimensión de la suma, y la suma directa.

TEOREMA 1.14

Sean E un espacio vectorial sobre un cuerpo K, F_1 y F_2 dos subespacios de E, de dimensión finita. Entonces F_1+F_2 es de dimensión finita y se cumple:

$$\dim(F_1+F_2)=\dim F_1+\dim F_2-\dim(F_1\cap F_2)$$

Demostración

La idea de la demostración es construir bases de F_1 y F_2, y a partir de ellas encontrar una base para F_1+F_2 y comparar el número de elementos que hay en cada una. Para construir las bases emplearemos el teorema de completamiento de base.

Como F_1 y F_2 son de dimensión finita, entonces $F_1\cap F_2$ también es de dimensión finita.

Sea $S=(e_1,e_2,\dots,e_k)$ una base de $F_1\cap F_2$. Como $(F_1\cap F_2)\subset F_1$, $(e_1,e_2,\dots,e_k)$ es un sistema linealmente independiente en F_1, y por el teorema de completamiento de base, existe una base de F_1 de la forma:

$$B_1=(e_1,e_2,\dots,e_k,v_1,v_2,\dots,v_p).$$

Análogamente, existe una base de F_2 de la forma:

$$B_2=(e_1,e_2,\dots,e_k,w_1,w_2,\dots,w_q).$$

Aquí hemos denotado:

$\dim(F_1\cap F_2)=k$
$\dim F_1=k+p$
$\dim F_2=k+q$

1) Vamos a demostrar que
$B=(e_1,e_2,\dots,e_k,v_1,v_2,\dots,v_p,w_1,w_2,\dots,w_q)$ es una base de F_1+F_2.

1ro. Demostremos que B genera a F_1+F_2.

Sea $a \in F_1+F_2$. Entonces existen $x \in F_1$, $y \in F_2$ tales que $a=x+y$; pero como B_1 y B_2 son bases de F_1 y F_2 respectivamente, tenemos:

$x=\alpha_1 e_1+\alpha_2 e_2+\ldots+\alpha_k e_k+\beta_1 v_1+\beta_2 v_2+\ldots+\beta_p v_p$

$y=\alpha_1' e_1+\alpha_2' e_2+\ldots+\alpha_k' e_k+\gamma_1 w_1+\gamma_2 w_2+\ldots+\gamma_q w_q$

con $\alpha_1,\ldots,\alpha_k, \alpha_1',\ldots, \alpha_k', \beta_1,\ldots,\beta_1,\ldots,\beta_p, \gamma_1,\ldots,\gamma_q \in K$.

Por tanto,

$$\begin{aligned} a&=x+y\\ &=(\alpha_1 e_1+\ldots+\alpha_k e_k+\beta_1 v_1+\ldots+\beta_p v_p)+\\ &\quad+(\alpha_1' e_1+\ldots+\alpha_k' e_k+\gamma_1 w_1+\ldots+\gamma_q w_q)\\ &=(\alpha_1+\alpha_1')\, e_1+(\alpha_2+\alpha_2')\, e_2+\ldots+(\alpha_k+\alpha_k')\, e_k+\beta_1 v_1+\beta_2 v_2+\ldots+\\ &\quad+\beta_p v_p+\gamma_1 w_1+\gamma_2 w_2+\ldots+\gamma_q w_q \end{aligned}$$

y B genera a F_1+F_2.

2do. Demostremos que B es linealmente independiente.

Consideremos una combinación lineal de B que tenga como resultado el vector nulo:

$\alpha_1 e_1+\alpha_2 e_2+\ldots+\alpha_k e_k+\beta_1 v_1+\beta_2 v_2+\ldots+\beta_p v_p+\gamma_1 w_1+\gamma_2 w_2+\ldots+\gamma_q w_q=0$

Vamos a emplear la notación de sumatoria:

$$\sum_{i=1}^{k} \alpha_i e_i+\sum_{j=1}^{p} \beta_j v_j+\sum_{h=1}^{q} \gamma_h w_h=0.$$

Despejando tenemos:

$$-\sum_{h=1}^{q} \gamma_h w_h=\sum_{i=1}^{k} \alpha_i e_i+\sum_{j=1}^{p} \beta_j v_j,$$

de donde resulta que $\sum_{h=1}^{q} \gamma_h w_h$ pertenece a F_2 por ser una combinación lineal de elementos de B_2, y también pertenece a F_1 por ser una combinación lineal de elementos de B_1; por tanto, pertenece a $F_1 \cap F_2$ y se puede expresar, en forma única, como una combinación lineal de elementos de S.

$$-\sum_{h=1}^{q} \gamma_h w_h=\sum_{i=1}^{k} \lambda_i e_i,$$

y de aquí,

$$\sum_{i=1}^{k} \lambda_i e_i+\sum_{h=1}^{q} \gamma_h w_h=0.$$

Pero esta es una combinación lineal de elementos de B_2, que es una base de E_2 y, por tanto, un sistema linealmente independiente en E. Esto significa que

$\lambda_1=0,\ \lambda_2=0,\ldots,\lambda_k=0,\ \gamma_1=0,\ \gamma_2=0,\ldots,\gamma_q=0.$

Luego la combinación lineal inicial se reduce a

$$\sum_{i=1}^{k} \alpha_i e_i + \sum_{j=1}^{p} \beta_j v_j = 0,$$

y esta es una combinación lineal de B_1, que es una base de F_1 y, por tanto, un sistema linealmente independiente en E, lo cual implica que
$\alpha_1=0, \alpha_2=0, \ldots, \alpha_k=0$, $\beta_1=0$, $\beta_2=0, \ldots, \beta_p=0$ y B es linealmente independiente.
Como B genera a F_1+F_2 y es linealmente independiente, B es una base de F_1+F_2.

2) Veamos ahora la relación entre las dimensiones.
$\dim F_1 = k+p$
$\dim F_2 = k+q$
$\dim (F_1+F_2) = k+p+q$
$\dim (F_1 \cap F_2) = k$

$$\begin{aligned} \dim F_1 + \dim F_2 &= k+p+k+q \\ &= (k+p+q)+k \\ &= \dim (F_1+F_2) + \dim (F_1 \cap F_2). \end{aligned}$$

Despejando:

$$\dim (F_1+F_2) = \dim F_1 + \dim F_2 - \dim (F_1 \cap F_2).$$

Además del resultado cuantitativo que hemos obtenido este teorema indica, por el método de su demostración, la forma de construir una base para la suma de dos subespacios.

De esta forma, obtenemos el corolario siguiente:

COROLARIO

Sean E un espacio vectorial sobre K, F_1 y F_2 dos subespacios de dimensión finita de E. Entonces $F_1 \cap F_2 = \{0\}$ si y solo si

$$\dim (F_1+F_2) = \dim F_1 + \dim F_2.$$

Demostración

Si $F_1 \cap F_2 = \{0\}$, por definición, $\dim F_1 \cap F_2 = \dim \{0\} = 0$ y de la fórmula del teorema 1.14 se obtiene el resultado.

Si se cumple la fórmula del corolario entonces por el teorema 1.14:

$$\dim (F_1 \cap F_2) = 0$$

y esto implica
$F_1 \cap F_2 = \{0\}$.

Es importante señalar que si dos subespacios F, F' cumplen

$$\dim E = \dim F + \dim F',$$

esto no es suficiente para afirmar que $E = F \oplus F'$.

Ejemplo

En $\mathbb{R}^3$, consideremos $u=(0,1,0)$ y los subespacios

$$F=\{(x,y,z)/z=0\},$$
$$F'=\mathbb{R}u.$$

En este caso $F'\subset F$ y, por tanto, $F\cap F'\neq\{0\}$ y la suma de F y F' no es directa. Más aún, en este caso

$$F+F'=F.$$

Sin embargo, dim $F=2$ y dim $F'=1$. Luego, se cumple:

$$\dim F+\dim F'=3=\dim \mathbb{R}^3.$$

Es importante tener claro que ninguna relación entre las dimensiones sustituye las condiciones para que un espacio dado sea igual a la suma directa de dos subespacios.

El corolario solamente establece condiciones sobre el carácter de que la suma sea directa y no dice nada sobre el espacio resultado de la suma. Este corolario sugiere investigar una nueva caracterización de la suma directa, a partir de la posibilidad de encontrar una base de $F_1\oplus F_2$.

TEOREMA 1.15

Sean E un espacio vectorial sobre K, F_1 y F_2 dos subespacios de dimensión finita de E. Entonces la suma de F_1 y F_2 es directa si y solo si se puede obtener una base de F_1+F_2 a partir de la unión de una base de F_1 y una base de F_2.

Demostración

Para la demostración vamos a emplear la caracterización de que la suma de F_1 y F_2 es directa si y solo si todo elemento de F_1+F_2 se puede expresar en forma única como un elemento de F_1 más un elemento de F_2.

1) Supongamos que la suma de F_1 y F_2 es directa, y sean $B_1=(v_1,v_2,\dots,v_p)$ $B_2=(w_1,w_2,\dots,w_q)$ bases de F_1 y F_2, respectivamente. Tenemos que probar que
$B=(v_1,v_2,\dots,v_p,w_1,w_2,\dots,w_q)$
es una base de $F_1\oplus F_2$.
Sea $x\in F_1\oplus F_2$. Entonces existen $x_1\in F_1$, $x_2\in F_2$ *únicos* tales que
$x=x_1+x_2$,
y como B_1,B_2 son bases de F_1 y F_2, respectivamente, por el teorema 1.6 existen escalares únicos
$\alpha_1,\alpha_2,\dots,\alpha_p,\beta_1,\beta_2,\dots,\beta_q$
tales que
$x_1=\alpha_1v_1+\alpha_2v_2+\dots+\alpha_pv_p$,
$x_2=\beta_1w_1+\beta_2w_2+\dots+\beta_qw_q$.
Por tanto, todo vector $x\in F_1\oplus F_2$ se expresa en forma *única* en la forma: $x=x_1+x_2=\alpha_1v_1+\alpha v_2+\dots+\alpha_pv_p+\beta_1w_1+\beta_2w_2+\dots+\beta_qw_q$, y por el teorema 1.6, B es una base de $F_1\oplus F_2$.

2) Veamos ahora la demostración en el otro sentido.

Si B es una base de F_1+F_2, entonces para todo $x\in F_1+F_2$ existen, por el teorema 1.6, escalares *únicos:*

$\alpha_1, \alpha_2, \ldots, \alpha_p, \beta_1, \beta_2, \ldots, \beta_q \in K$,

tales que

$x=\alpha_1 v_1+\alpha_2 v_2+\ldots+\alpha_p v_p+\beta_1 w_1+\beta_2 w_2+\ldots+\beta_q w_q.$

Si denotamos

$x_1=\alpha_1 v_1+\alpha_2 v_2+\ldots+\alpha_p v_p \in F_1,$

$x_2=\beta_1 w_1+\beta_2 w_2+\ldots+\beta_q w_q \in F_2,$

como para $x\in F_1+F_2$ los escalares son únicos, entonces los vectores $x_1\in F_1$ y $x_2\in F_2$ son únicos y se cumple:

$x=x_1+x_2$

y, por tanto, la suma de F_1 y F_2 es directa.

Una nueva caracterización de una base, a partir de la suma directa, es la siguiente:

Si E es un espacio vectorial de dimensión finita sobre K y $(e_1, e_2, \ldots, e_n)$ es una base de E, entonces todo vector $x\in E$ se escribe en forma única como:

$$x=\lambda_1 e_1+\lambda_2 e_2+\ldots+\lambda_n e_n,$$

con $\lambda_1, \lambda_2, \ldots, \lambda_n \in K$.

Si analizamos cada uno de los términos de esta suma, vemos que $\lambda_i e_i$ pertenece al subespacio generado por e_i: la recta Ke_i. Por tanto, todo vector del espacio E se puede escribir en forma *única* como la suma de n vectores, cada uno de los cuales pertenece a uno de los subespacios $Ke_1, Ke_2, \ldots, Ke_n$. Esto significa que

$$E=Ke_1 \oplus Ke_2 \oplus \ldots \oplus Ke_n.$$

Inversamente, si existen vectores $e_1, e_2, \ldots, e_n$ no nulos tales que

$$E=Ke_1 \oplus Ke_2 \oplus \ldots \oplus Ke_n$$

esto significa que para todo $x\in E$ existen $x_1\in Ke_1, x_2\in Ke_2, \ldots, x_n\in Ke_n$ *únicos* tales que

$$x=x_1+x_2+\ldots+x_n.$$

Para cada $x_i\in Ke_i$, $i=1,2,\ldots,n$, existe un *único* escalar $\lambda_i\in K$ tal que $x_i=\lambda_i e_i$. Así, tenemos que $x\in E$ se puede expresar como

$$x=\lambda_1 e_1+\lambda_2 e_2+\ldots+\lambda_n e_n$$

en forma única. Por tanto $(e_1, e_2, \ldots, e_n)$ es una base de E.

Obtenemos así el siguiente teorema que caracteriza una base en términos de la suma directa.

TEOREMA 1.16

Sea E un espacio vectorial de dimensión finita, n, sobre K. Un sistema de vectores de E $(e_1,e_2,...,e_n)$, es una base de E si y solo si

$$E=Ke_1 \oplus Ke_2 \oplus ... \oplus Ke_n.$$

Otro resultado que podemos obtener está relacionado con la existencia de un suplementario de un subespacio vectorial.

PROPOSICIÓN 1.9

Sea E un espacio vectorial de dimensión finita. Todo subespacio de E admite un suplementario.

Demostración

Sea F un subespacio de E, si $F=\{0\}$ es trivial que tiene un suplementario, que es el espacio total E. Supongamos, por tanto, que $F\neq\{0\}$.

Como $F\neq\{0\}$ por el teorema 1.12, sea $B_1=(e_1,e_2,...,e_k)$ una base de F; entonces B_1 es un sistema linealmente independiente en E, y por el teorema de completamiento de base, podemos encontrar una base de E:

$$(e_1,e_2,...,e_k,e'_1,e'_2,...,e'_p).$$

Sea $F'=\langle e'_1,e'_2,...,e'_p\rangle$. Entonces, como la unión de una base de F y una base de F' es una base de E (teorema 1.15), tenemos:

$$E=F\oplus F',$$

y F tiene a F' como suplementario.

En la demostración de esta propiedad se puede ver más claramente lo que ya mostramos mediante un ejemplo: el suplementario no es único, pues hemos empleado el teorema de completamiento de base y la forma de completar una base no es única. Sin embargo, todos los suplementarios de un subespacio tienen la misma dimensión, pues si F' es un suplementario de F, $E=F\oplus F'$, entonces por el teorema 1.14.

$$\dim E=\dim F+\dim F'$$
$$\dim F'=\dim E-\dim F$$

y este valor numérico es el mismo para todos los suplementarios de F. Esto nos lleva a introducir la definición siguiente:

DEFINICIÓN 1.16

Sean E un espacio vectorial de dimensión finita, F un subespacio de E. Se denomina *codimensión* de F y se denota por $\mathrm{cod}_K F$, o simplemente por $\mathrm{cod}\ F$, a la dimensión de cualquiera de los suplementarios de F con respecto a E.

Ejemplos

1. En $\mathbb{R}^3$ una recta es de codimensión 2 y un plano es de codimensión 1.

2. En $\mathbb{R}^4$ un plano es de codimensión 2.

En un espacio vectorial E, un subespacio de codimensión 1 se denomina un *hiperplano*.

1.14 Coordenadas de un vector y cambio de base

La aplicación fundamental del concepto de base en un espacio vectorial, es la existencia de las coordenadas de un vector, según vimos en el teorema 1.6. Esta propiedad de las bases adquiere todo su significado cuando se demuestra que siempre es posible encontrar una base en un espacio vectorial de dimensión finita.

Como conclusión de los teoremas que hemos estudiado, en un espacio vectorial de dimensión finita n, dada una base, todo vector se identifica con n números que son sus coordenadas con respecto a dicha base, y todas las operaciones pueden efectuarse con estas coordenadas.

Si $x \in E$ es un vector de un espacio vectorial de dimensión n y $B=(e_1,e_2,...,e_n)$ una base de E, las coordenadas de x en la base B, se escriben como una matriz columna, y entonces hablamos del *vector de coordenadas* de x en la base B y lo denotamos por $[x]_B$.

Es importante observar que ahora es cuando podemos justificar el haber empleado sistemas de vectores; el orden en que se toman los vectores de una base es fundamental porque de lo contrario, las coordenadas de un vector son distintas. Comprobémoslo con un ejemplo.

Ejemplo

En $\mathbb{R}^2$ consideremos las bases siguientes:

$B=((1,1),\ (1,0))$
$B'=((1,0),\ (1,1))$

Dado el vector $v=(2,3) \in \mathbb{R}^2$, su vector de coordenadas en la base B es

$$[v]_B = \begin{pmatrix} 3 \\ -1 \end{pmatrix}$$

y en la base B' es

$$[v]_{B'} = \begin{pmatrix} -1 \\ 3 \end{pmatrix}$$

Las coordenadas de un vector dependen de la base en que estemos trabajando, ya que a un mismo vector le corresponden coordenadas diferentes en bases diferentes.

¿Qué relación existe entre las coordenadas de un vector en dos bases diferentes? Si conocemos las coordenadas de un vector en una base, ¿podemos encontrar las coordenadas en otra base dada?

Vamos a dar respuesta a estas preguntas.

En un espacio vectorial E de dimensión finita n, sean $B=(e_1,e_2,...,e_n)$ y $B'=(e'_1,e'_2,...,e'_n)$ dos bases, v un vector de E con los vectores de coordenadas:

$$[v]_B=\begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \cdot \\ \cdot \\ \cdot \\ \alpha_n \end{pmatrix} \quad \text{y} \quad [v]_{B'}=\begin{pmatrix} \beta_1 \\ \beta_2 \\ \cdot \\ \cdot \\ \cdot \\ \beta_n \end{pmatrix}$$

Esto quiere decir que

$$v=\alpha_1 e_1+\alpha_2 e_2+...+\alpha_n e_n,$$
$$v=\beta_1 e'_1+\beta_2 e'_2+...+\beta_n e'_n.$$

Ahora buscaremos una expresión tal, que a partir de los α_i, podamos encontrar los β_i.

Como B' es una base de E, cada vector de la base B se puede expresar como una combinación lineal de los elementos de la base B'. Así, tenemos:

$$e_j=\sum_{i=1}^{n} a_{ij}\, e'_i,$$

donde $a_{ij}\in K$.

Entonces:

$$v=\sum_{j=1}^{n}\alpha_j e_j=\sum_{j=1}^{n}\alpha_j\Big(\sum_{i=1}^{n}a_{ij}\,e'_i\Big)$$
$$=\sum_{j=1}^{n}\sum_{i=1}^{n}a_{ij}\alpha_j e'_i=\sum_{i=1}^{n}\Big(\sum_{j=1}^{n}a_{ij}\alpha_j\Big)e'_i$$
$$=\Big(\sum_{j=1}^{n}a_{1j}\alpha_j\Big)\,e'_1+\Big(\sum_{j=1}^{n}a_{2j}\alpha_j\Big)e'_2+...+\Big(\sum_{j=1}^{n}a_{nj}\alpha_j\Big)e'_n$$
$$=\beta_1 e'_1+\beta_2 e'_2+...+\beta_n e'_n$$

Como las coordenadas de un vector en una base son únicas, entonces

$$\beta_1=\sum_{j=1}^{n}a_{1j}\alpha_j,\ \beta_2=\sum_{j=1}^{n}a_{2j}\alpha_j,...,\beta_n=\sum_{j=1}^{n}a_{nj}\alpha_j,$$

es decir,

$$\beta_i=\sum_{j=1}^{n}a_{ij}\alpha_j,\quad i=1,2,...,n.$$

Esta relación puede expresarse en notación matricial.

Sea $A=(a_{ij})$. Entonces escribimos:

$$\begin{pmatrix}\beta_1\\ \beta_2\\ \cdot\\ \cdot\\ \cdot\\ \beta_n\end{pmatrix}=\begin{pmatrix}a_{11} & a_{12} & \dots & a_{1n}\\ a_{21} & a_{22} & \dots & a_{2n}\\ \dots & \dots & \dots & \dots\\ \\ \\ a_{n1} & a_{n2} & \dots & a_{nn}\end{pmatrix}\begin{pmatrix}\alpha_1\\ \alpha_2\\ \cdot\\ \cdot\\ \cdot\\ \alpha_n\end{pmatrix}$$

$$[v]_{B'}=A\ [v]_B.$$

A es una matriz que permite, conocidas las coordenadas de un vector en la base B, calcular las coordenadas del mismo vector en la base B'.

La matriz A se denomina la *matriz de cambio de coordenadas* de la base B a la base B' y se obtiene expresando los elementos de la base B como combinación lineal de los elementos de la base B' y escribiendo esos coeficientes como columnas en la matriz. La forma de construir esta matriz garantiza que es la única que establece esta relación.

Ejemplos

1. Consideremos en $\mathbb{R}^2$ las bases:

   $B=((1,0),\ (0,1))$,
   $B'=((1,1),\ (-1,1))$.

   Vamos a calcular la matriz de cambio de coordenadas de la base B a la base B'. Para ello procedemos a escribir $(1,0)$ y $(0,1)$ como combinación lineal de B':

$$(1,0)=\frac{1}{2}\ (1,1)-\frac{1}{2}\ (-1,1)$$

$$(0,1)=\frac{1}{2}\ (1,1)+\frac{1}{2}\ (-1,1).$$

La primera columna de la matriz son los coeficientes de $(1,0)$ en B' y la segunda, los coeficientes de $(0,1)$ en B'.

$$A=\begin{pmatrix}\frac{1}{2} & \frac{1}{2}\\ -\frac{1}{2} & \frac{1}{2}\end{pmatrix}$$

Si ahora tomamos el vector $(3,7)$, cuyo vector de coordenadas en B es $\begin{pmatrix}3\\7\end{pmatrix}$, y queremos encontrar sus coordenadas en B' tenemos que aplicar la matriz A.

$$\begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix} \begin{pmatrix} 3 \\ 7 \end{pmatrix} = \begin{pmatrix} 5 \\ 2 \end{pmatrix}.$$

A B

El vector de coordenadas de (3,7) en la base B' es (5,2).

2. En $\mathbb{R}^3$ consideremos las bases:
 $B=((1,0,0), (0,1,0), (0,0,1))$,
 $B'=((1,1,1), (1,1,0), (1,0,0))$.
 Busquemos la matriz de cambio de coordenadas.
 $(1,0,0) = 0\ (1,1,1) + 0\ (1,1,0) + \ (1,0,0)$
 $(0,1,0) = 0\ (1,1,1) + \ (1,1,0) - \ (1,0,0)$
 $(0,0,1) = \ (1,1,1) - \ (1,1,0) + 0\ (1,0,0)$.
 La matriz que buscamos es:

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix}.$$

$P_B \longrightarrow B'$

$Id_B \Rightarrow B$

$P_{B'} \Rightarrow B$

Si queremos conocer las coordenadas del vector $(1,2,-1)$ en la base B', calculamos el producto:

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix} \begin{pmatrix} 1 \\ 2 \\ -1 \end{pmatrix} = \begin{pmatrix} -1 \\ 3 \\ -1 \end{pmatrix}$$

El vector de coordenadas de $(1,2,-1)$ en la base B' es $(-1,3,-1)$.
Tenemos, pues, la matriz de cambio de coordenadas de la base B a la base B', que establece la relación:
$[v]_{B'} = A[v]_B$
para todo vector $v \in E$.

DEFINICIÓN 1.17

Sean E un espacio vectorial de dimensión n, B y B' dos bases de E. Se denomina *matriz de cambio de coordenadas* de la base B a la base B' a la matriz cuya i-ésima columna está formada por las coordenadas del i-ésimo vector de la base B en la base B'.

PROPOSICIÓN 1.10

Sean E un espacio vectorial de dimensión n, B y B' dos bases de E, A la matriz de cambio de coordenadas de la base B a la base B'. Entonces A es inversible.

Demostración

A es una matriz cuadrada de orden n. Para demostrar que A es inversible tenemos que probar que

det $A \neq 0$

o que

$$\operatorname{rang} A = n.$$

Sea $v \in E$. Analicemos la relación que existe entre $[v]_B$ y $[v]_{B'}$. Sabemos que

$$[v]_{B'} = A[v]_B.$$

Vamos a interpretar esta relación como un sistema de ecuaciones. Supongamos $[v]_{B'}$ conocido; el problema es hallar $[v]_B$ como incógnita. Entonces la relación

$$[v]_{B'} = A[v]_B$$

es un sistema de n ecuaciones con n incógnitas, que tiene solución única pues las coordenadas de un vector en una base son únicas. Por la teoría de los sistemas de ecuaciones lineales, un sistema de este tipo tiene solución única si y solo si $\operatorname{rang} A = n$, y esto significa que A es inversible.

Si A es la matriz de cambio de coordenadas de la base B a la base B', cumple la relación:

$$[v]_{B'} = A[v]_B,$$

y como A es inversible, tenemos:

$$[v]_B = A^{-1}[v]_{B'}.$$

Por tanto, A^{-1} es la matriz de cambio de coordenadas de la base B' a la base B.

Observemos que A^{-1} se obtiene expresando los elementos de la base B' como combinación lineal de los elementos de la base B y los coeficientes de la combinación lineal forman las columnas de la matriz.

Ejemplos

1. En $\mathbb{R}^2$ consideremos las bases:
   $B = ((1,0), (0,1)) = (e_1, e_2)$,
   $B' = ((1,1), (-1,1))$.
   La matriz de cambio de coordenadas de B' a B se obtiene a partir de
   $(1,1) = e_1 + e_2$
   $(-1,1) = -e_1 + e_2$.
   Por tanto, la matriz es:

$$\begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}.$$

2. En $\mathbb{R}^3$ consideremos las bases:
   $B = ((1,0,0), (0,1,0), (0,0,1)) = (e_1, e_2, e_3)$,
   $B' = ((1,1,1), (1,1,0), (1,0,0))$.

La matriz de cambio de coordenadas de B' a B se obtiene a partir de

$(1,1,1)=e_1+e_2+e_3$,
$(1,1,0)=e_1+e_2$,
$(1,0,0)=e_1$.

Por tanto, la matriz es:

$$\begin{pmatrix} 1 & 1 & 1 \\ 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$$

La forma en que están planteados estos ejemplos no es casual. Por lo general, al trabajar en K^n los vectores se dan como n números ordenados:

$$(a_1, a_2, \dots, a_n),$$

y estos números son las coordenadas del vector en la base $((1,0,\dots,0),(0,1,0,\dots,0),\dots,(0,0,\dots,1))$, que denotaremos por $(e_1,e_2,\dots,e_n)$. El vector e_i es el vector de K^n que tiene todas sus componentes nulas, excepto la i-ésima componente que es 1. Los n vectores que así obtenemos, forman una base de K^n que usualmente recibe el nombre de *base canónica* de K^n.

Al trabajar con K^n se supone que estamos trabajando en la base canónica y es fácil encontrar las coordenadas de un vector cualquiera en dicha base ya que son sus componentes. Es por eso que al plantearnos otra base surge un problema:

es más fácil escribir la nueva base como combinación lineal de la canónica que escribir la canónica en función de la nueva base; es decir, en lugar de buscar la matriz de cambio de coordenadas de la base canónica a la nueva base, buscamos su inversa. En algunos casos será mejor hacer precisamente esto: encontrar esta matriz y después calcular su inversa.

Nota. Algunos autores emplean la siguiente notación: si B y B' son dos bases de un espacio vectorial, a la matriz que se obtiene al expresar los elementos de B' como combinación lineal de los elementos de B se denomina *matriz de cambio de base de B a B'*, y a la inversa de ésta, que es la que se obtiene al expresar los elementos de la base B como combinación lineal de los elementos de la base B', se denomina *matriz de cambio de coordenadas de B a B'*.
Según esta notación hay dos matrices:

Matriz de cambio de base de B a B'
Matriz de cambio de coordenadas de B a B'.

Una matriz es la inversa de la otra. Pero aquí se forma un "pequeño trabalenguas"

La matriz de cambio de base de B a B' es la matriz de cambio de coordenadas de B' a B, y es la inversa de la matriz de cambio de coordenadas de B a B', la cual es a su vez la matriz de cambio de base de B' a B.

La formulación de estas relaciones resulta algo complicada, al menos, para comprenderla inicialmente.
Mantendremos solamente la terminología de matriz de cambio de coordenadas.

En resumen:

Para obtener la matriz de cambio de coordenadas de la base B a la base B' tenemos dos métodos.

1ro. Se escriben los elementos de la base B como combinación lineal de los vectores de la base B', y los coeficientes que se obtienen se colocan en columnas, por su orden. La matriz A que se obtiene cumple la relación:

$$[v]_{B'}=A[v]_B.$$

2do. Se escriben los elementos de la base B' como combinación lineal de los elementos de la base B, y los coeficientes que se obtienen se colocan en columnas, por su orden. La matriz P que se obtiene se invierte y P^{-1} cumple:

$$[v]_{B'}=P^{-1}\ [v]_B.$$

La relación que hemos encontrado entre el cambio de coordenadas y una matriz inversible se completa con la proposición siguiente:

PROPOSICIÓN 1.11

Sean E un espacio vectorial sobre K, de dimensión n, P una matriz inversible de orden n con coeficientes en K, y sea B una base de E. Entonces existe una única base B' de E tal que para todo vector $v \in E$ se cumple:

$$[v]_{B'}=P^{-1}\ [v]_B$$

$$[v]_B=P\ [v]_{B'}.$$

Es decir, P es la matriz de cambio de coordenadas de la base B' a la B y, por tanto, P^{-1} es la matriz de cambio de coordenadas de la base B a la B'.

Demostración

Supongamos que $P=(a_{ij})$ y sea $B=(e_1,e_2,...,e_n)$ la base dada en E.

Si analizamos la relación que cumplían las bases, cuando estudiamos el problema del cambio de base, podemos darnos cuenta de cuál es el camino a seguir.

Vamos a construir los vectores $e'_1,e'_2,...,e'_n$ mediante la fórmula:

$$e'_j=\sum_{i=1}^{n} a_{ij}\ e_i,\ j=1,2,...,n.$$

Si los vectores así obtenidos forman una base de E, que llamaremos B', entonces la matriz P cumple las relaciones que plantea el enunciado.

Vamos a demostrar que el sistema de vectores $B'=(e'_1,e'_2,...,e'_n)$ es una base de E.

Como E es de dimensión n, basta demostrar que el sistema B' es un generador o que el sistema B' es linealmente independiente. Demostremos que B' genera a E.

Sea $P^{-1}=(\beta_{ij})$. Vamos a combinar adecuadamente los vectores del sistema B' con los coeficientes de la matriz P^{-1}.

$$\sum_{k=1}^{n} \beta_{kj}\, e'_k = \sum_{k=1}^{n} \beta_{kj} \left(\sum_{i=1}^{n} a_{ik}\, e_i \right)$$

$$= \sum_{k=1}^{n} \sum_{i=1}^{n} \beta_{kj}\, a_{ik}\, e_i$$

$$= \sum_{i=1}^{n} \left(\sum_{k=1}^{n} a_{ik}\, \beta_{kj} \right) e_i$$

$$= \sum_{i=1}^{n} \delta_{ij}\, e_i = e_j.$$

δ_{ij} es el delta de Kronecker, porque $P=(a_{ij})$ es la inversa de $P^{-1}=(\beta_{ij})$. Esto quiere decir que cada elemento de B se puede expresar como combinación lineal de los elementos de B' y, por tanto, el subespacio generado por B' contiene al subespacio generado por B. Como B es una base, genera a todo el espacio y como conclusión B' es un generador, lo cual demuestra que B' es una base y la matriz P es la matriz de cambio de coordenadas de la base B' a la B; luego B' satisface las relaciones que plantea el enunciado de la proposición.

La base B' es única porque su construcción se realiza a partir de combinaciones lineales de una base, y el teorema 1.6 garantiza la unicidad.

Ejemplo

La matriz de coeficientes reales

$$P = \begin{pmatrix} 1 & 0 & 2 \\ -1 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

es una matriz inversible, y su inversa es:

$$P^{-1} = \begin{pmatrix} 1 & 0 & -2 \\ 1 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Si consideramos en el espacio $\mathbb{R}^3$ la base canónica

$(e_1=(1,0,0),\ e_2=(0,1,0),\ e_3=(0,0,1))$,

entonces, según la proposición 1.11, los vectores

$a_1=(1,-1,0),\ a_2=(0,1,0),\ a_3=(2,-2,1)$

forman una base de $\mathbb{R}^3$. La matriz de cambio de coordenadas de la base canónica a la base (a_1,a_2,a_3) es la matriz P^{-1}.

Es probable que al estudiar este problema de cambio de bases, que es algo complicado, algunos estudiantes se pregunten: ¿qué necesidad hay de cambiar la base si con la base canónica en K^n es muy fácil hallar las coordenadas de un vector?

En este momento no es posible dar un ejemplo no muy complejo que muestre la necesidad, en ciertos casos, de seleccionar bases "adaptadas" al problema. Más adelante veremos ejemplos en los cuales, al escoger una base en cierta forma, todos los cálculos que hay que realizar se simplifican notablemente. Sin embargo, el estudiante conoce un problema de la Geometría analítica, donde el hacer un cambio de coordenadas simplifica la expresión del problema; nos referimos a la reducción de la ecuación de una cónica a su forma canónica.

En general, en el plano, espacio $\mathbb{R}^2$, una ecuación de la forma $Ax^2+Bxy+Cy^2+F=0$ representa una cónica, donde (x, y) son las coordenadas de un punto en un sistema de coordenadas cartesiano. Podemos interpretar esta situación de la manera siguiente:

En el espacio vectorial $\mathbb{R}^2$ con la base $(\vec{i}, \vec{j})$, si los componentes del vector $\vec{v}=x\vec{i}+y\vec{j}$ satisfacen la ecuación $Ax^2+Bxy+Cy^2+F=0$, entonces el extremo del vector $\vec{v}$ describe una cónica.

En este problema se buscaba un nuevo sistema de coordenadas, por tanto, una nueva base del espacio $\mathbb{R}^2$, de forma tal que en el nuevo sistema de coordenadas, la ecuación fuese de la forma más simple posible. En este caso, al cambiar las coordenadas, buscamos una nueva base (a_1, a_2), de $\mathbb{R}^2$ tal que si $v=x'a_1+y'a_2$, entonces la ecuación de la curva se escribe de una forma más simple, por ejemplo:

$$A'x'^2+B'x'y'+C'y'^2+F'=0$$

donde uno de los coeficientes A', B', C' es cero.

Aquí tenemos un caso donde al hacer un cambio de base y, por tanto, un cambio de coordenadas, se simplifica la expresión del problema. Esto quiere decir que para estudiar la curva representada por la ecuación, es más conveniente emplear la base (a_1, a_2), que está adaptada al problema concreto de esta curva, y no la base usual, que no está relacionada con este problema.

1.15 Bases de un subespacio de K^n

De los resultados obtenidos en relación con las bases, sabemos que siempre existe una base de un espacio vectorial. El problema que nos planteamos ahora es: ¿Cómo proceder en la práctica para encontrar una base en el caso de un subespacio de K^n?

Vamos a analizar los dos casos que hemos planteado para los subespacios de K^n.

1ro. El subespacio está dado como la solución de un sistema de ecuaciones lineales homogéneo.

2do. El subespacio está dado a partir de un generador.

Antes de entrar a analizar ambos casos, precisaremos un concepto que debemos utilizar y que ya ha sido empleado con las matrices: el *rango.*

Rango de un sistema de vectores

El concepto *rango* es muy similar al que ya conoce el estudiante para las matrices.

DEFINICIÓN 1.18

Sean E un espacio vectorial sobre K, S un sistema de vectores de E. Se denomina *rango* de S y se denota por rang S a la dimensión del subespacio generado por S.

Es evidente, a partir de los conceptos que hemos estudiado, que también podríamos decir:

El rango de un sistema finito de vectores es el número máximo de vectores linealmente independientes que se pueden extraer de él.

En cursos anteriores estudiamos el concepto *rango de una matriz.* A partir de la definición dada -ver el libro de A. Kurosh -podemos relacionar este concepto con el de rango de un sistema de vectores mediante lo que se conoce en la literatura como el espacio de los vectores columna de una matriz A.

Sea A una matriz de orden $m \times n$, y llamemos $a_1, a_2, \ldots, a_n$ a los n vectores de K^m que podemos formar con sus columnas. Sea $\langle A \rangle$ el subespacio de K^m generado por los vectores $a_1, a_2, \ldots, a_n$:

$$\langle A \rangle = \langle a_1, a_2, \ldots, a_n \rangle,$$

donde $a_i = (\alpha_{1i}, \alpha_{2i}, \ldots, \alpha_{mi})$ y $A = (\alpha_{ij})$.

Está claro, entonces, que el rango de la matriz A es igual al rango del sistema de vectores $(a_1, a_2, \ldots, a_n)$.

Pero podemos trabajar en sentido contrario, o sea, partir de un sistema de n vectores de K^m y con sus $m \cdot n$ componentes formar una matriz A de orden $m \times n$ tal, que cada una de las columnas de la matriz esté formada por las m componentes de uno de los vectores del sistema. En este caso el rango de la matriz A coincide con el rango del sistema de vectores.

Ejemplos

1. Sea A la matriz de coeficientes reales

$$A = \begin{pmatrix} 1 & 2 & 4 \\ 0 & 3 & 9 \\ 2 & -1 & -7 \\ -1 & 1 & 5 \end{pmatrix}$$

y consideremos el sistema de vectores de $\mathbb{R}^4$:

$(a_1 = (1,0,2,-1),\ a_2 = (2,3,-1,1),\ a_3 = (4,9,-7,5))$.

Entonces

rang A = rang $(a_1, a_2, a_3) = 3$.

2. Consideremos el sistema de vectores:

$(a_1 = (1,2,2,1),\ a_2 = (0,2,0,1),\ a_3 = (-2,0,-4,0))$

y formemos la matriz

$$A = \begin{pmatrix} 1 & 0 & -2 \\ 2 & 2 & 0 \\ 2 & 0 & -4 \\ 1 & 1 & 0 \end{pmatrix}$$

Entonces

rang (a_1, a_2, a_3) = rang $A = 2$.

Esta equivalencia entre rango de un sistema de vectores de K^n y rango de la matriz que se puede formar con ellos, será muy útil para calcular la dimensión del subespacio generado por un conjunto de vectores. Como el rango del conjunto de vectores es igual a la dimensión del subespacio generado por ellos, y este número es igual al rango de la matriz columna de los vectores, podemos calcularlo por cualquiera de las vías posibles. Por lo general, la vía más empleada es mediante la matriz, porque por medio de transformaciones elementales, reducimos la matriz A, a la forma escalonada, y en esta forma es muy fácil conocer el rango.

Ejemplo

En $\mathbb{R}^5$ consideremos los vectores:

$a_1 = (1,1,0,2,0)$, $a_2 = (2,2,0,4,0)$, $a_3 = (0,-1,1,1,0)$, $a_4 = (3,-1,4,10,0)$, $a_5 = (0,0,0,1,1)$.

Queremos conocer la dimensión del subespacio de $\mathbb{R}^5$ generado por $(a_1, a_2, a_3, a_4, a_5)$.

Formamos la matriz que tiene a a_1, a_2, a_3, a_4, a_5 por columnas y la llevamos a la forma escalonada a partir de transformaciones elementales.

$$A = \begin{pmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

La forma escalonada de A es:

$$\begin{pmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Aquí vemos que el rango, que es igual al número de filas no nulas, es 3. Entonces

$$S=\langle a_1, a_2, a_3, a_4, a_5\rangle$$
$$\dim S=3.$$

Hasta aquí tenemos un método para encontrar la dimensión, pero ¿cómo buscamos una base en este caso?

La respuesta se encuentra mediante el llamado teorema del rango. (Ver *Álgebra Superior* de A.G. Kurosh, página 69.)

El orden superior de los menores, diferentes de cero, de una matriz A, es igual al rango de esta matriz.

La demostración se basa en que si en la matriz A tomamos las columnas que forman un menor de orden máximo no nulo, todas las otras columnas de la matriz se pueden expresar como combinación lineal de ellas.

Hablando en términos de vectores, si consideramos los vectores que forman las columnas de un menor de orden máximo no nulo de la matriz A, todos los otros vectores del sistema se pueden expresar como combinación lineal de ellos. Por tanto, los vectores que forman las columnas de ese menor de orden máximo no nulo forman un sistema generador del espacio de los vectores columnas de la matriz A, y su número es igual a la dimensión de dicho espacio; luego forman una base de este.

Como ninguno de los resultados sobre los valores de los menores de una matriz se alteran por realizar operaciones elementales, generalmente se toma la matriz que tiene por columnas a los vectores dados y mediante operaciones elementales se reduce a la forma escalonada. En esta forma escalonada se selecciona un menor de orden máximo no nulo. Entonces las columnas de la matriz original correspondientes a aquellas que en la matriz escalonada forman el menor no nulo, constituyen la base del espacio dado.

Ejemplo

En el ejemplo anterior teníamos cinco vectores de $\mathbb{R}^5$ y la matriz A.

$$A=\begin{pmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

Su forma escalonada es:

$$\begin{pmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Las columnas 1, 3 y 5 forman un menor:

$$\begin{vmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

que es de orden máximo no nulo. Esto significa que las columnas 1, 3 y 5 de la matriz A forman una base del espacio S generado por $(a_1, a_2, a_3, a_4, a_5)$:

$$(a_1 = (1,1,0,2,0),\ a_3 = (0,-1,1,1,0),\ a_5 = (0,0,0,1,1)).$$

Observemos que podemos encontrar *otras* bases por este método. Por ejemplo, otro menor no nulo lo pueden formar las columnas 2, 4, 5:

$$\begin{vmatrix} 2 & 3 & 0 \\ 0 & 4 & 0 \\ 0 & 0 & 1 \end{vmatrix}$$

y, por tanto (a_2, a_4, a_5) es también una base de S.

De esta forma podemos resolver el problema de encontrar una base y la dimensión de un subespacio de K^n que esté dado como el subespacio generado por un sistema (finito) de vectores. Pero también podemos tener un subespacio de K^n, dado como el espacio solución de un sistema de ecuaciones lineales homogéneo $AX=0$. ¿Cómo procedemos en este caso?

Vamos a dar un método de trabajo, que no vamos a justificar pues será mucho más fácil, simple y comprensible hacerlo al finalizar el próximo capítulo.

Supongamos que tenemos una matriz A de orden $m \times n$ y rang $A=k$, que es la matriz del sistema homogéneo de ecuaciones lineales $AX=0$. Sea S el espacio solución de este sistema. Entonces

$$\dim S = n - \text{rang } A = n-k$$

Para encontrar una base, hay que encontrar $n-k$ vectores que generen a S. Para ello llevamos el sistema, por el método de Gauss, a la forma escalonada y despejamos las variables dependientes.

Supongamos que las variables independientes son:

$$x_{k+1}, x_{k+2}, \ldots, x_n.$$

Tenemos $n-k$ variables independientes.

Para cada conjunto de valores que demos a estas variables, obtendremos un conjunto de valores de las variables dependientes. Para obtener los $n-k$ vectores que buscamos, damos a las variables independientes las $n-k$ conjuntos de valores siguientes:

$$\begin{array}{ll}
1) & x_{k+1}=1,\ x_{k+2}=0,\ x_{k+3}=0, \ldots,\ x_n=0 \\
2) & x_{k+1}=0,\ x_{k+2}=1,\ x_{k+3}=0, \ldots,\ x_n=0 \\
3) & x_{k+1}=0,\ x_{k+2}=0,\ x_{k+3}=1, \ldots,\ x_n=0 \\
 & \cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots \\
n-k) & x_{k+1}=0,\ x_{k+2}=0,\ x_{k+3}=0, \ldots,\ x_n=1
\end{array}$$

De esta forma se obtienen $n-k$ vectores del espacio S que forman una base de este.

Ejemplo

Consideremos el sistema de ecuaciones lineales $AX=0$, donde A es la matriz:

$$\begin{pmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

y llamemos S al espacio solución del sistema, que es un subespacio de $\mathbb{R}^5$.

Al reducir el sistema de ecuaciones a la forma escalonada, obtenemos el sistema equivalente:

$$\begin{cases} x_1+2x_2+3x_4=0 \\ \qquad x_3+\ 4x_4=0 \\ \qquad\qquad x_5=0 \end{cases}$$

El rango de la matriz del sistema es 3, por tanto, hay dos variables independientes y la dimensión de S es 2.

Hallemos ahora una base de S. Tomemos como variables independientes x_2 y x_4. Al expresar las otras variables en función de x_2 y x_4, obtenemos:

$$x_1=-2x_2\ -3x_4$$
$$x_3=-4x_4$$
$$x_5=0$$

Damos a x_2 y x_4 los dos pares de valores:

$$x_2=1,\ x_4=0;\ x_2=0,\ x_4=1$$

y obtenemos los vectores de S:

$$v_1=(-2,1,0,0,0),\ v_2=(-3,0,-4,1,0).$$

Estos dos vectores forman una base de S.

Podemos generalizar más este método, pero no vamos a ganar mucho desde el punto de vista práctico. En general, hay que asignar a las $n-k$ variables independientes, $n-k$ conjuntos de valores, tales que la matriz cuadrada de orden $(n-k)$ que forman estos valores tenga determinante diferente de cero. Así, en el ejemplo dado, en lugar de asignar a x_2 y x_4 los valores que forman la matriz:

$$\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix},$$

podemos asignarles, por ejemplo, cualquiera de los valores que dan lugar a las matrices siguientes:

$$\begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix}, \begin{pmatrix} -1 & 2 \\ -3 & -1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 2 & 3 \end{pmatrix}$$

y así obtendremos bases diferentes para el mismo subespacio.

No obstante, recomendamos hacer lo sugerido inicialmente, por ser más simple.

Ya podemos resolver el problema de dar una base y la dimensión de un subespacio de K^n para cualquiera de los dos casos que nos interesan:

1ro. El subespacio está dado como la solución de un sistema de ecuaciones lineales homogéneo.

2do. El subespacio está dado a partir de un generador.

Un problema típico es:

Dados dos subespacios V y W de K^n, hallar una base y la dimensión de V, W, $V \cap W$ y $V+W$.

Este problema se presenta en tres variantes:

1ra. Los dos subespacios están dados como subespacios solución de sistemas de ecuaciones lineales homogéneos.

2da. Los dos subespacios están dados como subespacios generados por un conjunto de vectores.

3ra. Uno de los espacios es el espacio solución de un sistema de ecuaciones lineales homogéneo, y el otro es el subespacio generado por un conjunto de vectores.

La tercera variante se reduce a una de las otras dos, pues o bien al espacio solución del sistema de ecuaciones le hallamos una base y ya lo podemos expresar como el subespacio generado por esa base, y por tanto, se trata de la segunda variante o para el espacio que está dado por un sistema de generadores, hallamos un sistema de ecuaciones, según el procedimiento que ya vimos, y por tanto, se trata de la primera variante.

Veamos mediante ejemplos cómo podemos trabajar en las dos primeras variantes.

Obsérvese que verdaderamente cada una de las variantes se puede reducir a la otra. De esta forma el estudiante puede escoger la que le resulte más cómoda para trabajar, según sus habilidades y según el problema.

Ejemplos

1. Sean V y W los subespacios de $\mathbb{R}^5$ determinados por la solución de los sistemas de ecuaciones:

$$V\{(x_1,x_2,x_3,x_4,x_5) \in \mathbb{R}^5/x_1-x_2-x_3=0 \text{ y } 3x_1-x_2-x_4+x_5=0\},$$
$$W=\{(x_1,x_2,x_3,x_4,x_5) \in \mathbb{R}^5/x_1+x_2-x_3=0 \text{ y } 3x_1+5x_2-x_4+x_5=0\}.$$

Vamos a hallar primero una base y la dimensión de V y de W.

Para V:

Aplicando el método de Gauss al sistema que define a V, obtenemos:

$$\begin{cases} x_1-x_2-x_3=0 \\ 2x_2+3x_3-x_4+x_5=0 \end{cases}$$

El sistema tiene rango 2 y, por tanto, V tiene dimensión 3.
Para encontrar una base tomamos a x_2, x_3 y x_5 como variables independientes, y tenemos:

$$x_1=x_2+x_3$$
$$x_4= 2x_2+3x_3+x_5$$

Dando sucesivamente los valores 1 y 0 a estas variables, obtenemos los vectores:

$e_1=(1,1,0,2,0)$, $e_2=(1,0,1,3,0)$, $e_3=(0,0,0,1,1)$

que constituyen una base de V.

Para W:

Al aplicar el método de Gauss al sistema que define a W, obtenemos:

$$\begin{cases} x_1+x_2-x_3=0 \\ 2x_2+3x_3-x_4+x_5=0 \end{cases}$$

El sistema tiene rango 2 y, por tanto, W tiene dimensión 3. Tomamos como variables independientes a x_1, x_2, x_4, y para encontrar la base damos los valores apropiados obtenidos a partir de

$$x_3=x_1+x_2,$$
$$x_5=-3x_1-5x_2+ x_4$$

Por tanto, la base de W está formada por los vectores:

$e_1'=(1,0,1,0,-3)$, $e_2'=(0,1,1,0,-5)$, $e_3'=(0,0,0,1,1)$.

Para $V\cap W$:

$V\cap W$ es el subespacio solución del sistema de ecuaciones que se obtiene al reunir los sistemas de ecuaciones que satisfacen a V y W. Por tanto, $V\cap W$ es el subespacio solución del sistema:

$$\begin{cases} x_1-x_2-x_3=0 \\ 3x_1-x_2-x_4+x_5=0 \\ x_1+x_2-x_3=0 \\ 3x_1+5x_2-x_4+x_5=0 \end{cases}$$

Aplicando el método de Gauss al sistema, obtenemos:

$$\begin{cases} x_1-x_2-x_3=0 \\ \qquad\quad x_2=0 \\ 3x_3-x_4+x_5=0 \end{cases}$$

Este sistema tiene rango 3, por tanto, $V\cap W$ tiene dimensión 2. Tomamos x_3 y x_5 como variables independientes y despejamos:

$x_1=x_3$
$x_2=0$
$x_4=3x_3+x_5$.

Dando los valores apropiados a las variables independientes, obtenemos una base de $V\cap W$ formada por los vectores:

$\bar{e}_1=(1,0,1,3,0)$, $\bar{e}_2=(0,0,0,1,1)$.

Para $V+W$:

Podemos calcular la dimensión de $V+W$ a partir de la fórmula:

$$\begin{aligned} \dim(V+W) &= \dim V+\dim W-\dim(V\cap W) \\ &= 3+3-2=4. \end{aligned}$$

Para encontrar una base de $V+W$ procedemos según la demostración de esta fórmula, o sea, a partir de una base de $V\cap W$ completamos una base de V y otra de W.
Para ello veamos para qué valores de las variables independientes del sistema de V se obtienen los vectores $\bar{e}_1$ y $\bar{e}_2$.
La relación de las variables en V es:

$x_1=x_2+x_3$
$x_4=2x_2+3x_3+x_5$

El vector $\bar{e}_1$ se obtiene para $x_2=0, x_3=1$, $x_5=0$ y el $\bar{e}_2$, para $x_2=0$, $x_3=0, x_5=1$.
Para completar una base damos a las variables independientes un sistema de valores que forme un determinante no nulo con los otros dos; en este caso, $x_2=1$, $x_3=0$, $x_5=0$, y obtenemos:

$\bar{e}_3=(1,1,0,2,0)$.

La base de V completada a partir de la de $V\cap W$ es:

$(\bar{e}_1=(1,0,1,3,0)$, $\bar{e}_2=(0,0,0,1,1)$, $\bar{e}_3=(1,1,0,2,0))$.

(Esta base podíamos haberla obtenido por simple inspección de la base para V.)
Veamos ahora la relación entre las componentes de los vectores de W

y qué valores de las variables independientes dan estos vectores.

$x_3=x_1+x_2$
$x_5=-\ 3x_1-5x_2+x_4$

El vector $\bar{e}_1$ se obtiene para $x_1=1$, $x_2=0$, $x_4=3$ y el $\bar{e}_2$, para $x_1=0$, $x_2=0$, $x_4=1$.
Para completar una base damos a las variables independientes un sistema de valores que forme un determinante no nulo; en este caso es fácil comprobar que los valores $x_1=0$, $x_2=1$, $x_4=0$, satisfacen esta condición, y así obtenemos:

$\bar{e}_4=(0,1,1,0,-5)$

Por tanto, la base de $V+W$ está formada por:

$\bar{e}_1=(1,0,1,3,0)$, $\bar{e}_2=(0,0,0,1,1)$
$\bar{e}_3=(1,1,0,2,0)$, $\bar{e}_4=(0,1,1,-5)$.

Supongamos que quisiéramos obtener un sistema de ecuaciones que represente a $V+W$. Empleando el procedimiento que estudiamos y aplicando transformaciones:

$$\begin{pmatrix} 1 & 0 & 1 & 0 & x \\ 0 & 0 & 1 & 1 & y \\ 1 & 0 & 0 & 1 & z \\ 3 & 1 & 2 & 0 & s \\ 0 & 1 & 0 & -5 & t \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 0 & x \\ 0 & 0 & 1 & 1 & y \\ 0 & 0 & -1 & 1 & z-x \\ 0 & 1 & -1 & 0 & s-3x \\ 0 & 1 & 0 & -5 & t \end{pmatrix}$$

$$\begin{pmatrix} 1 & 0 & 1 & 0 & x \\ 0 & 1 & 0 & -5 & t \\ 0 & 0 & 1 & 1 & y \\ 0 & 0 & -1 & 1 & z-x \\ 0 & 0 & -1 & 5 & s-3x-t \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & 1 & 0 & x \\ 0 & 1 & 0 & -5 & t \\ 0 & 0 & 1 & 1 & y \\ 0 & 0 & 0 & 2 & z-x+y \\ 0 & 0 & 0 & 6 & s-3x-t+y \end{pmatrix} \sim$$

$$\sim \begin{pmatrix} 1 & 0 & 1 & 0 & x \\ 0 & 1 & 0 & -5 & t \\ 0 & 0 & 1 & 1 & y \\ 0 & 0 & 0 & 2 & z-x+y \\ 0 & 0 & 0 & 0 & s-t-2y-3z \end{pmatrix}$$

El sistema de ecuaciones está dado por la ecuación

$s-t-2y-3z=0$

y $V+W$ es el espacio solución de esta ecuación.
Otra forma de encontrar una base de $V+W$ es a partir de la siguiente propiedad, que se encuentra como ejercicio al final de este capítulo.

Si S_1 es un sistema generador de V y S_2 un sistema generador de W, entonces $S_1 \cup S_2$ es un sistema generador de $V+W$

A partir de esta propiedad podemos trabajar de la manera siguiente:

1) (e_1, e_2, e_3) genera a V porque es una base de V.
2) (e'_1, e'_2, e'_3) genera a W porque es una base de W.
   Por tanto, $(e_1, e_2, e_3, e'_1, e'_2, e'_3)$ genera a $V+W$, y tenemos que buscar una base de $V+W$ dentro de este sistema generador.
   Formamos una matriz cuyas columnas sean las componentes de los vectores $e_1, e_2, e_3, e'_1, e'_2, e'_3$, y obtenemos:

$$\begin{pmatrix} 1 & 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 2 & 3 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & -3 & -5 & 1 \end{pmatrix}$$

Reducimos esta matriz a la forma escalonada:

$$\begin{pmatrix} 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & -3 & 1 & 1 \\ 0 & 0 & 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

De aquí concluimos que los vectores

e_1, e_2, e_3, e'_2

forman una base de $V+W$.

2. Sean V y W subespacios generados por las columnas de las matrices A y B, respectivamente, donde

$$A = \begin{pmatrix} 1 & 2 & 0 & 3 & 0 \\ 1 & 2 & -1 & -1 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 2 & 4 & 1 & 10 & 1 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$B = \begin{pmatrix} 1 & 1 & 0 & -1 & 0 \\ 0 & -2 & 0 & 1 & 1 \\ 1 & -1 & 0 & 0 & 1 \\ 0 & 1 & 1 & 2 & 0 \\ -3 & 8 & 1 & 0 & -5 \end{pmatrix}$$

Resolvamos el mismo problema que en el ejercicio anterior.

Para V:

Llevamos la matriz A a la forma escalonada y obtenemos:

$$\begin{pmatrix} 1 & 2 & 0 & 3 & 0 \\ 0 & 0 & 1 & 4 & 0 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

La dimensión de V es igual al rango de la matriz A, por tanto, es 3. Buscando un menor de orden 3 que tenga determinante no nulo, tomamos el que forman las columnas 1, 3, 5 de la matriz A y, por tanto, los vectores:

$$e_1=(1,1,0,2,0),\; e_2=(0,-1,1,1,0),\; e_3=(0,0,0,1,1),$$

que son los que están en esas columnas, forman una base de V.

Para W:

Llevamos la matriz B a la forma escalonada y obtenemos:

$$\begin{pmatrix} 1 & 1 & 0 & -1 & 0 \\ 0 & -1 & 1 & 2 & 0 \\ 0 & 0 & 2 & 5 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

La dimensión de W es igual al rango de B, que es 3. Buscando un menor de orden 3 no nulo, podemos tomar el que forman las columnas 1, 2 y 3 y, por tanto, los vectores que están en esas columnas en B forman una base de W. Esta base está formada por

$$e'_1=(1,0,1,0,-3),\; e'_2=(1,-2,-1,1,8),\; e'_3=(0,0,0,1,1).$$

Para $V\cap W$:

Vamos a encontrar el espacio $V\cap W$.
Por definición, $V\cap W$ es el subespacio de los vectores de $\mathbf{R}^5$ (en este caso) que están en V y en W. Si $v\in V\cap W$, entonces $v\in V$ y $v\in W$; por tanto, se expresa como combinación lineal de (e_1,e_2,e_3) y de (e'_1,e'_2,e'_3):

$$v=a_1e_1+a_2e_2+a_3e_3=\beta_1e'_1+\beta_2e'_2+\beta_3e'_3.$$

Luego hay que buscar vectores que sean a la vez combinación lineal de ambos sistemas, uno que genera a V y el otro que genera a W. Calculamos

$$a_1e_1+a_2e_2+a_3e_3=\beta_1e'_1+\beta_2e'_2+\beta_3e'_3$$
$$(a_1, a_1-a_2, a_2, 2a_1+a_2+a_3, a_3)=(\beta_1+\beta_2, -2\beta_2, \beta_1-\beta_2, \beta_2+\beta_3, -3\beta_1+8\beta_2+\beta_3).$$

Por tanto:

$$\begin{aligned} a_1 &= \beta_1+\beta_2 \\ a_1-a_2 &= -2\beta_2 \\ a_2 &= \beta_1-\beta_2 \\ 2a_1+a_2+a_3 &= \beta_2+\beta_3 \\ a_3 &= -3\beta_1+8\beta_2+\beta_3 \end{aligned}$$

Así obtenemos un sistema de ecuaciones lineales homogéneo con seis incógnitas:

$$\left\{\begin{aligned} a_1-\beta_1-\beta_2 &= 0 \\ a_1-a_2+2\beta_2 &= 0 \\ a_2-\beta_1+\beta_2 &= 0 \\ 2a_1+a_2+a_3-\beta_2-\beta_3 &= 0 \\ a_3+3\beta_1-8\beta_2-\beta_3 &= 0 \end{aligned}\right.$$

Aplicando Gauss obtenemos:

$$\left\{\begin{aligned} a_1-\beta_1-\beta_2 &= 0 \\ a_2-\beta_1-3\beta_2 &= 0 \\ a_3+3\beta_1+4\beta_2-\beta_3 &= 0 \\ \beta_2 &= 0 \end{aligned}\right.$$

Resolviendo:

$a_1=a_2=\beta_1,\ \beta_2=0,\ a_3=\beta_3-3\beta_1.$

De aquí llegamos a la conclusión de que $V\cap W$ está generado por (e'_1, e'_3), pues un vector que tenga la segunda componente nula, en la base (e'_1, e'_2, e'_3) de W, tiene una expresión en términos de (e_1, e_2, e_3) y, por tanto, también está en V.

En este caso $V\cap W$ tiene dimensión 2 y la base que seleccionamos es $(\bar{e}_1=(1,0,1,0,-3),\ \bar{e}_2=(0,0,0,1,1))$.

Hay que tener cuidado con la interpretación del sistema de ecuaciones que obtenemos al analizar $V\cap W$. Aquí hay que interpretar las condiciones que se imponen a unos coeficientes en términos de los otros y las restricciones que estas implican en los subespacios V y W, que son las que dan la condición de pertenencia a ambos subespacios.

Para $V+W$:

Aquí el método es el mismo que utilizamos en el ejemplo 1: partimos de la base de $V\cap W$ y completamos una base de V y otra de W.

En este ejemplo completar la base de W es evidente, pues si tomamos

$\bar{e}_3=(1,-2,-1,1,8)$,

el sistema $(\bar{e}_1, \bar{e}_2, \bar{e}_3)$ es una base en W.

Tratemos de completar la base de V.

Tenemos que encontrar un vector de V que sea linealmente independiente con $(\bar{e}_1$ y $\bar{e}_2)$. Para ello hay varios caminos; vamos a emplear el de cambio de base. Como $\bar{e}_1$ y $\bar{e}_2$ son vectores de $V \cap W$, están en V y, por tanto, se pueden expresar como combinación lineal de e_1, e_2, e_3. En este caso tenemos:

$\bar{e}_1 = \alpha e_1 + \beta e_2 + \gamma e_3$

$$\begin{pmatrix} 1 \\ 0 \\ 1 \\ 0 \\ -3 \end{pmatrix} = \alpha \begin{pmatrix} 1 \\ 1 \\ 0 \\ 2 \\ 0 \end{pmatrix} + \beta \begin{pmatrix} 0 \\ -1 \\ 1 \\ 1 \\ 0 \end{pmatrix} \gamma \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \\ 1 \end{pmatrix}$$

$$\begin{aligned} \alpha &= 1 \\ \alpha - \beta &= 0 \\ \beta &= 1 \\ 2\alpha + \beta + \gamma &= 0 \\ \gamma &= -3 \end{aligned}$$

Por tanto:

$\bar{e}_1 = e_1 + e_2 - 3e_3$

Para $\bar{e}_2$ obtenemos:

$\bar{e}_2 = 0e_1 + 0e_2 + e_3$

Los vectores de coordenadas de $\bar{e}_1$ y $\bar{e}_2$ en la base (e_1, e_2, e_3) son, respectivamente,

$$\begin{pmatrix} 1 \\ 1 \\ -3 \end{pmatrix} \text{ y } \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

Para buscar el tercer vector de una base, solo falta tomar tres coordenadas (a, b, c), de forma tal que la matriz

$$\begin{pmatrix} 1 & 0 & a \\ 1 & 0 & b \\ -3 & 1 & c \end{pmatrix}$$

sea una matriz inversible y, por tanto, una matriz de cambio de coordenadas en el espacio V, según la proposición 1.11. En este caso es fácil seleccionar los números. Escojamos $a=1$, $b=0$, $c=0$. Así, el tercer vector de la base será:

$\bar{e}_4 = e_1 + 0e_2 + 0e_3 = e_1 = (1,1,0,2,0)$.

La base que completamos en V es $(\bar{e}_1, \bar{e}_2, \bar{e}_4)$, y la base que buscamos en $V+W$ está formada por

$\bar{e}_1 = (1,0,1,0,-3)$, $\qquad \bar{e}_2 = (0,0,0,1,1)$

$\bar{e}_3 = (1,-2,-1,1,8)$ $\qquad \bar{e}_4 = (1,1,0,2,0)$.

En estos dos ejemplos observamos que aunque las bases obtenidas para $V+W$ son diferentes, los espacios V, W, $V \cap W$ y $V+W$ son los mismos. Las diferencias que obtenemos no solo se deben a que hemos empleado métodos diferentes, sino también, fundamentalmente, a que empleamos el teorema de completamiento de base, y para completar una base no hay un método único.

Si evaluáramos los dos métodos empleados, diríamos que en algunos casos es más conveniente trabajar con subespacios generados por un sistema de vectores, y en otros, con subespacios solución de sistemas de ecuaciones. El caso más evidente es $V \cap W$. Es mucho mas fácil, en general, trabajar cuando V y W están dados como solución de sistemas de ecuaciones. Verdaderamente, no se trata de que sea más fácil o más difícil; lo que sí está mucho más claro en este caso es cómo obtener la conclusión. Por lo general, en la otra forma, el estudiante tiende a cometer algún error. Alertamos, entonces, sobre esto. Trate de trabajar para $V \cap W$ con sistemas de ecuaciones, o tenga mucho cuidado al establecer las condiciones para $V \cap W$ si utiliza el otro método.

El estudiante puede observar también que este tipo de ejercicios, aunque es de los que llamamos numéricos o "prácticos" (en contraposición a los que se llaman "teóricos"), si bien tienen métodos de trabajo que pueden convertirse en un poco "mecánicos", llevan intrínsecos el empleo de toda la teoría desarrollada en este capítulo, y constituyen un ejemplo de la importancia de conocer perfectamente esta teoría.

Ejercicios

1. En $\mathbb{R}^2$ se definen las operaciones:

$$(x_1,y_1) \oplus (x_2,y_2) = (x_1+x_2,\ y_1+y_2),$$
$$\alpha \odot (x_1,y_1) = (\alpha x_1, y_1),$$

para (x_1,y_1), $(x_2,y_2) \in \mathbb{R}^2$, $\alpha \in \mathbb{R}$.
Analice si $\mathbb{R}^2$ con estas dos operaciones es un espacio vectorial sobre $\mathbb{R}$.

2. En $\mathbb{R}^n$ se definen las operaciones:

$$a \oplus b = a - b,$$
$$\alpha \odot a = -\alpha a,$$

para $a, b \in \mathbb{R}^n$, $\alpha \in \mathbb{R}$.
¿Cuáles de los axiomas de espacio vectorial satisface $\mathbb{R}^n$ con estas dos operaciones?

3. En $K^2 (K = \mathbb{R}$ o $\mathbb{C})$ se definen:

$$(a,b) \oplus (c,d) = (a+d, b+c),$$
$$\alpha \odot (a,b) = (\alpha a, \alpha b),$$

para (a,b), $(c,d) \in K^2$, $\alpha \in K$.

¿Constituye K^2 con estas operaciones un espacio vectorial sobre K?

4. Sean $E=\mathbb{R}_+^*$, $K=\mathbb{R}$. Se definen las operaciones:

$$a \oplus b = ab,$$
$$\alpha \odot a = a^{\alpha},$$

para $a, b \in E$, $\alpha \in \mathbb{R}$.

¿Constituye E un espacio vectorial sobre $\mathbb{R}$?

5. En el conjunto de las funciones complejas de variable real, F ($\mathbb{R}, \mathbb{C}$) considere el subconjunto:

$$V=\{ f \in \mathrm{F}(\mathbb{R}, \mathbb{C}) / f(-t) = \overline{f(t)} \}.$$

Demuestre que V con las operaciones de suma de funciones y producto de una función por un escalar, es un espacio vectorial sobre $\mathbb{R}$. ¿Puede ser V un espacio vectorial sobre $\mathbb{C}$ con las mismas operaciones?

6. En K^2 se definen las operaciones:

$$(x_1,y_1) \oplus (x_2,y_2) = (x_1+x_2,0),$$
$$\alpha(x_1,y_1) = (\alpha x_1,0),$$

para (x_1,y_1), $(x_2,y_2) \in K^2$, $\alpha \in K$. ¿Es K^2 con estas operaciones un espacio vectorial sobre K?

7. En $\mathbb{R}^3$, sea V el subconjunto de los vectores tales que el extremo del vector se encuentra en un plano fijo π. ¿Es V un subespacio vectorial de $\mathbb{R}^3$?

8. Sean $K=\mathbb{R}, \mathbb{C}$ y X un conjunto cualquiera. Se define:
$K^{(X)} = \{f : X \longrightarrow K / f\,(a) \neq 0$ para un número finito de elementos de $X\}$.
   a) Demuestre que $K^{(X)}$ es un espacio vectorial sobre K, para la suma usual de funciones y el producto de funciones por un escalar.
   b) Si $X=\mathbb{N}$, demuestre que $K^{(\mathbb{N})}$ es isomorfo al espacio de los polinomios en una indeterminada con coeficientes en K.

9. Diga cuáles de los siguientes subconjuntos de $\mathbb{R}^n$ son subespacios vectoriales de $\mathbb{R}^n$; $(a_1,a_2,\ldots,a_n) \in \mathbb{R}^n$, $n \geqslant 3$.
   a) $a_1 \geqslant 0$ b) $a_1+3a_2=a_3$ c) $a_1^2=a_2$
   d) $a_1 \cdot a_2=0$ e) a_3 es racional.

10. En el espacio vectorial $F(\mathbb{R}, \mathbb{R})$ de las funciones reales de variable real, diga cuáles de los siguientes subconjuntos son subespacios vectoriales de $F(\mathbb{R}, \mathbb{R})$.
El subconjunto de las funciones f que cumplen
   a) $f(x) = -f(-x)$ e) $f(3) = 1 + f(-5)$
   b) $f(x) = f(-x)$ f) $f(-2) = 0$
   c) $f(x^2) = f(x)^2$ g) f es integrable
   d) $f(2) = f(1)$ h) f es continua.

11. En el espacio de las matrices reales de orden 3, $M_3(\mathbb{R})$, diga cuáles de los siguientes subconjuntos son subespacios vectoriales de $M_3(\mathbb{R})$.

a) El de las matrices simétricas.
b) El de las matrices antisimétricas.
c) El de las matrices que tienen la diagonal principal nula.
d) El de las matrices triangulares superiores.
e) El de las matrices triangulares inferiores.
f) El de las matrices A que cumplen $A=A^2$.
g) El de las matrices nilpotentes, es decir, tales que existe un entero k tal que $A^k=0$.
h) El de las matrices de determinante 0.
i) El de las matrices de traza 0.
j) El de las matrices que conmutan con A, dada una matriz fija A.
k) El de las matrices inversibles.

12. Exprese, si es posible, el vector v como combinación lineal de los vectores del sistema (e_i).

a) En $\mathbb{R}^3$: $v=(1,-2,5)$; $e_1=(1,1,1)$, $e_2=(1,2,3)$, $e_3=(2,-1,1)$.
b) En $\mathbb{R}^3$: $v=(2,-5,3)$; $e_1=(1,-3,2)$, $e_2=(2,-4,-1)$, $e_3=(1,-5,7)$.
c) En $\mathbb{R}^4$: $v=(3,-1,0,-1)$; $e_1=(2,-1,3,2)$, $e_2=(-1,1,1,-3)$, $e_3=(1,1,9,-5)$.
d) En $\mathbb{R}^3$: $v=(-1,1,2)$; $e_1=(3,0,-3)$, $e_2=(4,2,-2)$, $e_3=(2,1,1)$.

13. ¿Para qué valores del parámetro k se puede escribir el vector $v=(1,k,-2)$ de $\mathbb{R}^3$ como combinación lineal de los vectores $a_1=(3,-2,0)$, $a_2=(2,-5,-1)$? ¿Qué interpretación geométrica se puede dar en este caso?

14. En el espacio $\mathbb{R}_3[x]$, escriba el polinomio $p(x)=-3+4x+x^2$ como combinación lineal de los polinomios: $q_1(x)=5-2x+x^2$, $q_2(x)=-3x+x^2$, $q_3(x)=3+x$.

15. En el espacio $M_2(\mathbb{R})$ exprese la matriz

$$A=\begin{pmatrix}1 & -1\\ 3 & 1\end{pmatrix}$$

como combinación lineal de las matrices:

$$E_1=\begin{pmatrix}1 & 0\\ 1 & 1\end{pmatrix},\ E_2=\begin{pmatrix}1 & 1\\ 0 & 0\end{pmatrix},\ E_3=\begin{pmatrix}0 & -1\\ 0 & 2\end{pmatrix}.$$

16. Determine si la siguiente aplicación constituye un isomorfismo del espacio vectorial K^2 (con las operaciones usuales) en sí mismo:

$$f\colon K^2 \longrightarrow K^2,\ f(x,y)=(x+y,\ x-y).$$

17. Sean E y F dos espacios vectoriales sobre K, φ un isomorfismo de E sobre F. Demuestre que la aplicación $\hat{\varphi}$ de $E\times F$ en $E\times F$ definida por

$$\hat{\varphi}(x,y)=(x,y-\varphi(x))$$

es un isomorfismo.

18. Diga cuáles de los siguientes sistemas de vectores son linealmente independientes:

a) En $\mathbb{R}^3$, el sistema (v, e_1, e_2, e_3) donde v, e_1, e_2, e_3 son los vectores del inciso a) del ejercicio 12.
b) En $\mathbb{R}^3$, el sistema (v, e_1, e_2, e_3) donde v, e_1, e_2, e_3 son los vectores del inciso b) del ejercicio 12.
c) En $\mathbb{R}^4$, el sistema $((1,-1,-4,0)), (1,1,2,4), (2,-1,-5,2), (2,1,1,6))$.
d) En $\mathbb{R}^4$, el sistema $((1,2,-1,-2), (2,3,0,-1), (1,2,1,4), (1,3,-1,0))$.

19. Encuentre un sistema de ecuaciones lineales homogéneo que represente cada uno de los subespacios siguientes:

a) En $\mathbb{R}^4$, el subespacio generado por:
α) $\{(2,1,-3,0), (3,0,2,-5)\ (1,-1,1,1)\}$.
β) $\{(1,-1,1,0), (1,1,0,1), (2,0,1,1)\}$.
b) En $\mathbb{R}^3$, el subespacio generado por:
α) $\{(1,2,0), (0,1,0)\}$.
β) $\{(2,-1,0), (1,3,0)\}$.

20. Sean E un espacio vectorial, V y W dos subespacios de E. Demuestre que $V \cup W$ es un subespacio de E si y solo si $V \subset W$ o $W \subset V$.

21. En $\mathbb{R}^3$, considere los subespacios:

$V = \{(x, y, z)/x = y = z\}$, $W = \{(0, y, z)\}$.

Demuestre que $\mathbb{R}^3 = V \oplus W$.

22. Sea E un espacio vectorial sobre K y sean U, V y W subespacios de E. Demuestre que

$(U \cap V) + (U \cap W) \subset U \cap (V + W)$.

Halle un contraejemplo para demostrar que los dos subespacios no son necesariamente iguales.

23. En $\mathbb{R}^3$, considere los subespacios:

$U = \{(x, y, z)/x + y + z = 0\}$,
$V = \{(x, y, z)/x = z\}$,
$W = \{(0, 0, z)/z \in \mathbb{R}\}$.

a) Calcule: $U+V$, $U+W$, $V+W$, $(U \cap V) + (U \cap W)$, $U \cap (V+W)$.
b) Diga en qué casos las sumas calculadas son directas.
c) Diga qué subespacios son suplementarios con respecto a $\mathbb{R}^3$.

24. Sean E y F dos espacios vectoriales sobre K. En el espacio $E \times F$ considere los subespacios:

$E' = \{(x, 0)/\ x \in E\}$,
$F' = \{(0, y)/y \in F\}$.

a) Demuestre que E' es isomorfo a E y F' es isomorfo a F.
b) Demuestre que $E \times F = E' \oplus F'$.

25. En el espacio de las matrices $M_2(\mathbb{R})$ se definen:

$$V=\left\{\begin{pmatrix} a & -a \\ b & c \end{pmatrix} / a, b, c \in \mathbb{R}\right\},$$

$$W=\left\{\begin{pmatrix} a & b \\ -a & c \end{pmatrix} / a, b, c \in \mathbb{R}\right\}.$$

a) Demuestre que V y W son subespacios de $M_2(\mathbb{R})$.
b) Halle $V \cap W$ y $V+W$.
c) Halle una base y la dimensión de: V, W, $V \cap W$ y $V+W$.

26. En el espacio $M_2(\mathbb{C})$,

$$W=\{A=(a_{ij}) \ / \ \text{Traza}\ (A)=a_{11}+a_{22}=0\}.$$

W es un subespacio de $M_2(\mathbb{C})$. Halle una base de W y calcule su dimensión.

27. En el espacio $M_n(\mathbb{C})$,

$$W=\{A/\ \text{Traza}\ (A)=0\}.$$

Calcule la dimensión de W.

28. Halle una base y la dimensión de $M_{m,n}(K)$.

29. Demuestre que $\mathbb{R}$ como $\mathbb{Q}$-espacio no es un espacio de dimensión finita.

30. Demuestre que en el espacio de las funciones continuas reales, los vectores e^t, sen t, cos t son linealmente independientes.

31. En el espacio de las funciones continuas reales, considere la familia infinita

$$F=\{\text{sen}\ kt,\ \cos kt\}_{k \in \mathbb{N}}.$$

a) Demuestre que cualquier parte finita de esta familia es linealmente independiente.
b) Lo anterior significa que la familia F es linealmente independiente. ¿Qué significaría que F genera a $C(\mathbb{R}, \mathbb{R})$? ¿Es esto posible?

32. En el espacio $\mathbb{C}^2$ considere los vectores:

$$v=(1+i,\ 2i) \text{ y } w=(1,\ 1+i).$$

a) Demuestre que v y w son linealmente independientes si se considera $\mathbb{C}^2$ como $\mathbb{R}$-espacio.
b) Demuestre que v y w son linealmente dependientes si se considera $\mathbb{C}^2$ como $\mathbb{C}$-espacio.

33. Sea S el siguiente sistema, constituido por vectores de $\mathbb{R}^4$:

$$S=((1,-4,-2,1),\ (1,-3,-1,2),\ (3,-8,-2,7),\ (2,-7,-3,3)).$$

a) Halle la dimensión de $\langle S \rangle$.
b) Extraiga de S una base de $\langle S \rangle$.

c) Complete, si es necesario, la base de $\langle S\rangle$ hallada en b), para obtener una base de $\mathbb{R}^4$.

34. Determine si los siguientes sistemas constituyen o no bases de $\mathbb{C}^3$:
 a) $((1,5,-6),\ (2i,1,8-i),\ (3,-1-2i,4))$.
 b) $((-1,3,-4),(1,4,-3),\ (2,3,-11))$.

35. a) Demuestre que si en un espacio vectorial E, los vectores u, v, w son linealmente independientes, los vectores $u+v$, $u-v$ y $u-2v+w$ son también linealmente independientes.
 b) ¿Cuál es el número máximo de vectores linealmente independientes que puede obtenerse a partir de combinaciones lineales de los vectores u, v y w?

36. Sean E un espacio vectorial sobre K, $(a_1,a_2,\dots,a_n)$ un sistema de vectores linealmente independientes. Sean, además, $v_1,v_2,\dots,v_k$ vectores que se obtienen como combinaciones lineales:

$$v_1=\alpha_{11}a_1+\alpha_{21}a_2+\dots+\alpha_{n1}a_n$$
$$v_2=\alpha_{12}a_1+\alpha_{22}a_2+\dots+\alpha_{n2}a_n$$
$$\dots\dots\dots\dots\dots\dots\dots\dots\dots\dots$$
$$v_k=\alpha_{1k}a_1+\alpha_{2k}a_2+\dots+\alpha_{nk}a_n$$

y A la matriz o $n\times k$:

$$\begin{pmatrix} \alpha_{11} & \alpha_{12} & \cdots & \alpha_{1k} \\ \alpha_{21} & \alpha_{22} & \cdots & \alpha_{2k} \\ \dots & \dots & \dots & \dots \\ \alpha_{n1} & \alpha_{n2} & \cdots & \alpha_{nk} \end{pmatrix}$$

 a) Demuestre que el sistema de vectores $(v_1,v_2,\dots,v_k)$ es linealmente independiente si y solo si el rango de la matriz A es igual a k y $k\leqslant n$.
 b) A partir del inciso a) demuestre que un sistema de n vectores $(v_1,v_2,\dots,v_n)$ es linealmente independiente si y solo si $\det A\neq 0$.

37. En $\mathbb{R}^4$ considere los subespacios:

$$V=\{(x,y,z,t)/x+y=0,\ z=2t\},$$
$$W=\{(x,y,z,t)/y+z+t=0\}.$$

 Halle una base y la dimensión de: V, W, $V\cap W$, $V+W$, $V\times W$.

38. En $\mathbb{R}^4$ considere los subespacios:
 a) $V=\langle(1,1,0,-1),\ (1,2,3,0),\ (2,3,3,-1)\rangle$
  $W=\langle(1,2,2,-2),(2,3,2,-3),\ (1,3,4,-3)\rangle$
 b) $V=\langle(1,1,2,4),\ (2,1,5,2),\ (1,-1,-4,0)\rangle$
  $W=\langle(1,0,-1,1),\ (2,0,1,1),\ (0,-1,1,0)\rangle$

 En cada caso halle una base y la dimensión de: V, W, $V\cap W$ y $V+W$.

39. Considere los subespacios de $\mathbb{R}^3$:

$$V=\langle(1,2,3),\ (2,3,8),\ (3,2,17)\rangle,$$
$$W=\{(x,y,z)/2z-y=0\}.$$

a) Halle una base y la dimensión de: $V, W, V \cap W$ y $V+W$.
b) Halle un suplementario de $V \cap W$ en W.

40. Sean E un espacio vectorial de dimensión 4 y (e_1, e_2, e_3, e_4) una base de E.

a) Determine qué valores debe poseer el parámetro k para que los vectores:

$v_1 = ku_2 + u_3 - ku_4$,
$v_2 = 2u_1 + u_2 + 3u_4$,
$v_3 = 2ku_2 + ku_3$,
$v_4 = u_1 - u_2$,

formen una base de E.

b) Para los valores de k tales que (v_1, v_2, v_3, v_4) no forman una base de E, halle una base y la dimensión del subespacio generado por ellos.

41. En el espacio vectorial del ejercicio 40, y para los casos en que (v_1, v_2, v_3, v_4) sea una base, escriba la matriz de cambio de coordenadas de la base (e_1, e_2, e_3, e_4) a la base (v_1, v_2, v_3, v_4).

42. En el espacio de las funciones $C(\mathbb{R}, \mathbb{C})$ como espacio vectorial sobre $\mathbb{C}$, considere el subespacio V generado por las funciones:

$f_1(x) = 1$, $f_2(x) = \text{sen } x$, $f_3(x) = \cos x$.

a) Demuestre que V tiene dimensión 3.
b) Demuestre que el sistema de vectores $(e^{ix}, e^{-ix}, 1)$ es una base de V y halle la matriz de cambio de coordenadas de (f_1, f_2, f_3) a $(e^{ix}, e^{-ix}, 1)$.

43. En el espacio de los polinomios de grado menor que 4, $K_4[x]$, $(1, x, x^2, x^3)$ es una base.
Sea $a \in K$. Demuestre que $(1, (x-a), (x-a)^2, (x-a)^3)$ es también una base de $K_4[x]$ y halle las coordenadas de un polinomio cualquiera $e_0 + e_1 x + e_2 x^2 + e_3 x^3$ en esta base.

44. En el espacio de las matrices simétricas de orden 2, $MS_2(\mathbb{R})$, demuestre que las matrices

$$\begin{pmatrix} 1 & -2 \\ -2 & 1 \end{pmatrix}, \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}, \begin{pmatrix} 4 & -1 \\ -1 & -5 \end{pmatrix}$$

forman una base y halle las coordenadas de la matriz

$$\begin{pmatrix} 4 & -11 \\ -11 & -7 \end{pmatrix}$$

en esta base.

45. En el espacio $MS_2(\mathbb{R})$ demuestre que las matrices

$$\begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$

forman una base y halle la matriz de cambio de coordenadas de la base del ejercicio 44 a esta nueva base.

46. Sea E un espacio vectorial real de dimensión 3. Sea $x \in E$, cuyo vector de coordenadas con respecto a una base (a_1, a_2, a_3) de E es

$$X = \begin{pmatrix} 2 \\ 0 \\ -3 \end{pmatrix}$$

Demuestre que existe una base (b_1, b_2, b_3) de E en la cual el vector de coordenadas de x es

$$\begin{pmatrix} 4 \\ -2 \\ 0 \end{pmatrix}$$

47. En $\mathbf{R}^5$ considere el subespacio:

$V = \langle (3,21,0,9,0),\ (1,7,-1,-2,-1),\ (2,14,0,6,1),\ (6,\ 42,-1,13,0) \rangle.$

a) Halle una base de V.
b) Halle un sistema de ecuaciones lineales, homogéneo tal que V sea el espacio solución del sistema.
c) A partir del sistema hallado, encuentre otra base de V.

48. En $\mathbb{C}^3$ considere el subespacio complejo:

$V = \langle (1,0,i),\ (1+i, 1-i, 1),\ (i,i,i) \rangle.$

a) Halle una base de V.
b) Halle un sistema de ecuaciones lineales homogéneos tal que V sea el espacio solución del sistema.
c) A partir del sistema hallado, encuentre otra base de V.

49. En el espacio vectorial dado en cada caso, investigue si el sistema de vectores (a_i) es una base del espacio y halle las coordenadas del vector v en la base (a_i) y la matriz de cambio de coordenadas de la base canónica a la base (a_i).

a) En $\mathbf{R}^3$: $v=(6,2,-7)$; $a_1=(2,1,-3)$, $a_2=(3,2,-5)$, $a_3=(1,-1,1)$.
b) En $\mathbf{R}^3$: $v=(6,9,14)$; $a_1=(1,1,1)$, $a_2=(1,1,2)$, $a_3=(1,2,3)$.
c) En $\mathbf{R}^3$: $v=(6,2,-7)$; $a_1=(2,1,-3)$, $a_2=(3,2,-3)$, $a_3=(1,-1,1)$.
d) En $\mathbf{R}^3$: $v=(6,\ 2,\ -7)$; $a_1=(0,0,1)$, $a_2=(1,0,0)$, $a_3=(0,1,0)$.
e) En $\mathbf{R}^4$: $v=(7,14,-1,2)$; $a_1=(1,2,-1,-2)$, $a_2=(2,3,0,-1)$, $a_3=(1,2,1,4)$, $a_4=(1,3,-1,0)$.

50. En un espacio vectorial E, sean L un sistema de vectores linealmente independiente, $\langle L \rangle$ el subespacio generado por L, a un vector de E tal que $a \notin \langle L \rangle$. Demuestre que $L \cup \{a\}$ es linealmente independiente.

51. Sea E un espacio vectorial de dimensión n. Demuestre las proposiciones siguientes:

a) Todo subconjunto de vectores de E que contiene más de n vectores, es linealmente dependiente.

b) Ningún subconjunto de E que contiene menos de n vectores, puede generar a E.

c) Si F es un subespacio de E, entonces F es de dimensión finita y $\dim F \leqslant \dim E$.

52. En $K[x]$ sea $p(x)$ un polinomio de grado n. Se define:

$[p(x)] = \{p(x) \cdot q(x) / q(x) \in K[x]\}$.

Demuestre que $[p(x)]$ es un subespacio de $K[x]$ y

$K[x] = [p(x)] \oplus K_n[x]$.

53. Sean V, W dos subespacios vectoriales de un espacio E sobre K. Demuestre que si S_1 es un sistema generador de V y S_2 es un sistema generador de W, entonces $S_1 \cup S_2$ es un sistema generador de $V+W$.

2 Aplicaciones lineales

Introducción

Muchas veces en la práctica es necesario analizar las relaciones entre dos espacios vectoriales E y F. Sabemos que, en general, las relaciones entre dos conjuntos se representan desde el punto de vista matemático mediante el concepto de aplicación. En el caso de dos espacios vectoriales, son de interés aquellas aplicaciones que establecen relaciones entre las estructuras de ambos espacios; estas son las llamadas *aplicaciones lineales*, que aparecen frecuentemente en la práctica y mediante las cuales muchas veces es posible "aproximar" aplicaciones arbitrarias.

2.1 Aplicaciones lineales

Al estudiar en el capítulo 1 los isomorfismos de espacios vectoriales, analizamos que si queremos que una aplicación f de un espacio E sobre K en un espacio F sobre K, "traslade" la estructura vectorial de E a F, las condiciones que debemos exigir a f son:

(1) $f(x+y)=f(x)+f(y) \quad \forall x,y \in E$

(2) $f(ax)=af(x) \quad \forall a \in K, \ \forall x \in E.$

En el caso de los isomorfismos analizamos no solo la transferencia de estructura, sino que F fuese una "copia" de E; es por esto que exigimos, además, que f fuese biyectiva.

Ahora bien, algunas veces surgen correspondencias entre espacios que no son idénticos estructuralmente, pero dichas correspondencias "trasladan", en cierta medida, la estructura de uno de los espacios al otro. Veamos dos ejemplos.

Ejemplos

1. Sean los espacios $\mathbb{R}^3$ y $\mathbb{R}^2$ y consideremos la aplicación $f: \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ dada por $(x, y, z) \rightarrow (x,y)$.

   f cumple las propiedades (1) y (2):

$$\begin{aligned} f[(x_1,y_1,z_1)+(x_2,y_2,z_2)] &= f(x_1+x_2, y_1+y_2, z_1+z_2)=(x_1+x_2, y_1+y_2) = \\ &=(x_1, y_1)+(x_2, y_2) \\ &= f[(x_1,y_1,z_1)]+f[(x_2,y_2,z_2)]. \end{aligned}$$

$$\begin{aligned} f[a(x_1,y_1,z_1)] &= f(ax_1, ay_1, az_1) = (ax_1, ay_1) = \\ &= a(x_1,y_1) = af[(x_1,y_1,z_1)] \end{aligned}$$

Pero f no es inyectiva, aunque sí sobreyectiva. Pudiéramos decir en este caso que una parte de $\mathbb{R}^3$ se transforma en $\mathbb{R}^2$.

2. Sean ahora los $\mathbb{R}$-espacios $\mathbb{R}$ y $M_2(\mathbb{R})$ y consideremos la aplicación $g: \mathbb{R} \to M_2(\mathbb{R})$ dada por

$$k \to \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix}.$$

g cumple también (1) y (2) pues

$$g(k+k') = \begin{pmatrix} k+k' & 0 \\ 0 & k+k' \end{pmatrix} = \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} + \begin{pmatrix} k' & 0 \\ 0 & k' \end{pmatrix} = g(k) + g(k').$$

$$g(ak) = \begin{pmatrix} ak & 0 \\ 0 & ak \end{pmatrix} = a \begin{pmatrix} k & 0 \\ 0 & k \end{pmatrix} = ag(k).$$

En este caso, g es inyectiva pero no es sobreyectiva. Intuitivamente decimos que g traslada la estructura de $\mathbb{R}$ a una parte de $M_2(\mathbb{R})$, la formada por las matrices escalares.

Existen también ejemplos de aplicaciones entre espacios vectoriales que satisfacen (1) y (2) y no son inyectivas ni sobreyectivas, pero que, sin embargo, establecen una relación entre las estructuras de dichos espacios. Consideremos la siguiente definición.

DEFINICIÓN 2.1

Se denomina *aplicación lineal* entre dos K-espacios vectoriales E y E', a una aplicación f de E en E' tal que:

$$f(x+y) = f(x) + f(y) \quad \forall x, y \in E,$$
$$f(ax) = af(x) \quad \forall a \in k, \quad \forall x \in E.$$

Ejemplos

1. En todo espacio vectorial E sobre K, dado $a \in K$, la aplicación $f_a: E \longrightarrow E$ dada por $x \to a.x$ es una aplicación lineal de E en E.
   Las aplicaciones f_a son biyectivas para $a \neq 0$; f_a se denomina *homotecia de razón a.*

2. En un espacio E, son ejemplos triviales, pero no por ello menos importantes, los siguientes:

$$f_0: E \longrightarrow E \text{ dada por } x \to 0 \text{ para todo } x \in E,$$
$$id_E: E \longrightarrow E \text{ dada por } x \to x \text{ para todo } x \in E.$$

Estas dos aplicaciones se llaman, respectivamente, la *aplicación nula* y la aplicación identidad de E, y pueden considerarse casos particulares de las homotecias para los valores $a=0$ y $a=1$.
Cuando no exista confusión posible, designaremos f_0 por 0. Así, afirmar que una aplicación lineal f es tal que $f=0$ equivale a $f(x)=0 \ \forall x \in E$.

3. Dada una matriz A fija de orden $m \times n$, la aplicación f de K^n en K^m dada por $X \rightsquigarrow AX$ es una aplicación lineal.
En efecto,

$$f(X+Y) = A(X+Y) = AX + AY = f(X) + f(Y),$$
$$f(aX) = A(aX) = a(AX) = af(X).$$

4. La aplicación $T: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ dada por $T(x,y) = (x+1, y+2)$ no es lineal. Por ejemplo, $f[(1,1)+(1,0)] = f(2,1) = (3,3)$, mientras que $f(1,1) + f(1,0) = (2,3) + (2,2) = (4,5)$.
5. En $\mathbb{C}$ como $\mathbb{R}$-espacio, la aplicación $g: \mathbb{C} \longrightarrow \mathbb{C}$ dada por $a+ib \rightsquigarrow a-ib$ es lineal.
Sin embargo, si consideramos $\mathbb{C}$ como $\mathbb{C}$-espacio, la aplicación g no es lineal ya que $f(i(a+ib)) = f(i a-b) = -b-ia$,
pero

$$if(a+ib) = i(a-ib) = b+ia.$$

6. Para $m<n$, la aplicación $p: K^n \longrightarrow K^m$ dada por

$$(x_1, x_2, \ldots, x_m, x_{m+1}, \ldots, x_n) \rightsquigarrow (x_1, x_2, \ldots, x_m)$$

es una aplicación lineal.
7. En $K[x]$ como espacio vectorial sobre K, la aplicación $D: K[x] \longrightarrow K[x]$ dada por $p \rightsquigarrow p'$ (p': polinomio derivada de p) es lineal, ya que

$$D(p+q) = (p+q)' = p' + q' = D(p) + D(q),$$
$$D(\lambda p) = (\lambda p)' = \lambda p' = \lambda D(p).$$

8. En el $\mathbb{R}$-espacio $\mathbb{R}^{\mathbb{N}}$ de las sucesiones reales, son lineales las aplicaciones R e I dadas por

$$R[(a_0, a_1, \ldots, a_n, \ldots)] = (0, a_0, a_1, \ldots),$$
$$I[(a_0, a_1, \ldots, a_n, \ldots)] = (a_1, a_2, \ldots).$$

9. En el espacio $C[0,1]$ de las funciones continuas definidas en $[0,1]$ son lineales:

a) $I: C[0,1] \longrightarrow C[0,1]$ dada por $I(f) = \displaystyle\int_0^x f(t)\,dt$

b) $I_d: C[0,1] \longrightarrow \mathbb{R}$ dada por $I_d(f) = \displaystyle\int_0^1 f(t)\,dt$

Compruebe el estudiante que I es efectivamente una aplicación de $C[0,1]$ en $C[0,1]$ y la linealidad de I y de I_d mediante las propiedades de la integral.

Un *isomorfismo* entre dos K-espacios E y E' no es más que una aplicación lineal biyectiva entre dichos espacios. En el caso particular en que $E=E'$, una aplicación lineal de E en E se denomina *endomorfismo*. Son endomorfismos las aplicaciones lineales de los ejemplos 1, 2, 5, 7, 8 y 9a.

A un endomorfismo biyectivo se le llama *automorfismo*. Obsérvese que la definición de aplicación lineal no está restringida a espacios de dimensión finita, y en los ejemplos hemos visto aplicaciones lineales entre espacios de dimensión infinita.

Si f es una aplicación lineal de E en F, entonces $f(0)=0$ ya que $f(0)=f(0+0)=f(0)+f(0)$.

Esta propiedad se emplea generalmente para verificar la linealidad en los casos en que intuitivamente suponemos que una aplicación no es lineal. En el ejemplo 4 hubiera bastado comprobar que $F(0,0)=(1,2)$ para poder afirmar que F no es lineal. Pero, cuidado, por supuesto que $f(0)=0$ no basta para afirmar que f es lineal, como lo demuestra el ejemplo 5, ya que $g(0)=0$ y, sin embargo, vimos que g no era una aplicación lineal de $\mathbb{C}$ en $\mathbb{C}$ como $\mathbb{C}$-espacio.

Es evidente que de la definición de linealidad resulta:

Si f es una aplicación lineal de E en F, entonces

$$f(\alpha x+\beta y)=\alpha f(x)+\beta f(y) \quad \forall\, \alpha, \beta \in K, \quad \forall\, x, y \in E.$$

Esta condición es suficiente para afirmar la linealidad de una aplicación, pues para $\alpha=\beta=1$ obtenemos:

$$f(x+y)=f(x)+f(y) \quad \forall\, x, y \in E,$$

y para $\beta=0$:

$$f(\alpha x)=\alpha f(x) \quad \forall\, \alpha \in K \quad \forall\, x \in E.$$

Entonces podemos decir que una aplicación f de E en F es lineal si y solo si

$$f(\alpha x+\beta y)=\alpha f(x)+\beta f(y) \quad \forall\, \alpha, \beta \in K, \ \forall\, x, y \in E.$$

Esta definición, equivalente a la 2.1, resulta útil en la práctica para comprobar la linealidad de una aplicación. Así, en el ejemplo 9a:

$$I(\alpha f+\beta g)=\int_0^x (\alpha f+\beta g)(t)\,dt=\alpha\int_0^x f(t)\,dt+\beta\int_0^x g(t)\,dt$$

$$=\alpha I(f)+\beta I(g).$$

Esta caracterización conduce a una propiedad que es lógico suponer que cumplan las aplicaciones lineales, y es que estas "conservan" las combinaciones lineales. Así, si f de E en E' es una aplicación lineal y $\sum_{i=1}^{n} \alpha_i x_i$, una combinación lineal de elementos de E, obtenemos:

$$f\left(\sum_{i=1}^{n} \alpha_i x_i\right)=\sum_{i=1}^{n} \alpha_i f(x_i).$$

Esta propiedad se demuestra por inducción; proponemos la demostración al estudiante.

Por último, una propiedad que también reafirma el porqué de la definición de aplicación lineal, es la siguiente:

Si f es una aplicación lineal del K-espacio E en el K-espacio F, y E' es un subespacio de E, entonces $f(E')$ es un subespacio de F y si G es un subespacio de F, entonces $f^{-1}(G)$ es un subespacio de E.

En efecto, si $x', y' \in f(E')$, debemos probar que $\alpha x' + \beta y' \in f(E')$ para todo $\alpha, \beta \in K$:

Como $x' = f(x)$ para algún $x \in E'$, $y' = f(y)$ para algún $y \in E'$, entonces $\alpha x' + \beta y' = \alpha f(x) + \beta f(y) = f(\alpha x + \beta y) \in f(E')$
pues $\alpha x + \beta y \in E'$ por ser este un subespacio de E.

De este modo queda probado que si E' es un subespacio de E, entonce $f(E')$ es un subespacio de F.

Si ahora $x, y \in f^{-1}(G)$, debemos probar que $\alpha x + \beta y \in f^{-1}(G)$ para todo α, $\beta \in K$.

$$\alpha x + \beta y \in f^{-1}(G) \iff f(\alpha x + \beta y) \in G \text{ y}$$
$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y) \in G$$

ya que G es un subespacio y $f(x)$ y $f(y)$ pertenecen a G. Luego $\alpha x + \beta y \in f^{-1}(G)$, y así queda demostrado que $f^{-1}(G)$ es un subespacio de E.
Vamos a resumir estas propiedades en las proposiciones siguientes:

PROPOSICIÓN 2.1

Sean E y F dos espacios vectoriales sobre K, f una aplicación de E en F. Entonces, f es lineal si y solo si para todo $x, y \in E$ y $\alpha, \beta \in K$ se cumple:

$$f(\alpha x + \beta y) = \alpha f(x) + \beta f(y).$$

PROPOSICIÓN 2.2

Sean E y F dos espacios vectoriales sobre K, f una aplicación lineal de E en F. Entonces se cumple:

(1) $f(0) = 0$.
(2) Si E' es un subespacio de E, $f(E')$ es un subespacio de F.
(3) Si G es un subespacio de F, $f^{-1}(G)$ es un subespacio de E.

2.2 Caracterización de las aplicaciones lineales por las imágenes de los elementos de una base del dominio

Hemos señalado que mediante el concepto *base de un espacio vectorial*, podemos reducir el estudio del comportamiento del espacio completo al de los elementos de una base. ¿Esta idea tendrá vigencia también en el estudio de las aplicaciones lineales?

Si tenemos, por ejemplo, una aplicación lineal f entre dos espacios vectoriales de dimensión finita E y F y $(a_i)_{1\leqslant i\leqslant n}$ es una base de E y $x\in E$, entonces $x=\sum_{i=1}^{n} x_i\, a_i$, y aplicando las propiedades de las aplicaciones lineales,

$$f(x)=f\left(\sum_{i=1}^{n} x_i a_i\right)=\sum_{i=1}^{n} x_i f(a_i).$$

Si tomamos otro $y\in E$ tal que $y=\sum_{i=1}^{n} y_i a_i$, encontramos que

$$f(y)=\sum_{i=1}^{n} y_i f(a_i).$$

Esto ocurrirá para todo vector de E; es decir, una vez fijada una base de E, la expresión de la imagen de un vector cualquiera de E depende de las coordenadas del vector en dicha base y de las imagénes por f de los elementos de la base. ¿Será, entonces, suficiente para tener perfectamente determinada una aplicación lineal de E en F, fijar una base de E y conocer las imágenes por f de dicha base? El siguiente teorema dá respuesta a este problema.

TEOREMA 2.1

Sean E y F dos espacios vectoriales sobre K de dimensiones n y m respectivamente, $(a_1, a_2,\ldots, a_n)$ una base de E y $b_1, b_2,\ldots, b_n$ vectores cualesquiera de F. Entonces existe una única aplicación lineal f de E en F tal que $f(a_i)=b_i$, $i=1,\ldots,n$.

Demostración

Probemos primero que existe una aplicación f que satisface las condiciones del teorema.

En general, las pruebas de existencia son más delicadas que otro tipo de demostración, y en este caso ayuda el trabajo en forma constructiva; esto es, si tenemos f lineal que cumple $f(a_i)=b_i$, $i=1,\ldots, n$, ¿cómo será $f(x)$? Veamos.

Sea $x=\sum_{i=1}^{n} x_i a_i$. Entonces

$$f(x)=f\left(\sum_{i=1}^{n} x_i a_i\right)=\sum_{i=1}^{n} x_i f(a_i)=\sum_{i=1}^{n} x_i b_i.$$

Esto indica cómo definir f.

Así, definimos $f\colon E\longrightarrow F$ dada por

$$f(x)=\sum_{i=1}^{n} x_i b_i, \text{ donde } x=\sum_{i=1}^{n} x_i a_i.$$

Es evidente que f está bien definida pues las coordenadas x_i de x en $(a_1, a_2, \ldots, a_n)$ son únicas.

Probemos que f es lineal.

Sean $x=\sum_{i=1}^{n} x_i a_i \quad y=\sum_{i=1}^{n} y_i a_i$. Entonces,

$$f(\alpha x+\beta y)=f\left(\sum_{i=1}^{n}(\alpha x_i+\beta y_i)a_i\right)=\sum_{i=1}^{n}(\alpha x_i+\beta\, y_i)\,b_i=$$

$$=\alpha\sum_{i=1}^{n}x_i b_i+\beta\sum_{i=1}^{n}y_i b_i=\alpha f(x)+\beta f(y).$$

Es evidente también de la definición de f que $f(a_i)=b_i$.

De ese modo queda probado que la aplicación f que hemos definido es una aplicación lineal de E en F, tal que la imagen por dicha aplicación de cada vector a_i de la base es b_i.

Supongamos que existe, además de f, otra aplicación lineal g de E en F tal que $g(a_i)=b_i$ para todo i. Entonces, para todo $x\in E$, $x=\sum_{i=1}^{n} x_i a_i$, se cumple:

$$g(x)=g\left(\sum_{i=1}^{n}x_i a_i\right)=\sum_{i=1}^{n}x_i g(a_i)=\sum_{i=1}^{n}x_i b_i=f(x).$$

De aquí resulta que $f=g$.

Este resultado confirma la idea que habíamos esbozado de que para conocer una aplicación lineal, es suficiente conocerla para los elementos de una base del dominio, y tiene numerosas aplicaciones. Por ejemplo, si queremos demostrar que dos aplicaciones lineales entre los espacios E y F son iguales, es suficiente comprobar si coinciden sus valores en una base cualquiera de E. Otra aplicación está presente cuando hay que definir una aplicación lineal entre dos espacios E y F; en este caso se define la aplicación para los elementos de una base de E.

Estos ejemplos reafirman la característica que poseen las bases de los espacios vectoriales, en el sentido de que muchos resultados se obtienen realizando el trabajo con los elementos de una base.

Veamos algunos ejemplos que permitirán profundizar en el significado del teorema.

Ejemplos

1. ¿Existe una aplicación lineal f de $\mathbb{R}^3$ en $\mathbb{R}^3$ tal que $f(1,1,1)=(0,0,1)$, $f(1,1,0)=(0,1,1)$, $f(1,0,0)=(1,0,0)$?
   El teorema 2.1 establece que es suficiente conocer los valores de f en una base, y en este caso, la aplicación f se conoce en los vectores $(1,1,1)$, $(1,1,0)$, $(1,0,0)$, los cuales forman una base de $\mathbb{R}^3$. Por tanto, existe una única aplicación lineal de $\mathbb{R}^3$ en $\mathbb{R}^3$ que cumple la condición dada.

2. ¿Existe una aplicación lineal f de $\mathbb{R}^3$ en $\mathbb{R}^2$ tal que $f(1,-1,1)=(1,0)$ y $f(1,1,1)=(0,1)$?

El teorema plantea que para poder afirmar la existencia de una aplicación lineal, es necesario conocer las imágenes de una base del dominio. $(1,-1,1)$ y $(1,1,1)$ no constituyen una base de $\mathbb{R}^3$, pero no es difícil comprobar que son linealmente independientes y, por tanto, podemos formar una base de $\mathbb{R}^3$ a partir de ellos.

Sean, por ejemplo, $b_1=(1,-1,1)$, $b_2=(1,1,1)$, $b_3=(0,0,1)$. Entonces, según el teorema 2.1, podemos definir f por

$$f(1,-1,1)=(1,0),$$
$$f(1,1,1)=(0,1),$$
$$f(0,0,1)=(1,-1).$$

En este caso, el valor que asignamos a $f(0,0,1)$ es arbitrario, pues no se exige ninguna condición adicional para f. Más adelante mostraremos ejemplos en los cuales se exige que f cumpla condiciones adicionales, lo que restringirá el valor que asignemos a $f(0,0,1)$.

Si queremos buscar la expresión general de f para un vector x de $\mathbb{R}^3$, debemos expresar primeramente x en la base (b_1, b_2, b_3).

Sea $x=(x_1,x_2,x_3)\in\mathbb{R}^3$.

$$(x_1,x_2,x_3)=\alpha\,(1,-1,1)+\beta\,(1,1,1)+\gamma\,(0,0,1)$$

$$\begin{cases} \alpha+\beta=x_1 \\ -\alpha+\beta=x_2 \\ \alpha+\beta+\gamma=x_3 \end{cases}$$

De aquí,

$$\alpha=\frac{x_1-x_2}{2},\quad \beta=\frac{x_1+x_2}{2},\quad \gamma=x_3-x_1.$$

$$f(x_1,x_2,x_3)=\left(\frac{x_1-x_2}{2}\right)f(b_1)+\left(\frac{x_1+x_2}{2}\right)f(b_2)+(x_3-x_1)f(b_3)$$

$$=\left(\frac{x_1-x_2}{2}\right)(1,0)+\left(\frac{x_1+x_2}{2}\right)(0,1)+(x_3-x_1)\;(1,-1)$$

$$=\left(\frac{2x_3-x_1-x_2}{2},\ \frac{3x_1+x_2-2x_3}{2}\right).$$

Luego la respuesta es afirmativa en este caso.

3. Consideremos los vectores de $\mathbb{R}^2$:

$$a_1=(1,-1),\ a_2=(2,-1),\ a_3=(-3,2);$$
$$b_1=(1,0),\quad b_2=(0,1),\quad b_3=(1,1).$$

¿Existirá una aplicación lineal f de $\mathbb{R}^2$ en $\mathbb{R}^2$ tal que $f(a_i)=b_i$, $i=1,2,3$?

Utilizando de nuevo el teorema 2.1, una aplicación lineal de $\mathbb{R}^2$ en $\mathbb{R}^2$ quedará determinada al fijar las imágenes, por ella, de una base de $\mathbb{R}^2$. En este caso (a_1,a_2,a_3) no es una base, pero de ella podemos extraer una base, por ejemplo (a_1,a_2). Entonces existirá una única aplicación lineal de $\mathbb{R}^2$ en $\mathbb{R}^2$ tal que $f(a_1)=b_1$, $f(a_2)=b_2$.
Pero tenemos otra condición más; esta es que $f(-3,2)=(1,1)$.
Está claro que ya f está condicionada por los valores que toma en a_1 y a_2; veamos si, además, satisface que $f(a_3)=b_3$. Busquemos la expresión general de f:

Sea $(x_1,x_2)\in\mathbb{R}^2$. Sus coordenadas en la base (a_1,a_2) están dadas por $(x_1,x_2)=\alpha(1,-1)+\beta(2-1)$:

$$(x_1,x_2)=(-x_1-2x_2)a_1+(x_1+x_2)\ a_2$$

Luego

$$\begin{aligned}(x_1,\ x_2)&=(-x_1-\ 2x_2)f(a_1)+(x_1+x_2)\ f(a_2)\\&=(-x_1-2x_2,\ x_1+x_2).\end{aligned}$$

Hallemos pues $f(-3,\ 2)$:

$$f(-3,2)=(3-4,\ \ -3+2)=(-1,-1).$$

De aquí que no se cumplan las condiciones pedidas y no exista la aplicación lineal buscada.

Nota. Hemos demostrado el teorema que caracteriza una aplicación lineal por las imágenes de los elementos de una base del dominio, en el caso de dos espacios de dimensión finita.
Un teorema análogo puede enunciarse y demostrarse para el caso en que los espacios no son de dimensión finita. Si E, F son espacios vectoriales sobre K, $(a_i)_{i\in I}$ una base de E y $(b_i)_{i\in I}$ una familia de vectores de F, entonces existe una única aplicación lineal f de E en F tal que $f(a_i)=b_i$ para todo $i\in I$.
La demostración no difiere en lo esencial de la realizada en el caso de dimensión finita; solo hay que tener cuidado al expresar las combinaciones lineales. El estudiante puede intentar realizarla siguiendo los pasos dados en la demostración del teorema 2.1. También son válidas para los espacios de dimensión infinita las observaciones hechas sobre cómo utilizar el teorema.

2.3 Representación matricial de una aplicación lineal entre espacios vectoriales de dimensión finita

El teorema estudiado en el epígrafe anterior, que caracteriza las aplicaciones lineales entre espacios vectoriales de dimensión finita mediante las imágenes de los elementos de una base del dominio, da lugar a una de las ideas más importantes del Álgebra lineal: la de la representación matricial de una aplicación lineal.

Esta idea consiste en asociar matrices a cada aplicación lineal, de forma tal que el cálculo con dichas matrices y el análisis de sus características permitan sacar conclusiones sobre las aplicaciones lineales. Esta idea tiene antecedentes en dos hechos: primero, todo K-espacio de dimensión n es isomorfo a K^n, y segundo, las matrices de tamaño $m \times n$ con coeficientes en K, según vimos en los ejemplos anteriores, determinan aplicaciones lineales de K^n en K^m, y esencialmente estas son las únicas aplicaciones lineales de un espacio E de dimensión n. Esta idea tendrá, entre otras, las ventajas de las posibilidades de implementación computacional que poseen las matrices.

Matriz asociada a una aplicación lineal

Sea f una aplicación lineal entre los espacios vectoriales E y F de dimensiones n y m, respectivamente. Sabemos que dada una base $(a_1,\ldots,a_n)$ de E, f está determinada únicamente por los valores de $f(a_j)$, $j=1,2,\ldots,n$, los cuales son vectores de F, y por ello podemos expresarlos en una base $(b_1,b_2,\ldots,b_m)$ de F.

Sea $f(a_j)=\sum_{i=1}^{m} a_{ij}b_i = a_{1j}b_1+\ldots+a_{ij}b_i+\ldots+a_{mj}b_m$.

Entonces, para $x\in E$, $x=\sum_{j=1}^{n} x_j a_j$

$$f(x)=f\left(\sum_{j=1}^{n} x_j a_j\right)=\sum_{j=1}^{n} x_j f(a_j)=\sum_{j=1}^{n} x_j \sum_{i=1}^{m} a_{ij}b_i=\sum_{i=1}^{m}\sum_{j=1}^{n} a_{ij}x_j b_i.$$

De aquí que la coordenada $i-$ésima y_i de $f(x)$ en la base $(b_1, b_2,\ldots, b_m)$ esté dada por

$$y_i=\sum_{j=1}^{n} a_{ij}x_j,\quad i=1,2,\ldots,m. \quad (*)$$

De lo anterior resulta que para todo $x\in E$ podemos calcular $f(x)$ a partir de los a_{ij}, $i=1,2,\ldots,m$, $j=1,2,\ldots,n$, y las coordenadas de x en la base $(a_1,a_2,\ldots, a_n)$. Por otra parte la expresión (*) sugiere la siguiente identidad matricial:

$$\begin{pmatrix} y_1\\ y_2\\ \cdot\\ \cdot\\ \cdot\\ y_m \end{pmatrix}=\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1n}\\ a_{21} & a_{22} & \ldots & a_{2n}\\ & & & \\ & & & \\ & & & \\ a_{m1} & a_{m2} & \ldots & a_{mn} \end{pmatrix}\begin{pmatrix} x_1\\ x_2\\ \cdot\\ \cdot\\ \cdot\\ x_n \end{pmatrix}$$

Si denotamos por

$$X=\begin{pmatrix} x_1\\ x_2\\ \cdot\\ \cdot\\ \cdot\\ x_n \end{pmatrix} \qquad Y=\begin{pmatrix} y_1\\ y_2\\ \cdot\\ \cdot\\ \cdot\\ y_m \end{pmatrix}$$

a los vectores columna formados por las coordenadas de x en la base $(a_1,a_2,\dots,a_n)$ y $f(x)$ en la base $(b_1,b_2,\dots,b_m)$, respectivamente, y por

$$A=(a_{ij})_{\substack{1\leq i\leq m\\ 1\leq j\leq n}},$$

tendremos la igualdad matricial $Y=AX$, que de hecho permite calcular $f(x)$ si se conoce x.

Los a_{ij} se obtienen al expresar los vectores $f(a_j)$ en la base (b_j). A la matriz A la llamaremos *matriz asociada* a f en las bases (a_j) y (b_i). Precisemos este concepto mediante la definición siguiente:

DEFINICIÓN 2.2

Sean f una aplicación lineal entre los K-espacios E y F de dimensiones n y m, respectivamente, $(a_1,a_2,\dots,a_n)$ una base de E, $(b_1,b_2,\dots,b_m)$ una base de F. Se denomina *matriz asociada a* f en las bases (a_j) y (b_i) y se denota por $M(f,(a_j),(b_i))$, a la matriz cuyo coeficiente (i,j) está dado por la coordenada i-ésima en base (b_i) del vector $f(a_j)$.

La matriz $A=M(f,(a_j),(b_i))$ permite calcular, conocidas las coordenadas x_j de x en $(a_j)_{1\leq j\leq n}$, las coordenadas y_i de $f(x)$ en $(b_i)_{1\leq i\leq m}$ mediante la expresión $Y=AX$, donde

$$X=\begin{pmatrix}x_1\\x_2\\ \cdot\\ \cdot\\ \cdot\\x_n\end{pmatrix},\quad Y=\begin{pmatrix}y_1\\y_2\\ \cdot\\ \cdot\\ \cdot\\y_m\end{pmatrix}$$

Obsérvese que las columnas de A son los vectores de coordenadas en base $(b_i)_{1\leq i\leq m}$ de los vectores $f(a_j)_{1\leq j\leq n}$, y que el número de columnas de A está dado por $n=\dim E$ y el número de filas por $m=\dim F$.

Ejemplo

Sea $f\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ dada por $f(x_1,x_2,x_3)=(x_1+x_2,2x_3-x_1)$.

Consideremos $(e_j)_{1\leq j\leq 3}$ la base canónica de $\mathbb{R}^3$ y $(e_i')_{1\leq i\leq 2}$ la base canónica de $\mathbb{R}^2$. Entonces para hallar $M(f,(e_j),(e_i'))$ calculamos:

$$f(e_1)=(1,-1)=e_1'-e_2'$$
$$f(e_2)=(1,0)=e_1'$$
$$f(e_3)=(0,2)=2e_2'$$

De aquí que

$$M(f,(e_j),(e_i''))=\begin{pmatrix}1&1&0\\-1&0&2\end{pmatrix},$$

y puede comprobarse fácilmente que

$$\begin{pmatrix}x_1+x_2\\2x_3-x_1\end{pmatrix}=\begin{pmatrix}1&1&0\\-1&0&2\end{pmatrix}\begin{pmatrix}x_1\\x_2\\x_3\end{pmatrix}$$

Si consideramos ahora en $\mathbb{R}^3$ y $\mathbb{R}^2$ las bases

$(a_j)_{1\leq j\leq 3}$, donde $a_1=(1,0,-1)$, $a_2=(1,1,1)$, $a_3=(1,0,0)$;
$(b_i)_{1\leq i\leq 2}$, donde $b_1=(0,1)$ $b_2=(1,0)$,

$M(f,(a_j),(b_i))$ se calculará como sigue:

$$f(a_1)=f(1,0,-1)=(1,-3)$$
$$f(a_2)=f(1,1,1)=(2,1)$$
$$f(a_3)=f(1,0,0)=(1,-1)$$

Los vectores $f(a_1)$, $f(a_2)$, $f(a_3)$ están expresados en base canónica, por lo que es necesario buscar sus coordenadas en la base (b_i), lo cual podemos hacer directamente o mediante la matriz de cambio de coordenadas adecuada, que en este caso será la de cambio de coordenadas de la base canónica a la base (b_i). Mostraremos cómo hacerlo por dos vías:

1ro. Por la vía directa, según vimos anteriormente:

$$(1,-3)=\alpha_1(0,1)+\beta_1(1,0)=\alpha_1 b_1+\beta_1 b_2$$
$$(2,1)=\alpha_2(0,1)+\beta_2(1,0)=\alpha_2 b_1+\beta_2 b_2$$
$$(1,-1)=\alpha_3(0,1)+\beta_3(1,0)=\alpha_3 b_1+\beta_3 b_2$$

Resolviendo los sistemas que resultan de las identidades anteriores, obtenemos:

$$f(a_1)=-3b_1+b_2$$
$$f(a_2)=b_1+2b_2$$
$$f(a_3)=-b_1+b_2$$

$$M(f,(a_j),(b_i))=\begin{pmatrix}-3 & 1 & -1\\ 1 & 2 & 1\end{pmatrix}.$$

2do. Por la vía de la matriz de cambio de coordenadas.

La matriz P de cambio de coordenadas de (e_i) a (b_i) es

$$P=\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}.$$

Aplicando P a $(1,-3)$, $(2,1)$ y $(1,-1)$, obtenemos:

$$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}1\\ -3\end{pmatrix}=\begin{pmatrix}-3\\ 1\end{pmatrix}$$

$$\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}\begin{pmatrix}2\\ 1\end{pmatrix}=\begin{pmatrix}1\\ 2\end{pmatrix}$$

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

Así llegamos al mismo resultado obtenido anteriormente, esto es

$$M(f,(a_i),(b_i)) = \begin{pmatrix} -3 & 1 & -1 \\ 1 & 2 & 1 \end{pmatrix}.$$

Queremos señalar que aunque en un ejemplo modelo como el anterior las dos vías implican un volumen de cálculo equivalente, la vía de las matrices de cambio de coordenadas es la idónea en problemas reales en los que el tamaño de las matrices involucradas exige la solución mediante computadoras.

Aún hay una tercera vía de cálculo de las coordenadas que señalamos al tratar los cambios de base y es la siguiente:

Tenemos

$$b_1 = e_2$$
$$b_2 = e_1.$$

Luego si

$$f(a_1) = e_1 - 3e_2,$$
$$f(a_2) = 2e_1 + e_2,$$
$$f(a_3) = e_1 - e_2,$$

resulta inmediatamente que

$$f(a_1) = -3b_1 + b_2,$$
$$f(a_2) = b_1 + 2\ b_2,$$
$$f(a_3) = -b_1 + b_2.$$

Obsérvese en este caso que la denominación de *matriz asociada a f* en las bases (a_j) y (b_i) no es superflua; esto es, que la matriz de f depende de las bases consideradas en los espacios E y F.

Un caso particular importante de matriz asociada a una aplicación lineal, es el de las matrices asociadas a los endomorfismos. En este caso, por razones prácticas, se considera la misma base en el dominio y el codominio. Así, si f es un endomorfismo de un espacio E de dimensión n y $(a_i)_{1\leqslant i\leqslant n}$ una base de E, consideramos $M(f,(a_i),(a_i))$ y la denotamos simplemente por $M(f,(a_i))$.

Ejemplos

1. Sean E el espacio de los polinomios de grado menor o igual que 3 y $(a_i)_{1\leqslant i\leqslant 4}$ la base de E dada por $a_1=1$, $a_2=t$, $a_3=t^2$, $a_4=t^3$ y $D: E \longrightarrow E$ dada por $D(p) = p'$. Entonces

$$M(D,(a_i)) = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

2. Sean E un espacio vectorial de dimensión n sobre K y $(a_i)_{1\leqslant i\leqslant n}$ una base de E; busquemos $M(id_E,(a_i))$.

Evidentemente, como $id_E(a_i)=a_i$, $i=1,\dots,n$, tenemos:

$$M(\mathrm{id}_E,(a_i))=\begin{pmatrix}1 & 0 & \dots & 0\\ 0 & 1 & \dots & 0\\ \dots & \dots & \dots & \dots\\ 0 & 0 & \dots & 1\end{pmatrix}=I_n.$$

Queremos destacar que pudiera considerarse también para los endomorfismos, la matriz asociada a estos en dos bases distintas del espacio, pero esto en general no tiene interés ya que complicaría inútilmente los cálculos. No obstante, para algunos fines se utiliza a veces la posibilidad de buscar la matriz asociada a un endomorfismo en dos bases distintas del espacio.

Por ejemplo, busquemos la matriz que resulta al considerar $M(\mathrm{id}_E,(a_i),(a_i'))$, donde (a_i) y (a_i') son dos bases cualesquiera de E. Las columnas de la matriz estarán formadas por las coordenadas en la base $(a_i')_{1\leqslant i\leqslant n}$ de los vectores de la base $(a_i)_{1\leqslant i\leqslant n}$, y esto no es más que la matriz de cambio de coordenadas de la base (a_i) a la base (a_i'). Así, podemos expresar la matriz de cambio de coordenadas de (a_i) a (a_i') por $M(\mathrm{id}_E,(a_i),(a_i'))$.

Está claro que $M(id_E,(a_i'),(a_i))$, será la matriz de cambio de coordenadas de (a_i') a (a_i).

Obsérvese que en estos casos la matriz de la aplicación identidad no es I_n, a menos que realicemos el cambio de coordenadas trivial de la base (a_i) a la base (a_i).

Una última observación: la matriz $M(id_E,(a_i),(a_i'))$, que conocemos por matriz de cambio de coordenadas de la base (a_i) a la base (a_i'), se denomina también *matriz de cambio de base* de la base (a_i') a la base (a_i) y se denota MCB (de (a_i') a (a_i)). Esta denominación de matriz de cambio de base resulta de que si consideramos el endomorfismo de E dado por $f(a_i')=a_i, i=1,\dots,n$, y buscamos $M(f,(a_i'))$, evidentemente

$$M(f,(a_i'))=M(id_E,(a_i),(a_i'))$$

y la acción de f es transformar la base (a_i'), en la base (a_i).

De la definición de matriz asociada a una aplicación lineal f de E en F en dos bases fijas (a_j) y (b_i) de dichos espacios, resulta que dicha matriz es única.

Por otra parte, en los ejemplos de aplicaciones lineales vimos la de $\mathbb{R}^n$ en $\mathbb{R}^m$, en la que dada una matriz A de orden $(m\times n)$, la aplicación que a $X\in\mathbb{R}^n$ le hace corresponder AX es lineal. En otras palabras, toda matriz de tamaño $m\times n$ determina una aplicación lineal de $\mathbb{R}^n$ en $\mathbb{R}^m$.

Este ejemplo lleva a considerar la siguiente pregunta más general:

Sean $A\in M_{m,n}(K)$, E,F, espacios vectoriales sobre K, y $(a_j)_{1\leqslant j\leqslant n}$ y $(b_i)_{1\leqslant i\leqslant m}$ bases de E y F, respectivamente. ¿Existirá una aplicación lineal de E en F tal que $M(f,(a_j),(b_i))=A$? La respuesta es afirmativa, ya que podemos considerar las columnas de la matriz como los vectores de coordenadas de n vectores $e_1, e_2, \ldots, e_n$ de F, esto es:

$$e_j = \sum_{i=1}^{m} a_{ij}\, b_i,\ j = 1,\ 2,\ \ldots,\ n.$$

Según el teorema de caracterización de las aplicaciones lineales por las imágenes de los elementos de una base, existe una única aplicación lineal de E en F tal que

$$f\,(a_j) = e_j,$$

y es evidente que de la definición de matriz asociada a una aplicación lineal resulta:

$$M\ (f,\ (a_j),\ (b_i)) = A.$$

Llegamos, pues, al resultado siguiente:

TEOREMA 2.2

Si se consideran dos espacios vectoriales E y F de dimensiones n y m, respectivamente, y (a_j), (b_i) bases fijas de E y F, la correspondencia dada por

$$f \rightsquigarrow M\ (f, (a_j),\ b_i))$$

que a cada aplicación lineal de E en F le hace corresponder su matriz asociada en las bases (a_j), (b_i), es una correspondencia biyectiva.

Es necesario destacar que la correspondencia entre aplicaciones lineales es biyectiva solamente si consideramos bases prefijadas en los espacios E y F.

Así, cuando consideramos la aplicación de $\mathbb{R}^n$ en $\mathbb{R}^m$ dada por $X \rightsquigarrow AX$, donde $A \in M_{(m,\ n)}\ (K)$, dicha expresión puede representar en realidad más de una aplicación lineal, según consideremos que X y AX representan vectores de coordenadas de un vector de $\mathbb{R}^n$ y su imagen, respectivamente, en diferentes bases de estos espacios. Claro está que si no se hace otra aclaración, consideraremos que la expresión $X \rightsquigarrow AX$ se refiere a una aplicación donde los vectores X y AX están expresados en las bases canónicas de $\mathbb{R}^n$ y $\mathbb{R}^m$, respectivamente, y en este sentido muchas veces en la práctica se identifican" las matrices de orden $m \times n$ con las aplicaciones lineales de $\mathbb{R}^n$ en $\mathbb{R}^m$.

La correspondencia señalada entre aplicaciones lineales de E en F, con dim $E = n$ y dim $F = m$, y matrices de orden $m \times n$ para bases fijas (a_j) y (b_i) de E y F, respectivamente, permitirá, según veremos, pasar de resultados obtenidos con las matrices a resultados sobre aplicaciones lineales; pero la dependencia de dichas correspondencias, de las bases tomadas en los espacios, exige ocuparse de ciertos detalles antes de poder sacar conclusiones para las aplicaciones lineales, a partir de las obtenidas sobre matrices.

En primer lugar será necesario establecer qué relación existe entre las matrices asociadas a una misma aplicación lineal en bases distintas del espacio.

Efecto del cambio de base sobre la matriz de una aplicación lineal

Sean E, F espacios vectoriales sobre K de dimensiones n y m, respectivamente, $f : E \longrightarrow F$ una aplicación lineal, $(a_j)_{1\leqslant j\leqslant n}$ $(a'_j)_{1\leqslant j\leqslant n}$ bases de E, $(b_i)_{1\leqslant i\leqslant m}$, $(b'_i)_{1\leqslant i\leqslant m}$ bases de F. Designemos $A=M(f, (a_j), (b_i))$, $A'=M(f, (a'_j), b'_j))$. Queremos encontrar la relación que existe entre A y A'.

Llamemos X al vector columna de las coordenadas de $x\in E$ en la base (a_j) y X' al de las coordenadas de x en la base (a'_j); Y será el vector columna de las coordenadas de $f(x)$ en base (b_i) y Y' el vector columna de las coordenadas de $f(x)$ en (b'_i). Entonces $Y=AX$ y $Y'=A'X'$.

La relación entre X y X' y Y y Y' estará dada por la matriz P de cambio de coordenadas de la base (a'_j) a la base (a_j) y la matriz Q de cambio de coordenadas de (b'_i) a (b_i) mediante

$$X=PX' \text{ y } Y=QY'.$$

De aquí que de $Y=AX$ obtenemos, utilizando las identidades anteriores:

$$QY'=APX'.$$

Por tanto, $Y'=Q^{-1}APX'$ ya que Q es inversible por ser una matriz de cambio de coordenadas. Pero como la matriz asociada a una aplicación lineal en un par de bases dadas del espacio es única, de la relación anterior entre los vectores de coordenadas de x y $f(x)$ en las bases (a'_j) y (b'_i), respectivamente, resulta:

$$A'=Q^{-1}AP.$$

En el caso particular de los endomorfismos, para los cuales hemos convenido tomar la misma base en el dominio y en el espacio de llegada, la relación entre las matrices de un endomorfismo al cambiar las bases del espacio es más simple.

Sean $f : E \longrightarrow E$ un endomorfismo, (a_j) y (a'_j) bases de E, $A=M(f, (a_j))$, $A'=M(f, (a'_j))$.

Aplicando la relación obtenida anteriormente tenemos:

$$A'=P^{-1}AP,$$

donde P es la matriz de cambio de coordenadas de la base (a'_j) a la base (a_j).

Ejemplos

1. Sea f: $\mathbf{R}^3 \longrightarrow \mathbf{R}^2$ dada por $f(x, y, z)=\left(\dfrac{2z-x-y}{2}, \dfrac{3x+y-2z}{2}\right)$.

Hallemos $M\ (f,\ (e_j\ ),\ (e_j'))$, donde (e_j) es la base canónica de $\mathbb{R}^3$ y (e_j') la canónica de $\mathbb{R}^2$.

$$f\,(1,\ 0,\ 0)=\left(-\frac{1}{2},\ \frac{3}{2}\right),\ f\,(0,\ 1,\ 0)=\left(-\frac{1}{2},\ \frac{1}{2}\right),$$

$$f\,(0,\ 0,\ 1)=(1,\ -1).$$

Luego

$$M\ (f,\ (e_i),\ (e_j'))=\begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & 1 \\ \frac{3}{2} & \frac{1}{2} & -1 \end{pmatrix}$$

Sean ahora las bases de $\mathbb{R}^3$ y $\mathbb{R}^2$ dadas por

$a_1=(1,\ -1,\ 1),\qquad a_2=(1,\ 1,\ 1),\qquad a_3=(0,\ 0,\ 1);$
$b_1=(1,\ 1),\qquad b_2=(0,\ 1).$

Podemos hallar $M\ (f,\ (a_j),\ (b_i))$ mediante le definición. Para ello debemos calcular $f\ (a_1),\ f\ (a_2),\ f\ (a_3)$ y expresar el resultado en base (b_i). Así,

$f\ (a_1)=(1,0),$
$f\ (a_2)=(0,1),$
$f\ (a_3)=(1,-1).$

$(1,\ 0)=\alpha\ (1,1)+\beta\ (0,1);\qquad \alpha=1,\quad \beta=-1$
$(0,1)=\gamma(1,1)\ +\delta\ (0,1);\qquad \gamma=0,\quad \delta=1$
$(1,-1)=\varphi\ (1,1)+\Psi\ (0,1);\qquad \varphi=1,\quad \Psi=-2$

De aquí resulta:

$f\ (a_1)=b_1-b_2$
$f\ (a_2)=-b_2$
$f\ (a_3)=b_1-2b_2$

$$A'=M\ (f,\ (a_j),\ (b_i))=\begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & -2 \end{pmatrix}$$

Pero también podemos hallar $A'=M(f,(a_j),\ (b_i))$ mediante la relación

$A'=Q^{-1}AP,$

donde P es la matriz de cambio de coordenadas de (a_j) a (e_j) y Q la matriz de cambio de coordenadas de (b_i) a (e'_i). En efecto, P está dada por

$$\begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

y Q por

$$\begin{pmatrix} 1 & 0 \\ 1 & 1 \end{pmatrix}.$$

Luego $Q^{-1}=\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix}.$

$$A'=\begin{pmatrix} 1 & 0 \\ -1 & 1 \end{pmatrix} \begin{pmatrix} -\frac{1}{2} & -\frac{1}{2} & 1 \\ \frac{3}{2} & \frac{1}{2} & -1 \end{pmatrix} \begin{pmatrix} 1 & 1 & 0 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

$$=\begin{pmatrix} 1 & 0 & 1 \\ -1 & 1 & -2 \end{pmatrix}.$$

La relación $A'=Q^{-1}AP$ no solo tiene el valor de establecer cómo cambia la matriz de una aplicación lineal al cambiar las bases del espacio, sino también desde el punto de vista del cálculo, pues aunque en el ejemplo que acabamos de realizar, las dos vías mostradas tienen prácticamente el mismo volumen de cálculo, no debemos olvidar que en un ejemplo real, las matrices involucradas pueden ser de gran tamaño y los métodos matriciales son entonces los adecuados para la implementación computacional.

Está claro que hemos estudiado el caso más general, esto es, cuando cambiamos las bases en el dominio y el codominio de la aplicación, pero puede ocurrir que solo nos interese el cambio de base en uno de estos espacios, lo cual constituye un caso particular, pues entonces una de las matrices de cambio de coordenadas es la identidad.

En el ejemplo anterior podemos hallar $A''=M\ (f,\ (a_j),\ (e_i'))$ mediante

$$A''=AP=\begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & -1 \end{pmatrix}.$$

El estudiante puede comprobar, aplicando la definición de matriz asociada a una aplicación lineal en dos bases dadas, que se llega al mismo resultado.

2. Sea $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ dada por $f\ (x,\ y)=(-x-2y,\ x+y)$.
   Aquí, si designamos por (e_i) a la base canónica de $\mathbb{R}^2$, entonces

$$A=M\ (f,\ (e_i))=\begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}.$$

Si buscamos ahora $A'=M\ (f,\ (a_i))$, donde $a_1=(1,-1)$, $a_2=(2,-1)$, tenemos

$$f(a_1)=(1,0)\,;\ f(a_1)=-a_1+a_2$$
$$f(a_2)=(0,1)\,;\ f(a_2)=-2a_1+a_2$$

$$A'=\begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}$$

En este caso volvemos a encontrar la misma matriz. Esto se explica al realizar el [illegible] por vía matricial. Tenemos que

$$A'=P^{-1}AP,$$

donde P es la matriz de cambio de coordenadas de (a_i) a (e_i).

$$P=\begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix}$$

pero

$$P^{-1}=\begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}.$$

De aquí que

$$A'=\begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}\begin{pmatrix} 1 & 2 \\ -1 & -1 \end{pmatrix}$$
$$=\begin{pmatrix} -1 & -2 \\ 1 & 1 \end{pmatrix}.$$

El hecho de obtener de nuevo la misma matriz para f en base (a_i) es puramente casual y se debe a las características de las bases tomadas y de la aplicación f.

En el cálculo matricial se pone de manifiesto que $P^{-1}=A$, de lo que resulta $P^{-1}AP=P^{-1}\,P^{-1}\,P=P^{-1}=A$.

2.4 Matrices equivalentes y matrices semejantes

En el epígrafe anterior, al establecer la relación existente entre matrices asignadas a una misma aplicación lineal en pares de bases distintas del dominio y el espacio de llegada de la aplicación, hemos encontrado una relación entre las matrices A y B dada por $B=RAS$, donde R y S son inversibles. Definamos dicha relación entre matrices.

DEFINICIÓN 2.3

Dos matrices A y B de $M_{m,n}(K)$ se denominan *matrices equivalentes* si y solo si existen $R\in M_m(K)$ y $S\in M_n(K)$ inversibles tales que $B=RAS$.

De acuerdo con esta definición, podemos decir que las matrices asociadas a una misma aplicación lineal en pares de bases distintas son equivalentes.

Nota: Esta relación entre matrices cumple las propiedades siguientes:

(1) Para toda matriz A, A es equivalente a A, ya que $A=I_m A I_n$.

(2) Si A es equivalente a B, entonces B es equivalente a A; pues
A es equivalente a $B \Longleftrightarrow B=RAS$ con R y S inversibles.
Entonces $A=R^{-1}BS^{-1}$ donde R^{-1} y S^{-1} son inversibles. De aquí que B sea equivalente a A.

(3) Si A es equivalente a B y B es equivalente a C, entonces A es equivalente a C.
En efecto:
A es equivalente a $B \Longleftrightarrow B=RAS$ con R y S inversibles,
B es equivalente a $C \Longleftrightarrow C=TBL$ con T y L inversibles.
Entonces $C=TRASL$, donde TR y SL son inversibles. Luego A es equivalente a C.
Puede demostrarse que si clasificamos los elementos de $M_{m,n}(K)$ mediante el criterio: *A y B están en una misma clase si y solo si A y B son equivalentes*, $M_{m,n}(K)$ quedará dividido en subconjuntos o clases disjuntos dos a dos y tales que su unión sea el conjunto $M_{m,n}(K)$.

Según los resultados del epígrafe anterior, es natural considerar el problema siguiente:

Si A y B son dos matrices de $M_{m,n}(K)$ equivalentes, ¿existirán, dados los espacios E y F de dimensiones n y m respectivamente, una aplicación lineal f de E en F y dos pares de bases (a_j) y (b_i) y (a_j') y (b_i') de E y F tales que

$$A=M(f,(a_j),(b_i)) \text{ y } B=M(f,(a_j'),(b_i'))?$$

Si recordamos que dadas una base en un espacio vectorial y una matriz inversible de orden n, existe otra base en el espacio vectorial tal que la matriz dada es la matriz de cambio de coordenadas la primera base a la segunda, entonces dada $B=RAS$, designamos $S=P$ y $R=Q^{-1}$.

Si seleccionamos una base (a_j) en E y otra base (b_i) en F, entonces existen otras dos bases (a_j') en E y (b_i') en F tales que P y Q son las matrices de cambio de coordenadas de (a_j') a (a_j) y de (b_i') a (b_i), respectivamente.

Sabemos, además, que dadas dos bases (a_j) y (b_i) de E y F, respectivamente, y una matriz A de orden $m\times n$, existe una única aplicación lineal f de E en F tal que

$$A=M(f,(a_j),(b_i)).$$

Demostraremos que $B=M(f\,(a_j'),(b_k'))$.

Sea X el vector de coordenadas de $x\in E$ en la base (a_j'). Entonces, APX dará el vector de coordenadas de $f(x)$ en (b_i'), y $Q^{-1}(APX)$ el vector de coordenadas de $f(x)$ en (b_i'). Luego por la unicidad de la matriz asociada a una aplicación lineal en bases fijas del espacio, se puede afirmar que

$$B = M(f, (a_j'), (b_i'))\,.$$

Podemos resumir los resultados hallados mediante el teorema siguiente.

TEOREMA 2.3

Dos matrices A y B son equivalentes si y solo si están asociadas a una misma aplicación lineal, en pares de bases distintas del dominio y el codominio de la aplicación.

En el caso particular de los endomorfismos, vimos que la relación existente entre las matrices asociadas a un endomorfismo en bases distintas era más simple; teníamos que si f es un endomorfismo de un espacio E y A y A' sus matrices en las bases (a_j) y (a_j') de E entonces,

$$A' = P^{-1}AP.$$

De manera similar a la definición de matrices equivalentes daremos la de matrices semejantes.

DEFINICIÓN 2.4

Dos matrices A y B de M_n (K) se denominan *matrices semejantes* si y solo si existe una matriz P de M_n (K) inversible tal que

$$B = P^{-1}AP.$$

Está claro que las matrices asociadas a un mismo endomorfismo en bases distintas del espacio son semejantes.

Nota. Para la relación de semejanza entre matrices pueden probarse las propiedades (1), (2) y (3) dadas anteriormente para la equivalencia de matrices. También es válida para la semejanza, la observación sobre la posibilidad de dividir mediante esta relación el conjunto M_n (K) en clases disjuntas y cuya unión es M_n (K). Dejamos al estudiante que realice el enunciado y la demostración de las propiedades (1), (2) y (3) para la semejanza de matrices.

Con un razonamiento completamente análogo al seguido en el caso de las matrices equivalentes, es posible demostrar que si dos matrices A y B de $M_n(K)$ son semejantes y E es un espacio de dimensión n sobre K, entonces existe un endomorfismo f de E y dos bases (a_j) y (a_j') de E tales que $A = M(f, (a_j))$ y $A' = M(f, (a_j'))$.

Podemos concluir:

TEOREMA 2.4

Dos matrices A y B de $M_n(K)$ son semejantes si y solo si están asociadas a un mismo endomorfismo f de E en dos bases distintas de dicho espacio.

Antes de continuar queremos llamar la atención del estudiante sobre lo siguiente:

La semejanza es solo aplicable a matrices cuadradas; ahora bien, aunque la semejanza es un caso particular de equivalencia, dos matrices cuadradas pueden ser equivalentes y no ser semejantes.

Ejemplo

Las matrices

$$\begin{pmatrix} 3 & -1 \\ 1 & 1 \end{pmatrix} \text{ y } \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$$

son equivalentes pero no semejantes.

En efecto, como

$$\begin{pmatrix} 3 & -1 \\ 1 & 1 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & -1 \end{pmatrix}$$

donde

$$\begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} \text{ y } \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \\ 1 & -1 \end{pmatrix}$$

son inversibles pero no mutuamente inversas, entonces las matrices dadas son equivalentes.

En cuanto a que no son semejantes, podemos plantear la existencia de P inversible tal que

$$P^{-1} = \begin{pmatrix} 3 & -1 \\ 1 & 1 \end{pmatrix}, \; P = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix} \text{ con } P = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \; ad - bc \neq 0,$$

y analizar el sistema no lineal de cuatro ecuaciones con cuatro incógnitas que resulta, para concluir que no existe tal P.

Este tratamiento implica dificultades de cálculo. Más adelante, en un próximo capítulo, estudiaremos un criterio para semejanza de matrices que permitirá decidir, en forma más sencilla, si dos matrices son o no semejantes.

En el próximo capítulo trataremos las matrices semejantes, ya que según vimos, son estas las que están asociadas a un mismo endomorfismo. El problema central que estudiaremos es la búsqueda de matrices lo más simples posible que representen a un endomorfismo f. Está claro que dichas matrices, una vez conocida una matriz A que represente a f, debemos buscarlas entre las semejantes a A. Esta idea está avalada por el hecho de que algunas características importantes de las matrices, como son sus determinantes y su rango, son invariantes por semejanza, lo que nos permite obtener datos importantes sobre los endomorfismos que ellas representan.

2.5 Operaciones con aplicaciones lineales

Para interpretar diversas situaciones, tanto teóricas como prácticas, es necesario utilizar operaciones con aplicaciones lineales, mediante las cuales construimos nuevas aplicaciones a partir de aplicaciones dadas.

Designemos por $L_K(E,F)$ al conjunto de las aplicaciones lineales del K-espacio E en el K-espacio F. Para definir operaciones con aplicaciones lineales, debemos considerar aplicaciones lineales de un cierto K-espacio E en otro K-espacio F; en otras palabras, trabajaremos con los elementos de $L_K(E,F)$. En el caso particular de los endomorfismos, designaremos al conjunto de los endomorfismos de un K-espacio E por $L_K(E)$ o End(E).

Las operaciones en $L_K(E,F)$ y $L_K(E)$ se definen de forma natural, como es usual hacerlo con las funciones.

DEFINICIÓN 2.5

Dadas $f,g\in L_K(E,F)$ se denomina *suma* de f y g a la aplicación $f+g$ dada por

$$(f+g)(x)=f(x)+g(x)\ \forall x\in E.$$

Evidentemente, $f+g$ es por definición, una aplicación de E en F; más aún, es lineal ya que para todo $\alpha,\beta\in K$, $x,y\in E$, se cumple:

$$\begin{aligned}(f+g)(\alpha x+\beta y)&=f(\alpha x+\beta y)+g(\alpha x+\beta y)\\&=\alpha f(x)+\beta f(y)+\alpha g(x)+\beta g(y)\\&=\alpha[f(x)+g(x)]+\beta[f(y)+g(y)]\\&=\alpha(f+g)(x)+\beta(f+g)(y).\end{aligned}$$

Esto es, $f+g\in L_k(E,F)\ \forall f,g\in L_K(E,F)$.

DEFINICIÓN 2.6

Dados $f\in L_K(E,F)$, $k\in K$, se denomina *producto* de k por f a la aplicación kf dada por

$$(kf)(x)=kf(x)\ \forall x\in E.$$

En este caso también es evidente que kf es una aplicación de E en F y que es lineal pues

$$\begin{aligned}(kf)(\alpha x+\beta y)&=kf(\alpha x+\beta y)\\&=k[\alpha f(x)+\beta f(y)]\\&=k\alpha f(x)+k\beta f(y)\\&=\alpha kf(x)+\beta kf(y)\\&=\alpha(kf)(x)+\beta(kf)(y)\ \forall x,y\in E,\forall\alpha,\beta\in K;\end{aligned}$$

De esta forma tenemos que $kf\in L_K(E,F)\ \ \forall k\in K$ y $\forall f\in L_K(E,F)$.

Las operaciones que acabamos de definir en $L_K(E,F)$ poseen las propiedades siguientes:

$\forall f,g,h\in L_K(E,F)$, $\forall k,k'\in K$ se cumple:

(1) $(f+g)+h=f+(g+h)$

(2) $f+g=g+f$

(3) $k(f+g)=kf+kg$

(4) $(k+k')f=kf+kf$

(5) $k(k'f)=(kk')f$

(6) $1\cdot f=f$

Demostraremos las propiedades (1) y (4) y dejaremos la demostración del resto como ejercicio para el estudiante.

Demostración de (1).

$$[(f+g)+h](x)=(f+g)(x)+h(x)=f(x)+g(x)+h(x)$$
$$=f(x)+[g(x)+h(x)]=f(x)+(g+h)(x)=[f+(g+h)](x)\ \forall\, x\in E.$$

Demostración de (4)

$$[(k+k')f](x)=(k+k')f(x)=kf(x)+k'(x)=(kf)(x)+(k'f)(x).$$

Obsérvese que para realizar las demostraciones se aplicaron solamente la definición de igualdad de aplicaciones y las definiciones de las operaciones, lo cual permite trabajar con las imágenes, que son vectores de F, y por tanto, usar las propiedades de las operaciones en F. Las propiedades restantes se demuestran por esta misma vía.

Si observamos las operaciones definidas en $L_K(E,F)$ y sus propiedades, encontramos que constituyen parte de los axiomas que harían de $L_K(E,F)$ un espacio vectorial sobre K para las operaciones de suma de aplicaciones y producto de un escalar por una aplicación; solo faltarían las propiedades relativas a la existencia de un vector nulo y un vector opuesto para la suma.

Sabemos que la aplicación nula dada por $0(x)=0\ \forall x\in E$, es una aplicación lineal y no es difícil demostrar que

$$0+f=f+0\ \ \forall f\in L_K(E,F).$$

Por otra parte, dada $f\in L_K(E,F)$, si definimos $-f$ de E en F por $(-f)(x)=-f(x)\ \forall x\in E$, esta será una aplicación lineal ya que

$$(-f)(\alpha x+\beta y)=-f(\alpha x+\beta y)=-\alpha f(x)-\beta f(y)=$$
$$=\alpha(-f(x))+\beta(-f(y))=\alpha(-f(x))+\beta(-f)(y)\ \forall\, \alpha\in K,\ x,y\in E.$$

$$f+(-f)=0 \text{ pues } (f+(-f))(x)=f(x)-f(x)=0\ \ \forall x\in E.$$

Podemos entonces afirmar que $L_K(E,F)$ es un espacio vectorial sobre K con las operaciones de suma y producto por un escalar de aplicaciones, y este hecho facilita el trabajo con las aplicaciones lineales y provee de un lenguaje muy adecuado para el tratamiento de diversos problemas.

Es posible hallar una base de $L_K(E,F)$ a partir de una base de E y una base de F. Lo haremos en el caso en que E y F son, ambos, de dimensión finita. Veremos, una vez más, la utilidad del teorema de caracterización de una aplicación lineal por las imágenes de los elementos de una base del dominio.

Base de $L_K(E,F)$

Consideremos que $\dim E=n$ y $(a_j)_{1\leq j\leq n}$ es una base de E, y que $\dim F=m$ y $(b_i)_{1\leq i\leq m}$ es una base de F.

Definamos en $L_K(E,F)$ las siguientes aplicaciones f_{kh}:

$f_{kh}(a_j) = 0$ si $j \neq h$, $k = 1, 2, \ldots, m$
$f_{kh}(a_j) = h_k$ si $j = h$, $h = 1, 2, \ldots, n$

Las aplicaciones f_{kh} están bien definidas y son lineales, ya que hemos dado las imágenes de los elementos de una base en E. Probemos que efectivamente constituyen una base de $L_K(E,F)$.

1) El conjunto $(f_{kh})_{\substack{1 \leqslant k \leqslant m \\ 1 \leqslant h \leqslant n}}$ es linealmente independiente.

Tenemos que probar que $\sum\limits_{\substack{1 \leqslant k \leqslant m \\ 1 \leqslant h \leqslant n}} a_{kh} f_{kh} = 0 \Rightarrow a_{kh} = 0$ para todo k, h.

Pero

$$\sum_{\substack{1 \leqslant k \leqslant m \\ 1 \leqslant h \leqslant n}} a_{kh} f_{kh} = 0 \Longleftrightarrow \sum_{\substack{1 \leqslant k \leqslant m \\ 1 \leqslant h \leqslant n}} a_{kh} f_{kh}(x) = 0 \quad \forall x \in E$$

En particular debe ocurrir que

$$\sum_{\substack{1 \leqslant k \leqslant m \\ 1 \leqslant h \leqslant n}} a_{kh} f_{kh}(a_j) = 0 \text{ para } 1 \leqslant j \leqslant n$$

$$\Rightarrow \sum_{k=1}^{m} a_{kj} b_k = 0 \Rightarrow a_{kj} = 0 \text{ para } 1 \leqslant k \leqslant m \text{ por ser } (b_i) \text{ una base.}$$

Como lo anterior podemos repetirlo para cada a_j, $1 \leqslant j \leqslant n$ y k recorre el mismo conjunto de índices que h, queda probado que $a_{kh} = 0$ para todo k, h.

2) Probemos ahora que $(f_{kh})_{\substack{1 \leqslant k \leqslant m \\ 1 \leqslant h \leqslant n}}$ es un conjunto generador de $L_K(E,F)$.

Sea $f \in L_K(E,F)$. Entonces

$$f(x) = f\left(\sum_{j=1}^{n} x_j a_j\right) = \sum_{j=1}^{n} x_j f(a_j).$$

Pero $f(a_j) \in F$, de aquí que

$$f(a_j) = \sum_{k=1}^{m} a_{kj} b_k \text{ y}$$

$$f(x) = \sum_{j=1}^{n} x_j \sum_{k=1}^{m} a_{kj} b_k = \sum_{\substack{1 \leqslant k \leqslant n \\ 1 \leqslant j \leqslant m}} a_{kj} x_j b_k.$$

Por otra parte,

$$f_{kj}(x) = f_{kj}\left(\sum_{j=1}^{n} x_j a_j\right) = \sum_{j=1}^{n} x_j f_{kj}(a_j) = x_j b_k,$$

de donde

$$\sum_{\substack{1\leq j\leq n\\1\leq k\leq m}} a_{kj}x_jb_k = \sum_{\substack{1\leq j\leq n\\1\leq k\leq m}} a_{kj}f_{kj}(x)$$

y $(f_{kj})_{\substack{1\leq k\leq m\\1\leq j\leq n}}$ es un conjunto generador en $L_K(E,F)$.

Queremos señalar que la búsqueda de la base de $L_K(E,F)$ es un poco más complicada que otros problemas de búsqueda de base que hemos resuelto en este texto; esto se debe a que los elementos de $L_K(E,F)$ son aplicaciones. El estudiante no necesita memorizar la demostración que acabamos de realizar; sin embargo, debe entenderla, ya que el trabajo con espacios cuyos elementos son aplicaciones es de gran importancia en estudios posteriores.

La búsqueda de la base que hemos realizado puede hacerse de forma análoga cuando E y F no son de dimensión finita, ya que nos basamos en el teorema de la caracterización de una aplicación lineal mediante las imágenes de una base del dominio, que como señalamos en una nota anterior, puede ser demostrado para espacios de dimensión no finita.

Una consecuencia inmediata de la demostración que acabamos de realizar es que $\dim L_K(E,F) = (\dim E) \cdot (\dim F)$ cuando $\dim E$ y $\dim F$ son finitas. Volveremos a hallar este resultado inmediatamente al estudiar qué ocurre con las matrices asociadas a aplicaciones lineales cuando se realizan operaciones entre estas, pero antes estudiaremos otra operación, que no puede definirse entre aplicaciones lineales en general, pero sí entre endomorfismos.

DEFINICIÓN 2.7

Sean $f, g \in \mathrm{End}(E)$. Entonces, se denomina *producto* de f por g a la aplicación $f\circ g$, dada por

$$f\circ g\ (x) = f(g(x))\ \forall x \in E.$$

Esta aplicación es un endomorfismo de E y es lineal pues

$$\begin{aligned} f\circ g\ (\alpha x+\beta y) &= f(g(\alpha x+\beta y)) = f[\alpha g(x)+\beta g(y)] \\ &= \alpha f(g(x)) + \beta f(g(y)) = \alpha f\circ g(x) + \beta f\circ g(y)\ \forall\, \alpha,\beta\in K,\ x,y\in E. \end{aligned}$$

Esta operación entre endomorfismos tiene las propiedades siguientes: $\forall f,g,h \in \mathrm{End}(E)$, $\forall\, \alpha \in K$ se cumple:

(1) $(f\circ g)\circ h = f\circ(g\circ h)$

(2) $(f+g)\circ h = f\circ h + g\circ h$

(3) $h\circ(f+g) = h\circ f + h\circ g$

(4) $\alpha(f\circ g) = (\alpha f)\circ g = f\circ(\alpha g)$

(5) $f\circ id_E = id_E\circ f = f\ \ \forall f \in \mathrm{End}\ (E)$

Demostraremos una de estas propiedades, y dejaremos las demás para que sean demostradas por el estudiante.

Demostración de (3):

$$[h\circ(f+g)](x)=[h(f+g)](x)=h[f(x)+g(x)]$$
$$=h(f(x))+h(g(x))$$
$$=h\circ f(x)+h\circ g(x)\ \forall x\in E$$

Por último, se cumple:

(6) Si f es inversible, entonces su inversa f^{-1} es lineal y se cumple que $f\circ f^{-1}=f^{-1}\circ f=\mathrm{id}_E$.
Probemos que f^{-1} es lineal.
Sean $z,w\in E$, $\alpha,\beta\in K$; entonces

$$f^{-1}(\alpha z+\beta w)=f^{-1}(\alpha f(x)+\beta f(y))=f^{-1}(f(\alpha x+\beta y))$$
$$=\alpha x+\beta y=\alpha f^{-1}(z)+\beta f^{-1}(w).$$

Hemos utilizado que f es sobreyectiva para escribir:
$z=f(x)$, $w=f(y)$ y, por tanto, $x=f^{-1}(z)$, $y=f^{-1}(w)$.

Una propiedad importante que, sin embargo, no cumple la composición de endomorfismos es la conmutatividad. En general, dados dos endomorfismos f y g de E no se cumple que $f\circ g=g\circ f$, lo cual no debe de extrañarnos pues lo mismo ocurre con la composición de aplicaciones.

A partir de la composición de endomorfismos podemos definir las potencias de un endomorfismo f de E como sigue:

Sea f un endomorfismo de E; entonces $f^n=\underbrace{f\circ f\circ\ldots\circ f}_{n \text{ veces}}$ y, por definición, $f^0=id_E$.

Esta definición de las potencias de un endomorfismo será de gran utilidad en lo adelante.

Analicemos ahora, en el caso de las aplicaciones lineales entre espacios de dimensión finita, cómo se comportan las matrices asociadas a dichas aplicaciones al realizar operaciones con estas.

Sean f, $g\in L_K(E,F)$, (a_j) una base de E, (b_i) una base de F y $A=M(f,(a_j),(b_i))$; $B=M(g,(a_j),(b_i))$.
¿Cuál será la matriz S de $f+g$ en las bases (a_j), (b_i)?

Sabemos que la columna j-ésima de S está dada por las coordenadas en base (b_i) de la imagen por $f+g$ de a_j, esto es,

$$(f+g)(a_j)=f(a_j)+g(a_j).$$

Si designamos $A=(\alpha_{ij})$ y $B=(\beta_{ij})$, tendremos:

$$(f+g)(a_j)=f(a_j)+g(a_j)=\sum_{i=1}^{m}\alpha_{ij}b_i+\sum_{i=1}^{m}\beta_{ij}b_i$$
$$=\sum_{i=1}^{m}(\alpha_{ij}+\beta_{ij})\,b_i.$$

Luego el elemento S_{ij} de S está dado por

$S_{ij}=\alpha_{ij}+\beta_{ij}$,

y podemos concluir que $S = A + B$. En otras palabras,

$$M(f+g, (a_j), (b_i)) = M(f, (a_j), (b_i)) + M(g, (a_j), (b_i)).$$

De manera similar, dada $A = M\ (f, (a_j), (b_i))$, busquemos $M(\alpha f, (a_j), (b_i))$, donde $\alpha \in K$.

$$(\alpha f)(a_j) = \alpha f(a_j) = \alpha \sum_{i=1}^{m} a_{ij} b_i = \sum_{i=1}^{m} (\alpha a_{ij}) b_i .$$

De aquí que si designamos por γ_{ij} al elemento ij de la matriz de f en bases (a_j), (b_i), tendremos:

$$\gamma_{ij} = \alpha a_{ij}$$
$$M(\alpha f,\ (a_j), (b_i)) = \alpha M(f,\ (a_j),\ (b_i)).$$

Ya hemos estudiado que si consideramos bases fijas en E y F, existe una correspondencia biyectiva entre las aplicaciones lineales de E en F y las matrices de orden $m \times n$, donde $n = \dim E$ y $m = \dim F$ (teorema 2.2).

Las propiedades que acabamos de estudiar sobre la correspondencia entre operaciones con aplicaciones lineales y las respectivas operaciones entre las matrices asociadas a dichas aplicaciones, llevan a considerar la correspondencia biyectiva entre $L_K(E,F)$ y $M_{m,n}(K)$ no solamente como aplicación, sino en su condición de isomorfismo: de ahí el resultado siguiente:

PROPOSICIÓN 2.3

Si se consideran en E y F las bases fijas (a_j) y (b_i) respectivamente, la aplicación φ: $L_K(E,F) \longrightarrow M_{m,n}(K)$ dada por $\varphi(f) = M(f, (a_j), (b_i))$ es un isomorfismo de espacios vectoriales.

Demostración

Hemos visto que la correspondencia dada por φ es biyectiva. Probemos entonces que

$$\varphi(f+g) = \varphi(f) + \varphi(g)\ \forall f, g \in L_K(E,F),$$
$$\varphi(\alpha f) = \alpha\varphi(f)\ \forall\ \alpha \in K.$$

En efecto,

$$\varphi(f+g) = M(f+g, (a_j), (b_i)) = M(f, (a_j), (b_i)) + M(g, (a_j), (b_i)) = \varphi(f) + \varphi(g),$$
$$\varphi(\alpha f) = M(\alpha f, (a_j), (b_i)) = \alpha M(f, (a_j), (b_i)) = \alpha\varphi(f).$$

De esta forma volvemos a hallar el resultado de que $\dim M_{m,n}(K) = m \cdot n$, el cual puede ser hallado directamente. Una forma de evitar la demostración directa de que $\dim L_K(E,F) = (\dim E)\ (\dim F)$, consiste en hallar directamente la dimensión de $M_{m,n}(K)$ y aplicar que $L_K(E,F)$ y $M_{m,n}(K)$ son isomorfos y, por tanto, tienen la misma dimensión, aunque la búsqueda de la base de $M_{m,n}(K)$ también tiene sus dificultades.

No obstante, queremos llamar la atención sobre el isomorfismo φ hallado entre $L_K(E,F)$ y $M_{m,n}(K)$. Este isomorfismo depende de las bases seleccionadas en E y F (de hecho hay un isomorfismo distinto para cada par de bases que escojamos); por ello, aunque es importante conocer la existencia de φ, debemos ser cuidadosos al utilizarlo, pues los resultados que se obtienen de su aplicación dependen no solamente de los espacios $L_K(E,F)$ y $M_{m,n}(K)$, sino también de las bases tomadas en E y F.

También queremos destacar, con el fin de reafirmar la idea de isomorfismo, cómo hemos encontrado para la suma y el producto por un escalar de aplicaciones lineales, propiedades que conocíamos de la suma y el producto por un escalar de matrices; esto se debe a que los espacios $L_K(E,F)$ y $M_{m,n}(K)$ son isomorfos y, por tanto, estructuralmente idénticos.

Por último, estudiaremos las matrices de los endomorfismos.

Está claro que para los endomorfismos son válidos los resultados encontrados para la suma y el producto por un escalar, es decir, si f y g son endomorfismos de un espacio vectorial E sobre K y (a_j) una base de E, entonces

$$M(f+g,(a_j))=M(f,(a_j))+M(g,(a_j))$$

y si $\alpha\in K$,

$$M(\alpha f,(a_j))=\alpha M(f,(a_j)).$$

Falta, pues, saber cuál es la matriz asociada al producto de f y g.

Sean $A=M(f,(a_j))$, $B=M(g,(a_j))$ y sea X el vector de coordenadas de $x\in E$ en (a_j). Entonces BX dará el vector de coordenadas de $g(x)$ en (a_j) y $A(BX)=(A\cdot B)X$ el vector de coordenadas de $f(g(x))$ en (a_j), es decir,

$$M(f\circ g,(a_j))=A\cdot B=M(f,(a_j))\cdot M(g,(a_j)),$$

ya que la matriz asociada a una aplicación lineal en bases fijas del espacio es única y, por tanto, la matriz de $f\circ g$ será la que haga corresponder al vector de coordenadas en la base (a_j) de $x\in E$, el vector de coordenadas de $f\circ g(x)$ en la base (a_j).

El estudiante habrá observado que en el caso de la matriz de un producto, utilizamos un método de demostración diferente al usado en el caso de la búsqueda de la matriz de la suma y el producto por un escalar. La diferencia de estos métodos depende del uso de caracterizaciones diferentes de la matriz asociada a una aplicación lineal; en el caso de la suma y el producto por un escalar, usamos la forma de los elementos de la matriz asociada a una aplicación lineal, y en el caso del producto buscamos cuál es la acción que realiza la matriz asociada a una aplicación lineal sobre el vector de coordenadas de un elemento del espacio.

Ambos métodos son válidos en todos los casos; sin embargo, hemos escogido vías distintas para que se observe el uso de las diferentes caracterizaciones y cómo en cada caso debe escogerse la más adecuada. En el caso del producto, usar la vía de buscar la forma de los elementos de la matriz es un poco más engorroso, pues hay que utilizar la forma de los elementos del producto de matrices.

Una consecuencia inmediata de la igualdad

$$M((f\circ g),\ (a_j))=M(f,(a_j))\cdot M(a_j)),$$

es la que se refiere a la matriz del endomorfismo inverso de un endomorfismo inversible.

Sea f un endomorfismo inversible de E, esto es, existe un endomorfismo f^{-1} de E que cumple:

$$f\circ f^{-1}=f^{-1}\circ f=id_E.$$

Entonces

$$M(f\circ f^{-1},\ (a_j))=M(f,(a_j))\cdot M(f^{-1},\ (a_j))=M(id_E,\ (a_j))=I_n,$$
$$M(f^{-1}\circ f,(a_j))=M(f^{-1},\ (a_j))\cdot M(f,(a_j))=M(id_E,\ (a_j))=I_n,$$

de donde resulta:

$$M(f^{-1},\ (a_j))=[M(f,(a_j))]^{-1}.$$

En resumen, en el conjunto $L_K(E)$ de los endomorfismos de un espacio E es posible realizar tres operaciones: La *suma* de endomorfismos, el *producto* de un escalar por un endomorfismo y la *composición* de endomorfismos, y por tanto, el cálculo de potencias de un endomorfismo.

Si analizamos el tipo de operaciones que implica el evaluar un polinomio, tiene sentido plantearnos la posibilidad de evaluar un polinomio en un endormorfismo.

Sea $P(x)=a_0+a_1x+\ldots+a_n\ x^n$, donde $p(x)\in K[x]$, y sea f un endomorfismo de un K-espacio E. Definimos

$$p(f)=a_0id_E+a_1f+\ldots+a_n\ f^n.$$

$p(f)$ es un endomorfismo de E y este tipo de expresión será de gran utilidad en un próximo capítulo.

De manera análoga, si A es una matriz de $M_n\ (K)$, definimos

$$p(A)=a_0I_n+a_1\ A+\ldots+a_n\ A^n$$

y $p(A)$ es una matriz de $M_n(K)$.

2.6 Núcleo e imagen de una aplicación lineal. Teorema del rango

Es conocida la importancia que tiene para el estudio de las aplicaciones en general, el análisis de su inyectividad y su sobreyectividad. Ahora bien, ¿cómo se manifiestan estas características en las aplicaciones lineales entre espacios vectoriales?

En este caso, al igual que en las aplicaciones en general, la sobreyectividad está dada por la relación entre el conjunto de las imágenes por la aplicación de los elementos del dominio y el codominio. En cuanto a la inyectividad, si suponemos una aplicación lineal $f: E \longrightarrow F$, para que sea inyectiva se debe cumplir que

$$f(x) = f(y) \Rightarrow x = y,$$

y de la linealidad resulta que esta conclusión se transforma en $f(x-y) = 0 \Rightarrow x-y=0$; de aquí que nos interese el conjunto de los elementos de E cuya imagen por f sea el 0 de F.

Las consideraciones anteriores motivan que estudiemos dos subconjuntos importantes relacionados con una aplicación lineal: el *núcleo* y la *imagen*.

DEFINICIÓN 2.8

Dada una aplicación lineal f de E en F, se denomina *imagen* de f y se denota por $\operatorname{Im} f$ o $f(E)$, al conjunto de las imágenes por f de los elementos de E, es decir,

$$\operatorname{Im} f = \{f(x) / x \in E\}.$$

Ejemplos

1. Para f_a: $E \longrightarrow E$ dada por $f_a(x) = ax$ con $a \neq 0$, $\operatorname{Im} f = \{ax/x \in E\}$.
2. La imagen de la aplicación nula, f_0, es evidentemente el vector nulo, mientras que $\operatorname{Im}\, id_E = E$.
3. En el espacio vectorial $\mathbb{R}^{\mathbb{N}}$ sobre $\mathbb{R}$, si consideramos $R(a_0, a_1, \ldots, a_n, \ldots) = (0, a_0, a_1, \ldots,),$
   entonces,
   $\operatorname{Im} R = \{(a_n)_{n \in \mathbb{N}} / a_0 = 0\},$
   o sea, el subconjunto de las sucesiones de números reales cuyo primer término es nulo.

DEFINICIÓN 2.9

Se denomina *núcleo* de una aplicación lineal f de E en F, al conjunto de los vectores de E cuya imagen es el vector nulo de F, y se denota por $\operatorname{Ker} f$, es decir,

$$\operatorname{Ker} f = \{x \in E / f(x) = 0\}.$$

En los ejemplos anteriores,

1. $\operatorname{Ker} f_a = \{x/ax = 0\} = \{0\}, \quad a \neq 0$
2. $\operatorname{Ker} f_0 = E$, $\operatorname{Ker}\, id_E = \{0\}$
3. $\operatorname{Ker} R = \{(a_n)/R[(a_n)] = (0,0,\ldots,0)\}$
   $= \{(a_0, 0, \ldots, 0) : a_0 \in \mathbb{R}\},$

   es decir, el conjunto de las sucesiones cuyo único término que puede ser no nulo es el primero.

En los ejemplos dados de imágenes y núcleos de aplicaciones lineales, observamos que en todos los casos estos son subespacios de F y de E, respectivamente. (Compruebe el estudiante esta afirmación en el ejemplo de la aplicación R de $\mathbb{R}^{\mathbb{N}}$.)

Este resultado no es casual, ya que en general el núcleo y la imagen de una aplicación lineal son subespacios vectoriales, como veremos a continuación.

PROPOSICIÓN 2.4

Sean E y F espacios vectoriales sobre K y sea f una aplicación lineal entre E y F. Entonces $\mathrm{Im}\, f$ es un subespacio de F y $\mathrm{Ker}\, f$ es un subespacio de E.

Demostración

Aunque este resultado es consecuencia directa de la proposición 2.2, vamos a hacer la demostración.

Probemos que $\mathrm{Im}\, f$ es un subespacio de F.

Sean $f(x)$, $f(x')$ dos elementos cualesquiera de $\mathrm{Im}\, f$ y $\alpha, \beta \in K$. Entonces $\alpha f(x) + \beta f(x') = f(\alpha x + \beta x') \in \mathrm{Im}\, f$. Así queda probado que $\mathrm{Im}\, f$ es un subespacio de F.

Probemos ahora que $\mathrm{Ker}\, f$ es un subespacio de E.

Sean $x, x' \in \mathrm{Ker}\, f$, $\alpha, \beta \in K$. Debemos probar que $\alpha x + \beta x' \in \mathrm{Ker}\, f$.

$$\alpha x + \beta x' \in \mathrm{Ker}\, f \Longleftrightarrow f(\alpha x + \beta x') = 0,$$

pero $f(\alpha x + \beta x') = \alpha f(x) + \beta f(x') = 0$ ya que $x, x' \in \mathrm{Ker}\, f$ implica que $f(x) = f(x') = 0$. Con esto queda demostrado que $\mathrm{Ker}\, f$ es un subespacio de E.

El hecho de que el núcleo y la imagen de una aplicación lineal sean subespacios, motiva que sea de interés la búsqueda de generadores y, con mayor razón, de bases de dichos subespacios.

En el caso del subespacio imagen, podemos obtener una propiedad general sobre la búsqueda de un sistema generador, que depende directamente del hecho de que una aplicación lineal de E en F transforma combinaciones lineales en combinaciones lineales.

PROPOSICIÓN 2.5

Sean f una aplicación lineal de E en F y $(x_1, x_2, \ldots, x_n)$ un sistema generador de E. Entonces $(f(x_1), f(x_2), \ldots, f(x_n))$ es un sistema generador de $\mathrm{Im}\, f$.

Demostración

Debemos probar que todo y de $\mathrm{Im}\, f$ se expresa como combinación lineal de $(f(x_1), f(x_2), \ldots f(x_n))$.

Sea $y \in \mathrm{Im}\, f$. Entonces $y = f(x)$ para algún $x \in E$ y

$$x = \sum_{i=1}^{n} a_i\, x_i$$

ya que $(x_1,x_2,\dots,x_n)$ es un sistema generador de E. Luego

$$y=f(x)=f\left(\sum_{i=1}^{n} a_i\, x_i\right)=\sum_{i=1}^{n} a_i\, f(x_i).$$

Con esto queda probada la proposición.

No es difícil ver que si la aplicación f es sobreyectiva, la propiedad anterior permite afirmar que *si* $(x_1,x_2,\dots,x_n)$ *es un sistema generador de* E, entonces $(f(x_1), f(x_2),\dots, f(x_n))$ *es un sistema generador de* F.

¿Será esta propiedad una caracterización de las aplicaciones sobreyectivas?, es decir, ¿bastará comprobar que el conjunto de las imágenes de cualquier sistema generador de E es un sistema generador de F, para poder afirmar que f es sobreyectiva? Veamos.

PROPOSICIÓN 2.6

Sean E y F espacios vectoriales sobre K. Si f es una aplicación lineal de E en F y si dado un sistema generador $(x_1,x_2,\dots,x_n)$ de E se tiene que $(f(x_1), f(x_2),\dots f(x_n))$ es un sistema generador de F, entonces f es sobreyectiva.

Demostración

Sea $y\in F$. Debemos probar que existe $x\in E$ tal que $y=f(x)$.

Podemos escribir y como

$$y=\sum_{i=1}^{n} a_i\, f(x_i),$$

ya que $(f(x_1),\dots f(x_n))$ es generador de F.

De aquí resulta que $y=f\left(\sum_{i=1}^{n} a_i\, x_i\right)$, y podemos tomar $x=\sum_{i=1}^{n} a_i x_i$.

Con esto queda probado que f es sobreyectiva.

Podemos decir, entonces, que para que una aplicación lineal f de E en F sea sobreyectiva, es necesario y suficiente que las imágenes por f de un sistema generador de E formen un sistema generador de F.

Nota: El estudiante observará que desde el inicio de este epígrafe no hemos hecho alusión a la dimensión de los espacios. Sin embargo, en las proposiciones anteriores trabajamos con un sistema generador finito. Esto tiene como objetivo facilitar la comprensión de las propiedades y su demostración, aunque las proposiciones anteriores son válidas también para un sistema generador infinito. Dejamos al estudiante que enuncie y demuestre las proposiciones correspondientes a dicho caso, para lo cual basta seguir los pasos dados en el caso tratado teniendo cuidado de trabajar con las combinaciones lineales correspondientes a un sistema generador infinito.

Ejemplos

1. Sea $f\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^2$ dada por

$$f(x,y,z)=(x-y+2z,\ 2x+y,\ -x\ -2y+2z).$$

Si queremos hallar un sistema generador de Im f basta hallar las imágenes de un sistema generador de $\mathbb{R}^3$. En este caso tomaremos el más simple, esto es, $e_1=(1,0,0)$, $e_2=(0,1,0)$, $e_3=(0,0,1)$. Entonces

$f(e_1)=(1,2,-1)$, $f(e_2)=(-1,1,-2)$, $f(e_3)=(2,0,2)$.

De aquí que $((1,2,-1), (-1,1,-2), (2,0,2))$ genere Im f. Este conjunto, sin embargo, no es una base de Im f, ya que no es un sistema de vectores linealmente independiente.

2. Sea $f: \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ dada por $f(x,y)=(x+y, y)$.
Un sistema generador de Im f está dado por

$f(1,0)=(1,0)$, $f(0,1)=(1,1)$.

Esto es, $((1,0), (1,1))$ es un sistema generador de Im f. Este sistema es realmente una base ya que $(1, 0)$ y $(1, 1)$ son linealmente independientes.

3. Consideremos en $K[x]$ la aplicación $D: K[x] \longrightarrow K[x]$ dada por $D(p(x))=p'(x)$, que, como sabemos, es lineal. Un sistema generador de $K[x]$ es

$(1,x,x^2,\ldots,x^n,\ldots)$.

Luego un sistema generador de Im D será

$(0,1,2x,\ldots,nx^{n-1},\ldots)$.

de hecho podemos tomar $(0,1,x,x^2,\ldots,x^n,\ldots)$, lo cual también hubiéramos podido inferir del hecho de que D es sobreyectiva. El sistema generador obtenido no es una base, ya que contiene al vector nulo.

4. En $K[x]$ la aplicación $T: K[x] \longrightarrow K[x]$ dada por

$T(a_0+a_1x+\ldots+a_n x^n)=a_0x+a_1 x^2+\ldots+a_nx^{n+1}$

es una aplicación lineal, lo que puede comprobarse fácilmente.
Si tomamos como sistema generador del conjunto de partida $K[x]$, la base usual de los polinomios, $(1, x,\ldots,x^n,\ldots)$, un sistema generador de $T(K[x])$ será

$(x,x^2,\ldots,x^{n+1},\ldots)$,

que en este caso es una base ya que es linealmente independiente.

En los ejemplos anteriores podemos observar que no siempre el sistema generador hallado es linealmente independiente y, por tanto, una base del subespacio imagen de la aplicación lineal. Si en dichos ejemplos tomamos en cuenta que hemos partido de bases, esto es, de sistemas de generadores linealmente independientes, es lógico investigar por qué dichos sistemas no se transforman, a su vez, en sistemas linealmente independientes. Si observamos las características de las aplicaciones estudiadas, vemos que en los casos en que hemos obtenido sistemas linealmente independientes, se trata de aplicaciones inyectivas y, en caso contrario, de aplicaciones no inyectivas, de lo cual intuimos una posible relación entre la inyectividad de una aplicación y el hecho de que transforme sistemas linealmente independientes del conjunto de partida en sistemas linealmente independientes de la imagen y, por tanto, del conjunto de llegada. Estudiemos pues la relación entre la inyectividad y la independencia lineal.

Comencemos por un caso muy simple, que se presentó en el ejemplo 3.

En este ejemplo es más evidente el hecho de que D es no inyectiva, que el hecho de que la imagen de un vector no nulo de $K[x]$, el polinomio constante 1, es el polinomio nulo, lo que origina que el sistema generador obtenido no sea linealmente independiente.

Ya habíamos señalado que existe relación entre la inyectividad de una aplicación lineal y el hecho de que vectores no nulos tengan imagen nula, esto es, que en el núcleo de dicha aplicación lineal existan vectores no nulos.

PROPOSICIÓN 2.7

Una aplicación lineal f entre los K-espacios E y F es inyectiva si y solo si $\operatorname{Ker} f=\{0\}$.

Demostración

Supongamos que f es inyectiva y probemos que entonces $\operatorname{Ker} f=\{0\}$.

Si existe $x\neq 0$ tal que $f(x)=0$, tenemos $x\neq 0 \Rightarrow f(x)=f(0)$, lo que contradice que f sea inyectiva.

Recíprocamente, sea $\operatorname{Ker} f=\{0\}$ y probemos que entonces f es inyectiva.

Si f no es inyectiva, existen x_1, $x_2 \in E$, $x_1\neq x_2$, tales que $f(x_1)=f(x_2)$, es decir, que existen $x_1-x_2\neq 0$ tales que $f(x_1-x_2)=f(x_1)-f(x_2)=0$ y esto contradice que el núcleo de f esté formado solamente por el vector nulo.

Esta proposición ofrece un criterio muy importante y simple para la inyectividad de una aplicación, y reafirma una vez más las características de los problemas lineales y la importancia del vector nulo en estos, pues reduce el análisis de la inyectividad de una aplicación al análisis de los vectores que se transforman en el vector nulo.

Volvamos ahora a la relación entre la inyectividad de una aplicación lineal y el hecho de si dicha aplicación transforma sistemas de vectores linealmente independientes del conjunto de partida en sistemas de vectores linealmente independientes de la imagen.

De los ejemplos intuimos que es suficiente que una aplicación sea inyectiva, para que transforme sistemas de vectores linealmente independientes en sistemas de vectores linealmente independientes; veremos que esta condición es también necesaria.

PROPOSICIÓN 2.8

Para que una aplicación lineal de E en F sea inyectiva, es necesario y suficiente que el sistema formado por las imágenes de todo sistema de vectores linealmente independientes de E, sea un sistema de vectores linealmente independiente de F.

Demostración

Sea f una aplicación lineal que transforma todo sistema linealmente independiente de E en un sistema linealmente independiente de F y probemos que entonces f es inyectiva.

Supongamos que f no es inyectiva. En este caso, según la proposición, existe $x \in E, x \neq 0$, tal que $f(x) = 0$. Esto contradice el hecho de que las imágenes de un sistema linealmente independiente de E deben formar un sistema linealmente independiente de F, pues (x) es un sistema linealmente independiente de E y $(f(x)) = (0)$ no es linealmente independiente en F.

Supongamos que f es inyectiva. Tomemos un sistema $(x_1, x_2, \ldots, x_n)$ linealmente independiente de E y probemos que $(f(x_1), f(x_2), \ldots, f(x_n))$ es linealmente independiente en F.

Sea $(f(x_1), f(x_2), \ldots, f(x_n))$ un sistema no linealmente independiente: entonces existen $a_1, a_2, \ldots, a_n$, no todos nulos, tales que

$$\sum_{i=1}^{n} a_i f(x_i) = 0.$$

Pero

$$\sum_{i=1}^{n} a_i f(x_i) = f\left(\sum_{i=1}^{n} a_i \, (x_i) \right) = 0,$$

donde $\sum_{i=1}^{n} a_i \, x_i$ debe ser no nulo, pues como los a_i, $i = 1, \ldots, n$, no todos son nulos, si $\sum_{i=1}^{n} a_i \, x_i$ fuese nulo, esto estaría en contradicción con el hecho de que $(x_1, x_2, \ldots, x_n)$ es linealmente independiente.

Hemos, pues, encontrado $x = \sum_{i=1}^{n} a_i \, x_i \neq 0$ tal que $f(x) = 0$, lo que contradice la inyectividad de f, con lo cual queda probada la proposición.

Nota. En la demostración de que si f es inyectiva, transforma sistemas de vectores linealmente independientes en sistemas de vectores linealmente independientes, nos hemos limitado una vez más a sistemas finitos de vectores, con el fin de simplificar. Esta propiedad, sin embargo, es válida también para el caso de sistemas infinitos. El estudiante puede realizar la demostración para este caso y comprobar que la diferencia está dada solamente por la consideración de las combinaciones lineales de un número finito de elementos.

Vamos a resumir las proposiciones 2.6 y 2.8.

Sean E y F dos espacios vectoriales sobre K, f una aplicación lineal de E en F:

Para que f sea *sobreyectiva* es necesario y suficiente que f transforme todo sistema generador de E en un sistema generador de F.

Para que f sea *inyectiva* es necesario y suficiente que f transforme todo sistema linealmente independiente de E en un sistema linealmente independiente de F.

Luego podemos concluir:

Para que f sea *biyectiva*, es decir, un *isomorfismo*, es necesario y suficiente que f transforme toda base de E en una base de F.

Tomemos de nuevo los ejemplos 1 y 2 del presente epígrafe.

En el ejemplo 1 encontramos que $((1,2,-1), (-1,1,-2), (2,0,2))$ es un sistema generador de $\operatorname{Im} f$, y una base de $\operatorname{Im} f$ será, por ejemplo, $(1,2,-1)$, $(2,0,2))$, es decir, $\dim \operatorname{Im} f=2$.

Hallemos $\operatorname{Ker} f=\{(x,y,z) : f(x,y,z)=(0,0)\}$.

$\operatorname{Ker} f$ está dado por el espacio solución de

$$\begin{cases} x_1-x_2+2x_3=0 \\ 2x_1+x_2=0 \\ -x_1-2x_2+2x_3=0 \end{cases}$$

Resolviendo el sistema, que dejamos al estudiante como ejercicio, obtenemos:

$\operatorname{Ker} f=\langle\langle(-2,4,1)\rangle\rangle$, esto es, $\dim \operatorname{Ker} f=1$.

Obsérvese que $\dim \operatorname{Ker} f+\dim \operatorname{Im} f=3=\dim \mathbb{R}^3$ y $\mathbb{R}^3$ es el dominio de f.

En el ejemplo 2 $\operatorname{Im} f=\mathbb{R}^2$, de aquí que $\dim \operatorname{Im} f=2$, mientras que $\operatorname{Ker} f$ está dado por la solución de

$$\begin{cases} x+y=0 \\ \quad\; y=0 \end{cases}$$

es decir, $\operatorname{Ker} f=\{0\}$.

De nuevo obtenemos que $\dim \operatorname{Ker} f+\dim \operatorname{Im} f=2=\dim \mathbb{R}^2$, y $\mathbf{R}^2$ es el dominio de f.

En estos ejemplos observamos cierta regularidad: una relación entre las dimensiones del núcleo y la imagen de una aplicación lineal entre espacios vectoriales de dimensión finita, y la dimensión del dominio de dichas aplicaciones.

El núcleo y la imagen son subespacios que determinan propiedades importantes de una aplicación lineal, y en los ejemplos dados, partiendo de una aplicación lineal conocida, hemos encontrado que su núcleo y su imagen satisfacen ciertas relaciones. Podemos preguntarnos si es posible invertir el proceso; esto es, prefijados dos subespacios, uno que sería el núcleo y el otro la imagen, ¿existirá una aplicación lineal cuyo núcleo e imagen sean los dados? Estudiemos esta posibilidad mediante los ejemplos siguientes.

Ejemplos

1. ¿Existirá una aplicación lineal de $\mathbf{R}^3$ en $\mathbf{R}^4$ tal que $\operatorname{Im} f=V=\langle\langle(1,2,0,-4), (2,0,-1,-3)\rangle\rangle$ y $\operatorname{Ker} f=W=\langle\langle(1,1,0), (0,1,1)\rangle\rangle$? Recordemos que para tener unívocamente determinada una aplicación lineal f,

es necesario y suficiente conocer las imágenes por f de los elementos de una base del dominio, en este caso es $\mathbb{R}^3$. Como tenemos, además, la condición que determina el núcleo de f, tomemos la base en $\mathbb{R}^3$ formada a partir de completar una base del núcleo prefijado; sea esta:

$$a_1=(1,1,0),\ a_2=(0,1,1),\ a_3=(0,0,1).$$

Hagamos

$$f(1,1,0)=(0,0,0,0),$$
$$f(0,1,1)=(0,0,0,0).$$

Falta aún definir la imagen del vector $(0,0,1)$ y, por supuesto, tener en cuenta la condición de que la imagen de f sea W, es decir, que toda imagen por f de un elemento de $\mathbb{R}^3$ debe ser una combinación lineal de $(1,2,0,-4)$ y $(2,0,-1,-3)$. Es evidente que esto no lo podemos lograr, pues queda solo una posibilidad para asignarle valores a los vectores de la base, y como queremos que esta imagen esté en el subespacio prefijado, tendrá que ser necesariamente uno de los vectores de ese espacio; sea este:

$$v_0=\alpha_0\,(1,2,0,-4)+\beta_0\,(2,0,-1,-3)$$

con $\alpha_0,\beta_0\in K$ fijos. Entonces

$$f(0,0,1)=v_0.$$

De esta manera, por el teorema de caracterización, hemos definido una aplicación lineal. Busquemos $\operatorname{Ker} f$ e $\operatorname{Im} f$.
$\operatorname{Ker} f$ es, evidentemente, $\langle\langle(1,1,0),(0,1,1)\rangle\rangle$. Calculemos $\operatorname{Im} f$. Sea $x\in\mathbb{R}^3$ y expresémoslo como una combinación de la base $((1,1,0),\ (0,1,1),\ (0,0,1))$:

$$x=\alpha(1,1,0)+\beta(0,1,1)+\gamma(0,0,1).$$

Calculando $f(x)$ tenemos:

$$f(x)\ =\alpha f((1,1,0)+\beta f((0,1,1))+\gamma f((0,0,1))=\gamma v_0$$

Luego,

$$\operatorname{Im} f=\langle v_0\rangle\neq\langle\langle(1,2,0,-4),\ (2,0,-1,-3)\rangle\rangle.$$

Por tanto, no podemos construir ninguna aplicación lineal con ese núcleo y esa imagen, pues si el núcleo es el prefijado, la imagen siempre tendrá dimensión 1 y el subespacio dado tiene dimensión 2. Quedaría otra posibilidad: prefijar la imagen, pero entonces el núcleo nunca sería el subespacio dado.

2. ¿Existirá una aplicación lineal de $\mathbb{R}^3$ en $\mathbb{R}^3$ tal que $\operatorname{Ker} f=$ $=V=\langle\langle(3,-1,1)\rangle\rangle$ e $\operatorname{Im} f=W=\langle\langle(1,0,1),\ (0,1,-1)\rangle\rangle$?
Sigamos el mismo procedimiento anterior. Tomemos en $\mathbb{R}^3$ la base

$$a_1=(3,-1,1),\ a_2=(0,1,0),\ a_3=(0,0,1)$$

y definamos

$$f(3,-1,1)=(0,0,0),$$
$$f(0,1,0)=(1,0,1),$$
$$f(0,0,1)=(0,1,-1).$$

Esta aplicación está bien determinada y podemos incluso hallar su expresión explícita. Para ello expresemos $(x,y,z)\in\mathbb{R}^3$ como combinación lineal de (a_1,a_2,a_3).

$$(x,y,z)=\alpha(3,-1,1)+\beta(0,1,0)+\gamma(0,0,1).$$

De aquí,

$$\begin{cases} 3\alpha=x \\ \alpha+\beta=y \\ \alpha+\gamma=z \end{cases}$$

y $\quad \alpha=\dfrac{x}{3},\ \beta=\dfrac{x}{3}+y,\ \gamma=z-\dfrac{x}{3}.$

Por tanto,

$$(x,y,z)=\frac{x}{3}(3,-1,1)+(\frac{x}{3}+y)(0,1,0)+\left(z-\frac{x}{3}\right)(0,0,1)$$

$$f(x,y,z)=f\left[\frac{x}{3}(3,-1,1)+\left(\frac{x}{3}+y\right)(0,1,0)+\left(z-\frac{x}{3}\right)(0,0,1)\right]$$

$$=\frac{x}{3}f(3,-1,1)+(x+y)f(0,1,0)+\left(z-\frac{x}{3}\right)f(0,0,1)$$

$$=\left(\frac{x}{3}+y\right)(1,0,1,)+\left(z-\frac{x}{3}\right)(0,1,-1)$$

$$=\left(\frac{x}{3}+y,0,\ \frac{x}{3}+y\right)+\left(0,\ z-\frac{x}{3},\ -z+\frac{x}{3}\right)$$

$$=\left(\frac{x}{3}+y,z-\frac{x}{3},\frac{2}{3}x+y-z\right).$$

En el ejemplo 1, donde el problema planteado no tiene solución, no se cumple la relación que hemos encontrado entre las dimensiones del núcleo, la imagen y el dominio de la aplicación; mientras que en el 2 sí se cumple dicha relación. Veremos que, efectivamente, esta relación se cumple en general para aplicaciones lineales entre espacios vectoriales de dimensión finita, y es por ello que el problema planteado en los ejemplos 1 y 2 no siempre tiene solución. Pasemos a enunciar dicha propiedad general.

TEOREMA 2.5

Sean E y F dos espacios vectoriales de dimensión finita sobre K, f una aplicación lineal de E en F. Entonces

$$\dim \operatorname{Ker} f + \dim \operatorname{Im} f = \dim E.$$

Demostración

La demostración consiste en fundamentar rigurosamente un hecho que es posible observar en los ejemplos numéricos resueltos, esto es, que la cantidad máxima de vectores linealmente independientes de la imagen, está dada por las imágenes de los vectores de la base de E que no pertenecen al núcleo de f.

Sea $\dim \operatorname{Ker} f = r$ $\dim E = n$ y tomemos una base $(a_i)_{1 \leq i \leq n}$ de E de forma tal, que $(a_1, a_2, \ldots, a_r)$ sea una base de $\operatorname{Ker} f$.

Probemos que $f(a_{r+1}), f(a_{r+2}), \ldots, f(a_n)$ constituyen una base de $\operatorname{Im} f$ probando que forman un sistema generador de $\operatorname{Im} f$ y que son linealmente independientes.

Veamos primero que $(f(a_{r+1}), \ldots, f(a_n))$ genera a $\operatorname{Im} f$.

Sea $y \in \operatorname{Im} f$; entonces $y = f(x)$ para algún $x \in E$. De aquí,

$$y = f\left(\sum_{i=1}^{n} x_i a_i\right) = \sum_{i=1}^{n} x_i f(a_i) = \sum_{i=r+1}^{n} x_i f(a_i),$$

pues $f(a_1) = f(a_2) = \ldots = f(a_r) = 0$ ya que los a_i, $i = 1, 2, \ldots, r$, pertenecen a $\operatorname{Ker} f$.

Hemos probado que $(f(a_{r+1}), \ldots, f(a_n))$ genera a $\operatorname{Im} f$. Probemos ahora que $f(a_{r+1}), \ldots, f(a_n)$ son linealmente independientes.

Sea $\sum_{i=r+1}^{n} \alpha_i f(a_i) = 0$. Entonces,

$$f\left(\sum_{i=r+1}^{n} \alpha_i a_i\right) = 0 \Longleftrightarrow \sum_{i=r+1}^{n} \alpha_i a_i \in \operatorname{Ker} f.$$

Por tanto,

$$\sum_{i=r+1}^{n} \alpha_i a_i = \sum_{i=1}^{r} \beta_i a_i,$$

de donde obtenemos:

$$\sum_{i=1}^{r} \beta_i a_i - \sum_{i=r+1}^{n} \alpha_i a_i = 0.$$

Luego $\beta_1 = \beta_2 = \ldots = \beta_r = \alpha_{i+1} = \ldots = \alpha_i = 0$ ya que $(a_1, a_2, \ldots, a_n)$ es una base de E, y del hecho que $\alpha_{i+1} = \ldots = \alpha_i = 0$ queda demostrado que $f(a_{r+1}), \ldots, f(a_n)$ son linealmente independientes y, por tanto, forman una base de $\operatorname{Im} f$.

DEFINICIÓN 2.10

Se denomina *rango* de una aplicación lineal f a la dimensión del subespacio imagen. El rango de f, se denota por rg f; luego,

rg $f=$dim Im f.

Nota. Realmente, en esta definición sobreentendemos que la dimensión de Im f es finita. En particular, podemos asegurar que dim Im f es finita siempre que dim E sea finita, ya que como el conjunto de las imágenes por una aplicación lineal de E en F de un sistema generador de E es un sistema generador de Im f, se cumple que dim Im $f \leqslant$ dim E.

El teorema 2.5 es conocido con el nombre de *teorema del rango.*

Del teorema del rango resulta que rg $f=$ dim $E-$ dim Ker f.

La dimensión del núcleo de f suele llamarse *nulidad* de f y se denota por Nul f; así, Nul $f=$dim Ker f.

En la demostración del teorema del rango observamos que no ha intervenido para nada el hecho de que la dimensión de F sea finita, ya que lo que interviene es la dimensión del subespacio imagen, que puede ser finita aunque la dimensión del conjunto de llegada no lo sea. Por ejemplo, en $K[x]$ el subespacio de los polinomios de grado menor que n es de dimensión finita, mientras que la dimensión de $K[x]$ no es finita. Es por ello que el enunciado del teorema del rango puede darse en forma más general, asumiendo que solamente el espacio de partida es de dimensión finita.

La demostración del teorema del rango se puede expresar en otro lenguaje que nos permite llegar a útiles conclusiones.

Utilizando la notación de la demostración, si consideramos el subespacio de E, $W=\langle a_{r+1},\ldots, a_n\rangle$, tendremos que $E=\text{Ker} f\oplus W$, y así hemos encontrado que dim $W=$dim Im f, esto es, que la imagen de f es isomorfa a un suplementario cualquiera de Ker f, pues la selección de los $a_{r+1},\ldots, a_n$, y por tanto, del suplementario de Ker f, se hizo de forma arbitraria.

Nota. Cuando el espacio E no es de dimensión finita, no tiene sentido el enunciado del teorema del rango. Sin embargo, puede ocurrir que una aplicación lineal de E en F sea tal, que aunque las dimensiones de E y F no sean finitas, la dimensión de la imagen de f sí lo sea.

Ejemplo.

Consideremos la aplicación T_n entre $C^\infty(\mathbb{R})=\{$funciones de $\mathbb{R}$ en $\mathbb{R}$ infinitamente derivables$\}$ y $\mathbb{R}[x]$ dada por $T_n(f)=p_n(f)$, donde $p_n(f)$ es el polinomio de Taylor de orden n de f. Esta aplicación es lineal y evidentemente su imagen es el espacio de los polinomios de grado menor o igual que n, cuya dimensión es finita.

En este caso está claro que es válida la definición de rango de f como la dimensión del subespacio Im f. En cuanto a la propiedad de que el subespacio Im f es isomorfo a cualquier suplementario de Ker f, es también válida en este caso, aunque para su demostración es necesario usar otros recursos que están fuera del alcance de este curso.

Aplicaciones del teorema del rango

El teorema del rango se utiliza en la demostración de numerosas propiedades. A continuación daremos algunos ejemplos.

PROPOSICIÓN 2.9

Sean E y F dos espacios vectoriales sobre K de dimensión finita, tales que $\dim E = \dim F$, f una aplicación lineal de E en F. Entonces

$$f \text{ inyectiva } \Longleftrightarrow f \text{ sobreyectiva } \Longleftrightarrow f \text{ biyectiva.}$$

Demostración

Probemos que: f inyectiva $\Longleftrightarrow$ f sobreyectiva.
f inyectiva $\Longleftrightarrow$ $\operatorname{Ker} f = \{0\}$ $\Longleftrightarrow$ $\dim \operatorname{Ker} f = 0$
Del teorema del rango resulta:

$$\dim E = \dim \operatorname{Im} f$$

y como $\dim E = \dim F$,

$$\dim \operatorname{Im} f = \dim F.$$

Por ser $\operatorname{Im} f$ un subespacio de F, tenemos

$\operatorname{Im} f = F$, es decir, f es sobreyectiva.

Recíprocamente, probemos que si f es sobreyectiva, entonces f es inyectiva.

f sobreyectiva $\Longleftrightarrow$ $\operatorname{Im} f = F$. Entonces,

$\dim \operatorname{Im} f = \dim F = \dim E$.
Del teorema del rango resulta:

$$\dim \operatorname{Ker} f = 0 \Longleftrightarrow \operatorname{Ker} f = \{0\} \Longleftrightarrow f \text{ es inyectiva.}$$

Está claro que una vez probado que f inyectiva $\Longleftrightarrow f$ sobreyectiva, toda aplicación lineal inyectiva o sobreyectiva será biyectiva, con lo que queda probada la proposición.

Obsérvese que la hipótesis $\dim E = \dim F$ es fundamental en la demostración de la proposición 2.9. El estudiante puede buscar en los ejemplos dados en este capítulo, aplicaciones lineales entre espacios vectoriales de distinta dimensión, inyectivas y no sobreyectivas, o viceversa.

Un caso particular importante de la proposición anterior es el de los endomorfismos, que solamente enunciaremos pues está incluido en la demostración de dicha proposición.

PROPOSICIÓN 2.10

Sea f un endomorfismo de un espacio vectorial de dimensión n sobre K. Entonces f inyectiva $\Longleftrightarrow f$ sobreyectiva $\Longleftrightarrow f$ es un automorfismo.

Nota. La proposición 2.10 no es válida para endomorfismos de espacios de dimensión infinita, como se muestra en los ejemplos siguientes:

Ejemplos

1. $I: K[x] \to K[x]$ dado por $I(a_0 + a_1x + \ldots + a_nx^n) = a_1 + a_2x + \ldots + a_nx^{n-1}$, es un endomorfismo sobreyectivo pero no inyectivo, ya que su núcleo está formado por los polinomios constantes.

2. $D\colon K[x] \to K[x]$ dado por
$D(a_0 + a_1x + \dots + a_nx^n) = a_0x + a_1x^2 + \dots + a_nx^{n+1}$ es inyectiva pero no sobreyectiva, pues los polinomios constantes no pertenecen a la imagen.

Otro ejemplo de una propiedad en cuya demostración se utiliza el teorema del rango es el siguiente:

Ejemplo

Demostremos la propiedad:

Sean E un espacio vectorial de dimensión n sobre K y f un endomorfismo de E. Entonces, $\operatorname{Ker} f = \operatorname{Im} f$ si y solo si $f^2=0$, $f \neq 0$, n es par y $\operatorname{rg} f = \dfrac{n}{2}$.

Probemos primero que si $\operatorname{Ker} f = \operatorname{Im} f$ se cumplen las condiciones dadas.

$\operatorname{Ker} f = \operatorname{Im} f \Rightarrow \dim \operatorname{Ker} f = \dim \operatorname{Im} f$.

Del teorema del rango resulta:

$$n = \dim E = \dim \operatorname{Ker} f + \dim \operatorname{Im} f = 2 \dim \operatorname{Ker} f,$$

de donde

$$\dim \operatorname{Ker} f = \dim \operatorname{Im} f = \frac{n}{2},$$

y por definición de rango de f,

$$\operatorname{rg} f = \dim \operatorname{Im} f = \frac{n}{2}.$$

Además, $f \neq 0$ pues $\dim \operatorname{Im} f = \dfrac{n}{2}$ y la única aplicación cuya imagen es el espacio $\{0\}$ es la nula.
Por último, como $\operatorname{Ker} f = \operatorname{Im} f$, resulta:

$$f(x) \in \operatorname{Ker} f \quad \forall x \in E.$$

Luego, $f(f(x)) = f^2(x) = 0 \; \forall x \in E$ y, por tanto, $f^2 = 0$.

Recíprocamente, probemos que si $f^2=0$, $f \neq 0$, n es par y $\operatorname{rg} f = \dfrac{n}{2}$, entonces $\operatorname{Ker} f = \operatorname{Im} f$.

De nuevo, por el teorema del rango, como

$$\dim E = \dim \operatorname{Ker} f + \dim \operatorname{Im} f$$

y $\dim E = n$, $\dim \operatorname{Im} f = \operatorname{rg} f = \dfrac{n}{2}$, resulta que $\dim \operatorname{Ker} f = \dfrac{n}{2}$ y, por tanto, $\dim \operatorname{Ker} f = \dim \operatorname{Im} f$.

Si probamos que $\operatorname{Ker} f \subset \operatorname{Im} f$ o $\operatorname{Im} f \subset \operatorname{Ker} f$, al ser iguales las dimensiones de ambos subespacios quedará probada la propiedad.

Probemos que $\operatorname{Im} f \subset \operatorname{Ker} f$, lo cual resulta más fácil dadas las hipótesis que poseemos y las definiciones de núcleo e imagen.

Sea $f(x) \in \mathrm{Im}\, f$. Entonces

$$f(x) \in \mathrm{Ker}\, f \Longleftrightarrow f(f(x)) = 0 \text{ y } f(f(x)) = f^2(x) = 0,$$

luego $f(x) \in \mathrm{Ker}\, f$.

Resulta, entonces, $\mathrm{Im}\, f \subset \mathrm{Ker}\, f$, y de aquí $\mathrm{Im}\, f = \mathrm{Ker}\, f$, que era lo que queríamos probar.

Existen muchas otras aplicaciones del teorema del rango. Algunas de ellas aparecerán enunciadas en los ejercicios.

Rango de aplicaciones lineales y rango de matrices

Hemos definido el rango de una aplicación lineal entre espacios vectoriales E y F como la dimensión del subespacio de f, cuando esta dimensión es finita. No es difícil pensar que en el caso en que las dimensiones de E y F sean finitas, en el cual podemos representar f matricialmente, la representación matricial nos proporcione un método de cálculo del rango de la aplicación lineal. (Recordemos que el rango de una matriz se define como el número máximo de vectores columna (o vectores fila) linealmente independientes, esto es, la dimensión del espacio generado por las columnas de la matriz.)

Por otra parte, hemos estudiado que un generador de la imagen de f se obtiene mediante el conjunto de las imágenes por f de un generador cualquiera de E, en particular, de una base $(a_i)_{1 \leq i \leq n}$ de E.

Si consideramos $\mathrm{Im}\, f = \langle f(a_1), f(a_2), \ldots, f(a_n) \rangle$, está claro que $\dim \mathrm{Im}\, f = \dim \langle f(a_1, f(a_2), \ldots, f(a_n) \rangle$ y en la práctica esta dimensión se calcula mediante el rango de la matriz cuyas columnas son los vectores de coordenadas de $f(a_1), \ldots, f(a_n)$ en cualquier base de F.

Sea, entonces, $A = M(f, (a_i), (b_i))$, cuyas columnas son los vectores de coordenadas en la base $(b_i)_{1 \leq i \leq m}$ de los vectores $f(a_1), f(a_2), \ldots, f(a_n)$.

De todo lo expuesto anteriormente resulta:

$$\mathrm{rg}\, f = \dim \mathrm{Im}\, f = \dim \langle f(a_1), \ldots, f(a_n) \rangle = \mathrm{rg}\, A.$$

Por tanto, podemos afirmar que *el rango de una aplicación lineal f entre espacios vectoriales E y F de dimensiones finitas n y m respectivamente, es igual al rango de la matriz asociada a f en cualquier par de bases de E y F.*

Ejemplo

Sea f: $\mathbb{R}^3 \longrightarrow \mathbb{R}^3$ la aplicación dada por

$$f(x, y, z) = \left(\frac{x}{3} + y,\ z - \frac{x}{3},\ \frac{2}{3}x + y - z \right).$$

Para hallar el rango de f, basta buscar el rango de su matriz asociada en cualquier par de bases del espacio; tomemos la base canónica:

$$A = M(f(e_L)) = \begin{pmatrix} \frac{1}{3} & 1 & 0 \\ -\frac{1}{3} & 0 & 1 \\ \frac{2}{3} & 1 & -1 \end{pmatrix}$$

Hallemos el rango de A:

$$\operatorname{rg} A = \operatorname{rg} \begin{pmatrix} \frac{1}{3} & 1 & 0 \\ 0 & 1 & 1 \\ 0 & 0 & 0 \end{pmatrix} = 2.$$

Luego $\operatorname{rg} f = 2$.

Una consecuencia inmediata de la afirmación anterior es que *dos matrices equivalentes tienen el mismo rango,* ya que hemos probado que dos matrices equivalentes A y B están asociadas a una misma aplicación lineal y, por tanto, su rango será el de dicha aplicación.

Este resultado puede obtenerse de forma directa de la definición de matrices equivalentes y usando propiedades del rango del producto de matrices.

Caracterización de las matrices de $M_{m,n}(K)$ de rango r

Una consecuencia más del teorema del rango consiste en una caracterización de las matrices de rango r.

En la demostración del teorema del rango vimos que si se escoge una base de E tal que contenga una base del núcleo de f, entonces la imagen de f está generada por las imágenes de los elementos de la base que no están en el núcleo.

Sea $\operatorname{rg} f = q$ $(a_i)_{1 \leq i \leq n}$ una base de E tal que $(a_{q+1}, \ldots, a_n)$ sea una base de $\operatorname{Ker} f$, y sea $(f(a_i), \ldots, f(a_q))$ una base de $\operatorname{Im} f$;

Tomemos en F la base formada por $b_1 = f(a_1)$, $b_2 = f(a_2), \ldots, b_q = f(a_q)$, $b_{q+1}, \ldots, b_m$, y hallemos $M(f, (a_j), (b_i))$. Obtenemos:

$$\begin{aligned} f(a_1) &= b_1 \\ f(a_2) &= b_2 \\ &\cdots\cdots \\ f(a_q) &= b_q \\ f(a_{q+1}) &= 0 \\ &\cdots\cdots \\ f(a_n) &= 0 \end{aligned}$$

y

$$M(f,(a_j),(b_i)) = \begin{array}{c} \\ q \\ \\ \end{array}\left(\begin{array}{ccc|c} 1 & & & \\ & \ddots & & 0 \\ & & 1 & \\ \hline & 0 & & 0 \end{array}\right)$$

esto es,

$$M(f,(a_j),(b_i)) = \left(\begin{array}{c|c} I_q & 0 \\ \hline 0 & 0 \end{array}\right)$$

Luego, toda aplicación lineal de rango q se puede representar mediante una matriz del tipo

$$\left(\begin{array}{c|c} I_q & 0 \\ \hline 0 & 0 \end{array}\right)$$

y, por tanto, *toda* matriz de $M_{m,n}(K)$ de rango q será equivalente a una matriz del tipo anterior, ya que bastará considerarla asociada a una aplicación lineal de rango q.

Como dos matrices equivalentes a una tercera son equivalentes entre sí, llegamos a la conclusión siguiente:

Para que dos matrices de $M_{m,n}$ (K) sean equivalentes es necesario y suficiente que sean del mismo rango.

Aunque la representación hallada anteriormente para una aplicación lineal es válida para endomorfismos, la matriz de forma sencilla hallada está dada en general en dos bases distintas del espacio. Esto significa que la relación entre dicha matriz sencilla y cualquier otra que represente al endomorfismo, será de equivalencia pero no de semejanza, por lo que no será útil para los endomorfismos.

2.7 Aplicaciones lineales y sistemas de ecuaciones

Al estudiar los sistemas de ecuaciones lineales vimos que era posible dar a estos una interpretación matricial. Así, dado el sistema S de m ecuaciones con n incógnitas

$$\left\{\begin{array}{l} a_{11}x_1 + a_{12}x_2 + \ldots + a_{1n}x_n = b_1 \\ a_{21}x_1 + a_{22}x_2 + \ldots + a_{2n}x_n = b_2 \\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots \\ a_{m1}x_1 + a_{m2}x_2 + \ldots + a_{mn}x_n = b_m \end{array}\right.$$

y considerando la matriz del sistema $A = (a_{ij})$ y los vectores columna

$$X=\begin{pmatrix} x_1 \\ \cdot \\ \cdot \\ \cdot \\ x_n \end{pmatrix} \text{ y } B=\begin{pmatrix} b_1 \\ \cdot \\ \cdot \\ \cdot \\ b_m \end{pmatrix},$$

el sistema se expresa por:

$$AX=B.$$

Podemos dar otra interpretación del sistema en términos de aplicaciones lineales, pues si consideramos X y B como vectores de K^n y K^m, respectivamente, y la aplicación f de K^n en K^m cuya matriz en la base canónica de dichos espacios es A, entonces el sistema tendrá solución si existe X de K^n tal que su imagen por f sea B, y el conjunto solución del sistema S será en este caso la preimagen por f de B.

Analicemos, en particular, el caso de un sistema homogéneo

$$AX=0.$$

Según la interpretación que acabamos de explicar el espacio solución del sistema $AX=0$ sería el núcleo de la aplicación lineal f de K^n en K^m determinado por la matriz A. De aquí que, por el teorema del rango

$$\dim \operatorname{Ker} f=\dim K^n-\dim \operatorname{Im} f=n-\operatorname{rg} f.$$

Esta relación ya la empleamos en el capítulo anterior, al plantearnos cómo buscar la dimensión y una base de un subespacio de K^n dado como el espacio solución del sistema $AX=0$.

Para encontrar una base del espacio $AX=0$ reducimos la matriz A a la forma escalonada. Supongamos que $x_1,x_2,\ldots,x_r$ son las variables dependientes y $x_{r+1},\ldots,x_n$ las independientes. Para encontrar la base del espacio solución, damos a las variables independientes, alternadamente, los valores:

$$\begin{array}{l} 1,\ 0,\ \ldots,\ 0 \\ 0,\ 1,\ \ldots,\ 0 \\ \cdots\cdots\cdots\cdots \\ 0,\ 0,\ \ldots,\ 1 \end{array}$$

Así obtenemos los vectores $a_1,\ a_2,\ldots,a_{n-r}$.

Como ya sabemos que la dimensión del espacio solución de $AX=0$ es $n-\operatorname{rg} f=n-r=k$, solo falta demostrar que el sistema de vectores $(a_1,a_2,\ldots,a_{n-r})$ es linealmente independiente. Al escribir la matriz de las componentes de estos vectores en la base canónica, obtenemos:

$$C=\begin{pmatrix} a_{11} & a_{12} & \ldots & a_{1k} \\ a_{21} & a_{22} & \ldots & a_{2k} \\ \cdots & \cdots & \cdots & \cdots \\ a_{r1} & a_{r2} & \ldots & a_{rk} \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \ldots & 1 \end{pmatrix}$$

y el rango de C es exactamente $k=n-r=\dim \operatorname{Ker} f$.

Por tanto, los vectores $(a_1, a_2, \ldots, a_{n-r})$ forman una base del espacio solución de $AX=0$.

Ejercicios

1. Analice si son lineales o no las siguientes aplicaciones. Justifique su respuesta.

   a) f: $\mathbb{R}^3 \longrightarrow \mathbb{R}^2$ dada por $f(x, y, z) = (z, x+y)$.
   b) f: $\mathbb{R} \longrightarrow \mathbb{R}^2$ dada por $f(x) = (x, 1)$.
   c) f: $\mathbb{R}^2 \longrightarrow \mathbb{R}$ dada por $f(x, y) = |x-y|$.
   d) f: $\mathbb{R}^3 \longrightarrow \mathbb{R}$ dada por $f(x, y, z) = 2x-3y+4z$.
   e) En $M_n(K)$, dada una matriz fija T;
   f: $M_n(K) \longrightarrow M_n(K)$ dada por $f(A) = AT-TA$.
   g: $M_n(K) \longrightarrow M_n(K)$ dada por $g(A) = T+A$.
   f) Dados un espacio vectorial E sobre K y un subespacio W de E:
   $i: W \longrightarrow E$ dada por $i(x) = x \; \forall x \in E$.

2. Sea f: $\mathbb{R}^2 \longrightarrow \mathbb{R}^3$ tal que
   $f(1,2) = (3, -1, 5)$,
   $f(0, 1) = (2, 1, -1)$.
   a) ¿Con estos datos bastará para hallar $f(1, -1)$?
   b) ¿Será posible hallar $f(x, y)$ para todo $(x, y) \in \mathbb{R}^2$?

3. Pruebe que si f es una aplicación lineal de E en F, entonces $f(-x) = -f(x)$ para todo $x \in E$.

4. Sean Eun espacio vectorial sobre K y $f: E \longrightarrow K$, $g: E \longrightarrow K$ dos aplicaciones lineales. Se define F: $E \longrightarrow K^2$ dada por $F(x) = (f(x), g(x))$.
   a) Pruebe que F es lineal.
   b) ¿Podrá generalizarse la definición de F al caso en que se tengan dos aplicaciones lineales: $f: E \longrightarrow F$ y $gE \longrightarrow G$?

5. Sean E un espacio vectorial sobre K y f un endomorfismo de E.
   Se dice que un subespacio W de E es estable por f si y solo si $f(W) \subset W$
   (recuerde que $f(W) = \{f(x) : x \in W\}$).
   Pruebe que f': $W \longrightarrow W$ definida por $f'(x) = f(x) \quad \forall x \in W$,
   es un endomorfismo de W. Este se denomina *restricción* de f a W.

6. Sean E_1 y E_2 espacios vectoriales sobre K y considérese el espacio producto $E_1 \times E_2$. Pruebe que
   p_1: $E_1 \times E_2 \longrightarrow E_1$ dada por $p_1(x_1, x_2) = x_1$,
   p_2: $E_1 \times E_2 \longrightarrow E_2$ dada por $p_2(x_1, x_2) = x_2$,
   son aplicaciones lineales de $E_1 \times E_2$ en E_1 y E_2, respectivamente.
   A p_1 y p_2 se les denomina *proyecciones* de $E_1 \times E_2$ sobre E_1 y E_2, respectivamente.

7. Sean E_1 y E_2 dos subespacios vectoriales de dimensión finita de un espacio vectorial E. Pruebe que la aplicación f: $E_1 \times E_2 \longrightarrow E$ dada por $f(x_1, x_2) = x_1 + x_2$ es lineal.

8. Sean E y F dos espacios vectoriales sobre K, f una aplicación lineal de E en F. Pruebe que φ: $E\times F \longrightarrow E\times F$ dada por $\varphi(x,y)=(x,y-f(x))$ es un endomorfismo de $E\times F$.
9. Sea f: $E \longrightarrow F$ lineal y suponga que $x_1,x_2,\ldots,x_k$ son vectores de E tales que $f(x_1), f(x_2),\ldots,f(x_k)$ son linealmente independientes en F. Pruebe que $x_1,x_2,\ldots,x_k$ son linealmente independientes en E.
10. a) ¿Existirá una aplicación lineal de $\mathbb{R}^3$ en $\mathbb{R}^3$ cuya imagen esté generada por $(1,1,0)$ y $(1,2,3)$? Si existe, hállela; ¿será única?
    b) Responda a las preguntas anteriores considerando una aplicación de $\mathbb{R}^4$ en $\mathbb{R}^2$ cuya imagen esté generada por $(1,1)$.
11. a) ¿Existirá una aplicación lineal de $\mathbb{R}^4$ en $\mathbb{R}^3$ cuyo núcleo esté generado por $(1,1,0,0)$ y $(2,0,3,-1)$? Si existe, hállela; ¿será única?
    b) Responda las preguntas anteriores considerando una aplicación de $\mathbb{R}^2$ en $\mathbb{R}^5$ cuyo núcleo esté generado por $(2,-1)$.
12. Sea E un espacio vectorial de dim n sobre K y $(a_1,a_2,\ldots,a_n)$ una base de E. Pruebe que r elementos ($r\leqslant n$) de E, $x_1, x_2,\ldots,x_r$, son linealmente independientes si y solo si existe un automorfismo f de E tal que $f(a_k)=x_k$ para todo $1\leqslant k\leqslant r$.
13. Halle la matriz asociada a las siguientes aplicaciones lineales en las bases dadas en cada caso
    a) f: $\mathbb{R}^3 \longrightarrow \mathbb{R}^2$ dada por $f(x_1,x_2,x_3)=(x_1+x_2,\ 2x_3-x_1)$:
       i) con respecto a las bases canónicas de $\mathbb{R}^3$ y $\mathbb{R}^2$,
       ii) con respecto a las bases (a_j) de $\mathbb{R}^3$ y (b_i) de $\mathbb{R}^2$, donde
       $a_1=(1,0,-1)$, $a_2=(1,1,1)$, $a_3=(1,0,0)$,
       $b_1=(0,1)$, $b_2=(1,0)$.
    b) f: $\mathbb{R}^3 \longrightarrow \mathbb{R}^3$ dada por $f(x,y,z)=(2y+z,\ x-4y,\ 3x)$:
       i) con respecto a las bases canónicas de $\mathbb{R}^3$, considerada en ambos espacios
       ii) con respecto a las bases $(a_i)_{1\leqslant i\leqslant 3}$ dadas por
       $a_1=(1,1,1)$, $a_2=(1,1,0)$, $a_3=(1,0,0)$
       y tomando por (b_i) la base canónica,
       iii) considerando (a_i) y (b_i) iguales a la primera base dada en ii).
    c) Dado $\mathbb{C}$ como espacio vectorial sobre $\mathbb{R}$, halle la matriz de g: $\mathbb{C} \longrightarrow \mathbb{C}$ dada por $g(a+ib)=a-ib$ en las bases $(1,i)$ y $(1+i, 1+2i)$ de $\mathbb{C}$.
14. Considere el subespacio del espacio de las funciones reales de variable real generado por $(1,\ t,\ e^t,\ te^t)$.
    a) Pruebe que $B=(1,t,e^t,te^t)$ es una base de dicho subespacio.
    b) Considere la aplicación D: $E \longrightarrow E$ dada por $D(f)=f'$. Pruebe que es un endomorfismo de E.
    c) Halle $M(D,B)$.
15. Sea $E=M_2(K)$ y $T=\begin{pmatrix} a & b \\ c & d \end{pmatrix}$ una matriz fija de E. Halle la matriz en la base usual de E

$$\left(\begin{pmatrix}1&0\\0&0\end{pmatrix}, \begin{pmatrix}0&1\\0&0\end{pmatrix}, \begin{pmatrix}0&0\\1&0\end{pmatrix}, \begin{pmatrix}0&0\\0&1\end{pmatrix}\right)$$

de los siguientes endomorfismos de E:

a) $f(A)=T\cdot A$

b) $g(A)=A\cdot T$

c) $h(A)=TA-AT$

16. Para cada número real θ, considere el endomorfismo F_θ: $\mathbf{R}^2 \longrightarrow \mathbf{R}^2$ cuya matriz en la base canónica de $\mathbb{R}^2$ es

$$\begin{pmatrix}\cos\theta & -\operatorname{sen}\theta\\ \operatorname{sen}\theta & \cos\theta\end{pmatrix}$$

Pruebe que si θ y θ' son números reales, entonces $F_\theta \circ F_{\theta'}=F_{\theta+\theta'}$ y que $F_\theta^{-1}=F_{-\theta}$.

17. Sean (a_1,a_2) una base de E y f un endomorfismo de E tal que $f(a_1)=3a_1-2a_2$, $f(a_2)=a_1+4a_2$, y sean $b_1=a_1+a_2$, $b_2=2a_1+3b_2$. Pruebe que (b_1,b_2) es una base de E y halle $M(f,\ (b_i))$.

18. Pruebe que si A es una matriz de $M_2\ (K)$ que solo es semejante a sí misma, entonces tiene la forma

$$A=\begin{pmatrix}a&0\\0&a\end{pmatrix}.$$

19. Sean f un endomorfismo de un K-espacio E y W un subespacio de E invariante por $f(f(W)\subset W)$. Si dim $W=m$, pruebe que f se puede representar mediante una matriz de la forma

$$\begin{pmatrix}A&B\\0&C\end{pmatrix},$$

donde A es una matriz de tamaño $m\times m$

Sugerencia: Tome una base de E que contenga una base de W.

20. Sea $E=V\oplus W$ y sean V y W invariantes por un endomorfismo f. Suponga que dim $V=n$ y dim $W=m$. Pruebe que T se puede representar mediante una matriz de la forma

$$\begin{pmatrix}A&0\\0&B\end{pmatrix},$$

donde A es de tamaño $n\times n$ y B de tamaño $m\times n$.

Sugerencia: Tome una base de E formada por la unión de una de V y una de W.

21. Sea A una matriz triangular superior de $M_n(K)$ en la cual todos los elementos diagonales son nulos ($a_{ij}=0$ para $i\geqslant j$). Pruebe que $A^n=0$.

Sugerencia: Considere A como la matriz asociada a un endormorfismo

f de un espacio E de dim n en una base (a_i) de dicho espacio. Observe que
$f(a_1)=0$ y que para $k>1$, $f(a_k)\in\langle a_1,a_2,\ldots,a_{k-1}\rangle$.

22. Sea A la matriz de un endomorfismo de $\mathbb{R}^3$ con respecto a la base canónica de dicho espacio, donde

$$A=\begin{pmatrix} a+b & a-b+c & a-c \\ a-b-c & a & a+b+c \\ a+c & a+b-c & a-b \end{pmatrix}.$$

Halle la matriz de f con respecto a la base (e_i') de $\mathbb{R}^3$, donde $e_1'=e_1+e_2+e_3$, $e_2'=e_2$, $e_3'=e_3$.

23. Sea f un endomorfismo de un espacio E de dimensión n sobre K, tal que existe un endomorfismo g de E para el cual $f\circ g=id_E$.
    a) Pruebe que f es inversible y que $g=f^{-1}$.
    Sugerencia: Pruebe que f es sobreyectiva.
    b) ¿Esta propiedad será válida si E no es de dimensión finita?

24. Sean k un elemento no nulo de K y f un endomorfismo de un K-espacio E.
    a) Pruebe que f es no inyectivo si y solo si kf es no inyectivo.
    b) Pruebe que f es no inyectivo si $-f$ es no inyectivo.

25. Sean $f\colon E\longrightarrow F$ y $g\colon F\longrightarrow E$ aplicaciones lineales.
    a) Pruebe que si f y g son no inyectivas, entonces $g\circ f$ es no inyectiva.
    b) Dé un ejemplo donde $g\circ f$ sea no inyectiva y g sea inyectiva.

26. Sea $f\colon \mathbb{R}^2\longrightarrow\mathbb{R}^2$ un endomorfismo tal que $f\neq 0$ pero $f^2=f\circ f=0$. Pruebe que existe una base (a_1,a_2) de $\mathbb{R}^2$ tal que $f(a_1)=a_2$ y $f(a_2)=0$.

27. Pruebe que son inversibles los siguientes endomorfismos de $\mathbb{R}^3$:
    a) $f(x,y,z)=(x-y,\ x+z,\ x+y+2z)$.
    b) $g(x,y,z)=(2x-y+z,\ x+y,\ 3x+y+z)$.

28. a) Sea f un endomorfismo de E tal que $f^2=0$. Pruebe que id_E-f es inversible.
    b) Sea f un endomorfismo de E tal que $f^3=0$. Pruebe que id_E-f es inversible.

29. Sea f un endomorfismo de E tal que $f^2-f+id_E=0$. Pruebe que f^{-1} existe y es igual a id_E-f.

30. a) ¿Puede existir una aplicación lineal inyectiva de $\mathbb{R}^3$ en $\mathbb{R}^2$?
    b) ¿Puede existir una aplicación lineal sobreyectiva de $\mathbb{R}^2$ en $\mathbb{R}^3$?
    c) Considere una aplicación lineal f de $\mathbb{R}^3$ en $\mathbb{R}^2$ y otra g de $\mathbb{R}^2$ en $\mathbb{R}^3$. Pruebe que $g\circ f$ es no inversible.

31. Halle dos endomorfismos f y g de $\mathbb{R}^2$ tales que $f\circ g=0$ pero $g\circ f\neq 0$.

32. Sea E el espacio vectorial sobre $\mathbb{R}$ de las funciones infinitamente derivables, y considere los endomorfismos de E:
    $\varphi : E\longrightarrow E$ dado por $\varphi(f)=f'$,
    $I\colon E\longrightarrow E$ dado por $I(f)=\int_0^t f(x)\,dx$.

Pruebe que si $f(0) \neq 0$, entonces $(\varphi \circ I)(f) \neq (I \circ \varphi)(f)$.

33. Considere los endomorfismos R e I del espacio de las sucesiones reales dados por

$R[(a_0, a_1, \ldots, a_n, \ldots)] = (0, a_0, a_1, \ldots)$,
$I[(a_0, a_1, \ldots, a_n, \ldots)] = (a_1, a_2, \ldots,)$
y halle $R \circ I$ e $I \circ R$.

34. Sea f un endomorfismo de un espacio E de dimensión n sobre K. Pruebe que $W = \{g \in L_k(E)$ tales que $f \circ g = 0\}$ es un subespacio vectorial de $L_k(E)$.

35. Determine el núcleo y la imagen de las siguientes aplicaciones, dando una base y la dimensión de cada uno.

a) $f\colon \mathbb{R}^3 \longrightarrow \mathbb{R}$ dada por $f(x, y, z) = 3x - 4y + 5z$.
b) $f\colon \mathbb{R}^2 \longrightarrow \mathbb{R}^2$ dada por $f(x, y) = (x + y,\ 2x + 2y)$.
c) $f\colon \mathbb{R}^3 \longrightarrow \mathbb{R}^4$ dada por $f(x, y, z) = (x + 2y,\ 2x,\ -y,\ -x - 3y)$.
d) $g\colon M^2(\mathbb{R}) \longrightarrow M_2(\mathbb{R})$ dada por $g(A) = TA - AT$, donde

$$T = \begin{pmatrix} 1 & 2 \\ 0 & 3 \end{pmatrix}.$$

e) $T\colon M_2(\mathbb{R}) \longrightarrow M_2(\mathbb{R})$ dada por $T(A) = BA$, donde

$$B = \begin{pmatrix} 1 & -1 \\ -2 & 2 \end{pmatrix}.$$

f) $p_1\colon E_1 \times E_2 \longrightarrow E_1$ dado por $(x_1, x_2) \rightsquigarrow x_1$ donde E_1, E_2 son espacios vectoriales sobre K.

36. Considere el espacio vectorial $K[x]$ y sea D el endomorfismo de $K[x]$ dado por $D(p) = p'$. Halle el núcleo y la imagen de D.

37. Halle una base y la dimensión del núcleo y de la imagen de las aplicaciones de K^4 en K^3 determinadas por:

a) $\begin{pmatrix} 1 & 2 & 0 & 1 \\ 2 & -1 & 2 & -1 \\ 1 & -3 & 2 & -2 \end{pmatrix}$ b) $\begin{pmatrix} 1 & 0 & 2 & -1 \\ 2 & 3 & -1 & 1 \\ -2 & 0 & -5 & 3 \end{pmatrix}$

38. Sean f una aplicación líneal de E en F y $k \in K$, $k \neq 0$. Pruebe que f y kf tienen el mismo núcleo y la misma imagen.

39. Sea $f\colon E \longrightarrow F$ una aplicación lineal y dado W un subespacio de E, considere la aplicación $f'\colon W \longrightarrow F$ dada por $f'(x) = f(x)\ \ \forall x \in W$ (restricción de f a W). Pruebe que f' es lineal, halle su núcleo y su imagen.

40. Sean E un espacio vectorial sobre K, p un endomorfismo de E tal que $p^2 = p \circ p = p$. Sea $V = \operatorname{Im} p$ y $W = \operatorname{Ker} p$. Pruebe que $E = V \oplus W$.
Sugerencia: Utilice la descomposición para todo x de E:
$x = (x - p(x)) + p(x)$.

41. Sean E un espacio vectorial sobre K y p_1, p_2 endomorfismos de E tales que
$p_1+p_2=id_E$,
$p_1 \circ p_2=p_2 \circ p_1=0$,
$p_1^2=p_1$, $p_2^2=p_2$.
a) Pruebe que $E=\text{Im } p_1 \oplus \text{Im } p_2$.
b) Pruebe que $\text{Im } p_1=\text{Ker } p_2$ y deduzca que $E=\text{Ker } p_2 \oplus \text{Im } p_2$.

42. Sean E un espacio vectorial de dimensión n sobre K y f un endomorfismo de E tal que $\text{Ker } f=\text{Im } f$. Pruebe que $f(-x)=f(x)$ para todo x de E. Dé un ejemplo de un endomorfismo que cumpla la condición anterior.

43. Sean E un espacio vectorial y f un endomorfismo de E. Pruebe que las proposiciones siguientes son equivalentes:
$\text{Ker } f \cap \text{Im } f=\{0\}$
y $f(f(f(x)))=0 \Rightarrow f(x)=0$.

44. Sea f una aplicación lineal de E en F. Pruebe que son equivalentes las propiedades siguientes:
f es inyectiva.
Para toda descomposición $E=E_1 \oplus E_2$ se tiene que $f(E)=f(E_1) \oplus f(E_2)$.

45. Sean E_1 y E_2 subespacios vectoriales de dimensión finita de un espacio vectorial E. Considere la aplicación $f: E_1 \times E_2 \longrightarrow E$ dada por $f(x_1, x_2)=x_1+x_2$. Pruebe que:
a) f es lineal;
b) $\text{Ker } f=\{(x, -x) \ : \ x \in E_1 \cap E_2\}$;
c) existe un isomorfismo entre $\text{Ker } f$ y $E_1 \cap E_2$;
d) $\text{Im } f=E_1+E_2$;
e) Recuerde que $\dim (E_1 \times E_2)=\dim E_1+\dim E_2$ y deduzca de nuevo que $\dim (E_1+E_2)+\dim (E_1 \cap E_2)=\dim E_1+\dim E_2$.

46. Dados f y g endomorfismos de un espacio vectorial E sobre K, pruebe que
$\text{Ker } (g \circ f)=f^{-1} (\text{Ker } g \cap \text{Im } f)$.

47. Sea $E=K[x]$ y considere la base $(1, x, x^2, \ldots, x^n, \ldots)$ de $K[x]$ y el endomorfismo f de E definido por:

$$\begin{aligned} f(0)&=1 \\ f(1)&=0 \\ \left.\begin{aligned} f(x^{2p})&=x^p \\ f(x^{2p+1})&=0 \end{aligned}\right\} & \text{ para } p>1 \end{aligned}$$

a) Pruebe que f es sobreyectiva pero no inyectiva.
b) Pruebe que existe un endomorfismo g inyectivo y no sobreyectivo tal que $f \circ g=id_E$.

48. Sean E y F dos espacios vectoriales de dimensión finita, V y W subespacios vectoriales de E y F, respectivamente, tales que $\dim V+$ $+\dim W=\dim E$. Pruebe que existe al menos una aplicación lineal de E en F tal que $\text{Ker } f=V$, $\text{Im } f=W$

Sugerencia: Considere una base de E formada por la unión de una base de V y otra de W.

49. Sean f y g dos endomorfismos de un espacio vectorial E tales que $f \circ g = g \circ f$. Pruebe que f (Ker g) $\subset$Ker g y que f (Im g) $\subset$Im g. (Se dice que el núcleo y la imagen de g son *estables por f*.)

50. Sean dos endomorfismos f y g de E tales que *Im* $f \subset$Ker g. Pruebe que $g \circ f = 0$.

51. Sea f un endomorfismo de un espacio vectorial E sobre K. Pruebe que

$f^2 = id_E$ si y solo si $[\frac{1}{2}(f + id_E)]^2 = \frac{1}{2}(f + id_E)$.

52. Sea f un endomorfismo de un espacio E de dimensión finita sobre K.
   a) Pruebe que $E = \text{Im}\, f \oplus \text{Ker}\, f \Longleftrightarrow \text{Im}\, f = \text{Im}\, (f^2)$.
   b) Pruebe que no necesariamente un endomorfismo f que cumpla las condiciones equivalentes dadas en a), debe cumplir que $f^2 = f$.

53. Sea f un endomorfismo de E de dimensión n sobre K tal que $f \neq 0$ y $f^2 = 0$ y suponga que rg $f = r$.
   a) Pruebe que Im $f \subset$Ker f y deduzca que $2r \leqslant n$.
   b) Sea $(a_1, a_2, \ldots, a_r)$ una base de un suplementario F de Ker f. Pruebe que $b_1, b_2, \ldots, b_r$, donde $b_i = f(a_i)$ $(1 \leqslant i \leqslant r)$, son independientes y que existen $n-2r$ vectores de Ker f: $\{c_1, c_2, \ldots, c_{n-2r}\}$ tales que $B = (a_1, a_2, \ldots, a_r, b_1, \ldots, b_r, c_1, \ldots, c_{n-2r})$ es una base de E.
   c) ¿Cuál es la matriz de f en la base B?
   d) ¿Qué se obtiene en el caso en que $n = 2r$?

54. Se dice que un endomorfismo f de un espacio vectorial E de dim n sobre K es *nilpotente* de índice p, si existe $p > 1$ tal que $f^{p-1} \neq 0$ y $f^p = 0$. Igualmente, una matriz A de $M_n(K)$ es *nilpotente* de índice p, si existe $p > 1$ tal que $A^{p-1} \neq 0$ y $A^p = 0$.
   a) Pruebe que

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

   y, por tanto, el endomorfismo de $\mathbb{R}^3$ cuya matriz en la base canónica de $\mathbb{R}^3$ es A, son nilpotentes de índice 3.
   b) Pruebe que si f es nilpotente y si existe $\lambda \in K$ y $x \neq 0$ de E tal que $f(x) = \lambda x$, entonces $\lambda = 0$.
   c) Pruebe que si f es nilpotente de índice p y x es tal que $f^{p-1}(x) \neq 0$, entonces los vectores
   $b_1 = x = f^0(x)$, $b_2 = f^1(x), \ldots, b_p = f^{p-1}$ (x)
   son linealmente independientes.
   d) Suponga que f es nilpotente de índice n y halle la matriz de f en la base $(b_1, b_2, \ldots, b_n)$ calculada en el inciso c).
   Sugerencia: En b), halle la expresión de $f^p(x)$ y deduzca una con-

dición para λ. En c), plantee la definición de vectores linealmente independientes y aplique f.

55. Halle todas las matrices de $M_2(\mathbb{R})$ tales que $A^2=I_2$.

56. Sea E un espacio de dimensión n sobre K y considere un endomorfismo de E tal que $f^2=id_E$. Tome los endomorfismos
$g=id_E+f, \qquad h=id_E-f.$
   a) Pruebe que $g(E)$ y $h(E)$ son estables por f, o sea, que $f(g(E))\subset g(E)$ y $f(h(E))\subset h(E)$ y que $g(E)$ y $h(E)$ son subespacios suplementarios en E.
   b) Halle las restricciones de f a $g(E)$ y $h(E)$, o sea,
   $f_g : g(E) \longrightarrow g(E)$ dada por $f_g(y)=f(y) \quad \forall y \in g(E)$,
   $f_h : h(E) \longrightarrow h(E)$ dada por $f_h(y)=f(y) \quad \forall y \in h(E)$.
   c) Forme una base de E mediante la unión de una base de $g(E)$ y una base de $h(E)$.
   d) Deduzca que toda matriz de $M_n(K)$ tal que $M^2=I_n$ es semejante a una matriz del tipo

$$\begin{pmatrix} I_p & 0 \\ 0 & -I_q \end{pmatrix}$$

con $p, q \in \mathbb{N}$ y $p+q=n$.

3 Vectores y valores propios

Introducción

En el capítulo anterior, al estudiar las transformaciones lineales de un espacio vectorial de dimensión finita en otro espacio vectorial, también de dimensión finita, vimos que estas podían representarse mediante matrices, pero que la matriz obtenida dependía de las bases que se fijaran en el dominio y el codominio de la aplicación. En particular, si se trataba de endomorfismos de un espacio vectorial, al fijar una base del espacio se obtenía una matriz cuadrada, pero esta matriz a su vez dependía no solamente del endomorfismo, sino también de la base que se prefijase en el espacio.

Ahora podemos plantearnos el problema siguiente:

De todas las matrices que pueden representar a un mismo endomorfismo, ¿cómo encontrar la más simple?

Por esta expresión, "la más simple", debe entenderse aquella matriz que tenga como elementos la mayor cantidad de ceros.

¿Cuál será el tipo más simple posible de matriz? Si exceptuamos la matriz que tiene todos sus elementos cero, y que solamente representa al endomorfismo nulo, las matrices más simples son aquellas que llamamos *diagonales*, y tienen la forma:

$$D=\begin{pmatrix} \lambda_1 & 0 & 0 \dots & 0 \\ 0 & \lambda_2 & 0 \dots & 0 \\ 0 & 0 & \lambda_3 \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 \dots & \lambda_n \end{pmatrix}$$

Es decir, todos los elementos que no están en la diagonal principal son cero, pero esto no significa que algún elemento en la diagonal principal no pueda ser también nulo.

Un caso particular de matrices diagonales son las llamadas *matrices escalares:*

$$\begin{pmatrix} \lambda & 0 & 0 \dots & 0 \\ 0 & \lambda & 0 \dots & 0 \\ 0 & 0 & \lambda \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & 0 \dots & \lambda \end{pmatrix} = \lambda I_n$$

Vamos a trabajar con una matriz diagonal de orden 3.

$$\begin{pmatrix} \lambda_1 & 0 & 0 \\ 0 & \lambda_2 & 0 \\ 0 & 0 & \lambda_3 \end{pmatrix}$$

y supongamos que ella representa un endomorfismo T de K^3 en K^3 en la base (a_1, a_2, a_3). En este caso,

$$T(a_1) = \lambda_1 a_1, \quad T(a_2) = \lambda_2 a_2, \quad T(a_3) = \lambda_3 a_3.$$

Los vectores de la base en que T tiene una forma diagonal, cumplen una propiedad muy especial; se transforman en múltiplos de ellos mismos. En otras palabras, las rectas generadas por a_i, Ka_i, se transforman en sí mismas, o sea, son subespacios "estables" para el endomorfismo T.

Podemos, entonces, plantearnos el problema siguiente:

Dado un endomorfismo T de K^n, ¿podrá este representarse por una matriz diagonal?

Seguramente el estudiante se imagina que este problema no puede ser tan simple como para responder afirmativamente. Por ello, vamos a investigar cuáles son las condiciones para que un endomorfismo se pueda representar mediante una matriz diagonal, es decir, para que sea "diagonalizable". La búsqueda debe comenzar por localizar las rectas que son estables por el endomorfismo.

3.1 Vectores y valores propios

Sean E un espacio vectorial de dimensión finita n y T un endomorfismo de E. Si T es diagonalizable, es decir, si existe una base de E, $(e_1, e_2, \ldots, e_n)$, tal que en esta base

$$M(T, (e_i)) = \begin{pmatrix} \lambda_1 & 0 & \ldots & 0 \\ 0 & \lambda_2 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

entonces los escalares que se encuentran en la diagonal están relacionados con los vectores de la base por la ecuación

$$T(e_i) = \lambda_i e_i.$$

Para resolver el problema planteado, debemos investigar si hay escalares y vectores del espacio que cumplen la relación

$$T(v) = \lambda v,$$

donde $\lambda \in K$, $v \in E$.

Los escalares y vectores que cumplen esta relación desempeñan un papel muy importante en el desarrollo del Álgebra lineal. Es por eso que les daremos un nombre especial.

DEFINICIÓN 3.1

Sean E un espacio vectorial sobre K, T un endomorfismo de E. Se denomina *valor propio* de T a todo escalar $\lambda \in K$ tal que existen vectores no nulos v de E que cumplen:

$$T(v) = \lambda v.$$

Es importante observar que la condición de que existan vectores no nulos es indispensable, pues de lo contrario todo escalar sería valor propio, porque

$$T(0) = 0 = \lambda 0,$$

y la definición no tendría interés.

Observemos que no hay ninguna restricción sobre los escalares.

Ejemplos

1. En $\mathbb{R}^3$ consideremos:
   $T(x, y, z) = (3x,\ x-y,\ 2x+y+z)$.
   Para T se cumple:
   $$\begin{aligned} T(8,2,9) &= (24,6,27) = 3(8,2,9) \\ T(0,2,-1) &= (0,-2,1) = -1(0,2,-1) \\ T(0,0,1) &= (0,0,1) = 1(0,0,1) \end{aligned}$$
   Por tanto, 3, -1 y 1 son valores propios del endomorfismo T.
2. Sean E un espacio vectorial cualquiera, Id_E su endomorfismo identidad. Entonces para todo $x \in E$ se cumple:
   $Id_E(x) = x$.
   Por tanto, 1 es el único valor propio de Id_E.
3. Sean E un espacio vectorial cualquiera, T un endomorfismo no inyectivo de E. Por tanto, Ker $T \neq \{0\}$, y si $x \in$ Ker T, $x \neq 0$, entonces $T(x) = 0 = 0 \cdot x$.
   Esto quiere decir que el escalar 0 es un valor propio de todo endomorfismo no inyectivo.
4. Sea N el endomorfismo de $\mathbb{R}^3$ dado por $N(x, y, z) = (x-z,\ x-z,\ y-z)$.
   El 0 es valor propio de N pues $N(1,1,1) = (0,0,0) = 0(1,1,1)$.
5. Sean F el espacio de las funciones diferenciables de $\mathbb{R}$ en $\mathbb{R}$ y D el operador derivada.
   $f(x) \in F$, $D(f(x)) = f'(x)$.
   Entonces todo número real es valor propio de D, pues si consideramos la función e^{kx}, se cumple:
   $D(e^{kx}) = ke^{kx}$.
6. En el espacio de las sucesiones con valores en K, $K^{\mathbb{N}}$, consideremos el endomorfismo
   $R : K^{\mathbb{N}} \longrightarrow K^{\mathbb{N}}$,
   $R(a_0, a_1, a_2, \ldots) = (0, a_0, a_1, \ldots)$.

Este endomorfismo no tiene valores propios, pues si $\lambda \in K$ fuera un valor propio, esto significaría que existe una sucesión no nula tal que

$$R(a_0, a_1, a_2, \ldots) = \lambda(a_0, a_1, a_2, \ldots)$$
$$(0, a_0, a_1, \ldots) = (\lambda a_0, \lambda a_1, \lambda a_2, \ldots).$$

Por tanto,

$0 = \lambda a_0,\ a_0 = \lambda a_1,\ a_1 = \lambda a_2, \ldots$

y entonces, o bien λ es nulo y todos los a_i son cero, o λ es no nulo y de la primera igualdad obtenemos $a_0=0$, que al sustituirla en la segunda da $a_1=0$, y así sucesivamente, todos los $a_i=0$, $i=1,2,\ldots$

Por tanto, para cualquier $\lambda \in K$ no existe una sucesión no nula que satisfaga

$R(a_0, a_1, a_2, \ldots) = \lambda(a_0, a_1, a_2, \ldots)$, y esto quiere decir que la transformación R de $K^{\mathbb{N}}$ no tiene valores propios.

7. En $\mathbb{R}^2$ consideremos el endomorfismo dado por la matriz

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

Investiguemos cuáles son los valores propios de A.
Si λ es valor propio, existe $(x, y) \in \mathbb{R}^2$, no nulo, tal que

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix},$$

$$\begin{pmatrix} y \\ -x \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}.$$

Obtenemos las ecuaciones
$y = \lambda x,\ -x = \lambda y.$
Sustituyendo obtenemos:
$y = \lambda x = \lambda(-\lambda y) = -\lambda^2 y,\ y(1+\lambda^2) = 0.$
Como $y \neq 0$, esto significa que un valor propio de A es un número real λ que satisface
$\lambda^2 + 1 = 0,$
lo cual es imposible.
Por tanto, el endomorfismo de $\mathbb{R}^2$ definido por la matriz A no tiene valores propios.

8. En $\mathbb{C}^2$ consideremos el endomorfismo dado por la matriz

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

Los valores propios de A son los números complejos λ tales que existe $(x, y) \in \mathbb{C}^2$, no nulo, que satisface:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \lambda \begin{pmatrix} x \\ y \end{pmatrix},$$

$$\begin{pmatrix} y \\ -x \end{pmatrix} = \begin{pmatrix} \lambda x \\ \lambda y \end{pmatrix}.$$

De nuevo tenemos las ecuaciones
$y = \lambda x$, $-x = \lambda y$,
pero ahora λ, x, y son números complejos.
Sustituyendo obtenemos:
$y = \lambda x = \lambda(-\lambda y) = -\lambda^2 y$,
$y(1 + \lambda^2) = 0$,
que tiene dos soluciones en $\mathbb{C}$: i, $-i$. Comprobemos.
Al buscar el vector no nulo para i, suponemos que
$x = x_1 + ix_2$, $y = y_1 + iy_2$

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = i \begin{pmatrix} x \\ y \end{pmatrix}.$$

Las ecuaciones son:
$y = ix \quad , \quad -x = iy$
$y_1 + iy_2 = ix_1 - x_2 \quad , \quad x_1 + ix_2 = -iy_1 + y_2$

y obtenemos las siguientes relaciones entre las partes reales y las partes imaginarias de x, y:
$y_1 = -x_2$, $y_2 = x_1$,
$y_2 = x_1$, $-y_1 = x_2$.
Por tanto,
$y_1 = -x_2$, $y_2 = x_1$,
y el vector que buscamos tiene la forma
$(a + ib, -b + ia)$.
Así por ejemplo, el vector de $\mathbb{C}^2$
$v = (1, i)$
es un vector no nulo que satisface la relación
$Av = iv$.
Proponemos al estudiante que realice los cálculos para el escalar $-i$, y encuentre la forma de los vectores que satisfacen la relación $Av = -iv$.

Como vemos, hay posibilidades muy diversas para los valores propios de los endomorfismos. Hay endomorfismos con un número finito de valores propios, otros que tienen todos los escalares como valores propios, los hay que no tienen valores propios, etc. Por ello tenemos que precisar un método para encontrar los valores propios.

Los ejemplos 1, 7 y 8, en los cuales empleamos matrices, indican que hay también una vía matricial de plantear este problema. Por eso vamos a definir los valores propios de una matriz.

DEFINICIÓN 3.2

Sea A una matriz de orden n con coeficientes en K. Se dice que $\lambda \in K$ es un *valor propio* de A, si λ es valor propio del endomorfismo de K^n que determina la matriz A.

Una terminología que proviene del Análisis funcional consiste en denominar al conjunto de los valores propios de un endomorfismo, el "espectro" del endomorfismo. En análisis funcional también los endomorfismos se llaman *operadores* y se emplea *el espectro del operador T*.

Asociada a la definición de valor propio, está la definición siguiente:

DEFINICIÓN 3.3

Sean E un espacio vectorial sobre K, T un endomorfismo de E y λ un valor propio de T. Se denomina *vector propio* asociado al valor propio λ, a todo vector v de E que satisface la ecuación

$$T(v) = \lambda v.$$

Observemos que para definir *vector propio* partimos de un valor propio, es decir, de la existencia de vectores no nulos que satisfacen la ecuación; sin embargo, la definición que damos no excluye al vector 0 como vector propio.

Nota: Algunos autores denominan *vector propio* solamente a vectores no nulos, y no dan al vector 0 la categoría de vector propio. Desde nuestro punto de vista, esto no es fundamental, y el considerar al vector 0 como vector propio resuelve una pequeña dificultad que se presenta en las propiedades sobre los vectores propios y que veremos más adelante. Es importante observar que en la definición partimos de que λ es valor propio, es decir, hay vectores no nulos que cumplen la ecuación $T(v) = \lambda(v)$, pues de lo contrario, no tiene sentido la definición.

Ejemplos

1. Si λ es un valor propio de un endomorfismo T de un espacio vectorial, el vector nulo es un vector propio asociado a λ.
2. En $\mathbb{R}^3$, para el endomorfismo $T(x, y, z) = (3x,\ x-y,\ 2x+y+z)$, el vector $(8,2,9)$ es vector propio del valor propio 3, el vector $(0,2,-1)$ es vector propio del valor propio -1 y el vector $(0,0,1)$ es vector propio del valor propio 1.
3. Todo vector x de un espacio vectorial es vector propio del valor propio 1 del endomorfismo identidad.
4. Si T es un endomorfismo no inyectivo de un espacio vectorial E, todo vector del núcleo de T es vector propio del valor propio 0.
5. Para el operador D definido en el espacio de las funciones diferenciables como la derivada de una función, la función e^{kx} es vector propio del valor propio $k \in \mathbb{R}$.

6. En el endomorfismo de $\mathbb{R}^3$:
$N(x,y,z)=(x-z,\ x-z,\ y-z)$,
el vector (1,1,1) es vector propio asociado al valor propio 0.

Si observamos la ecuación que cumplen los vectores propios asociados a un valor propio λ, podemos inferir las propiedades siguientes:

(1) 0 es vector propio de λ.

(2) Si v es vector propio de λ, $(-v)$ es también vector propio de λ pues $T(-v)=-T(v)=-(\lambda v)=\lambda(-v)$.

(3) Si v es vector propio de λ, todo vector de la recta generada por v, Kv, es también un vector propio de λ.
$w\in Kv$, $w=av$ para $a\in K$
$T(w)=T(av)=aT(v)=a(\lambda v)=(a\lambda)v)=(\lambda a)v=\lambda(av)=\lambda w$.

Todo parece indicar que el conjunto de los vectores propios de un valor propio dado, tiene propiedades que cumplen los subespacios vectoriales. Veamos

PROPOSICIÓN 3.1

Sean E un espacio vectorial, T un endomorfismo de E y λ un valor propio de T. Entonces el subconjunto de todos los vectores propios asociados al valor propio λ, es un subespacio vectorial de E.

Demostración

Denotemos por $E(\lambda)$ al conjunto de todos los vectores propios asociados al valor propio λ.

$$v\in E(\lambda) \Longleftrightarrow T(v)=\lambda v.$$

Probemos que $E(\lambda)$ es un subespacio de E.
Sean $v, w\in E(\lambda)$, a, $\beta\in K$.

$$\begin{aligned} T(av+\beta w) &= T(av)+T(\beta w)=aT(v)+\beta T(w)\\ &= a(\lambda v)+\beta(\lambda w)=(a\lambda)v+(\beta\lambda)w\\ &= \lambda\,(av+\beta w). \end{aligned}$$

Esto significa que $av+\beta w\in E(\lambda)$ y, por tanto, $E(\lambda)$ es un subespacio de E.

El subespacio de los vectores propios asociados a un valor propio λ, se denomina el *subespacio propio* asociado a λ y se denota por $E(\lambda)$.

Nota. Si no hubiéramos definido el vector propio de forma tal, que el vector 0 se considere también un vector propio, en esta proposición tendríamos que hablar del conjunto de los vectores propios y añadir el vector 0 para obtener un subespacio.

Una propiedad importante de los subespacios propios es que las imágenes de los vectores propios son a su vez vectores propios del mismo valor propio. Esta propiedad recibe el nombre de *estabilidad* y podemos definirla para un subespacio cualquiera.

DEFINICIÓN 3.4

Sean E un espacio vectorial sobre K, T un endomorfismo de E. Se dice que un subespacio E' de E es *estable por el endomorfismo* T si cumple:

$T(E') \subset E'$.

Podemos entonces enunciar:

PROPOSICIÓN 3.2

Sea T un endomorfismo de un espacio vectorial E sobre K. Los subespacios propios asociados a los valores propios de T son estables por el endomorfismo T.

Demostración

Sea $\lambda \in K$ un valor propio de T. Debemos probar que

$T(E(\lambda)) \subset E(\lambda)$.

Sea $v \in E(\lambda)$. Probemos que $T(v) \in E(\lambda)$. Para ello llamemos v' a $T(v)$. Entonces hay que probar que v' cumple

$T(v') = \lambda v'$.
$T(v') = T(T(v)) = T(\lambda v) = \lambda T(v) = \lambda v'$.

Por tanto, $E(\lambda)$ es estable por T.

No todo subespacio estable es un subespacio propio, pues la estabilidad de los subespacios propios es de un tipo muy especial, ya que un vector propio se transforma en otro vector múltiplo de sí mismo. En otras palabras, más que los subespacios propios, las rectas generadas por los vectores propios son estables por el endomorfismo.

3.2 Polinomio característico

Vamos a tratar de expresar en otras formas la ecuación que empleamos para definir los valores propios. λ es valor propio de un endomorfismo T si existen vectores no nulos $v \in E$ tales que

$T(v) = \lambda(v)$.

Empleando el endomorfismo identidad de E, escribimos:

$$T(v) = \lambda \; id_E(v)$$
$$T(v) - \lambda \; id_E(v) = 0$$
$$(T - \lambda \; id_E) \; (v) = 0.$$

Entonces, λ es valor propio de T si existen vectores no nulos tales que $(T - \lambda \; id_E) \; (v) = 0$. Esto significa que el endomorfismo $T - \lambda \; id_E$ no es inyectivo y por tanto, que

$\text{Ker}\,(T - \lambda \; id_E) \neq \{0\}$.

En resumen, λ valor propio de $T \Longleftrightarrow T-\lambda\ id_E$ no inyectivo
$\Longleftrightarrow$ Ker $(T-\lambda Id_E) \neq \{0\}$.

Para el caso en que E sea un espacio de dimensión finita, sabemos que un operador es inyectivo si y solo si es biyectivo o inversible. Por tanto, vamos a enunciar esta caracterización para el caso de dimensión finita.

TEOREMA 3.1

Sean E un espacio vectorial de dimensión finita sobre K, T un endomorfismo de E. Las siguientes proposiciones son equivalentes:

(1) $\lambda \in K$ es valor propio de T.
(2) Ker $(T-\lambda Id_E) \neq \{0\}$
(3) El endomorfismo $T-\lambda\ Id_E$ no es inyectivo.
(4) El endomorfismo $T-\lambda\ Id_E$ no es inversible.

Veamos qué sucede para las matrices.

Para estudiar los valores propios de una matriz A, tenemos que estudiar los valores propios de la transformación que ella determina en K^n. El escalar λ es valor propio si existen vectores de K^n, $X \in K^n$, no nulos tales que

$$AX = \lambda X.$$

Empleando la matriz identidad I_n, obtenemos:

$$\begin{aligned} AX = \lambda X &\Longleftrightarrow AX = \lambda(I_n X) \\ &\Longleftrightarrow AX - \lambda(I_n X) = 0 \\ &\Longleftrightarrow (A - \lambda I_n)X = 0. \end{aligned}$$

Por tanto,

$$\begin{aligned} \lambda \text{ es valor propio de } A &\Longleftrightarrow \text{Ker } (A - \lambda I_n) \neq \{0\} \\ &\Longleftrightarrow A - \lambda I_n \text{ no es inyectiva} \\ &\Longleftrightarrow A - \lambda I_n \text{ no es inversible} \\ &\Longleftrightarrow \det (A - \lambda I_n) = 0. \end{aligned}$$

Tenemos que buscar los valores propios de la matriz A entre las raíces de la ecuación

$$\det (A - xI_n) = 0,$$

que desempeña un papel muy importante en este capítulo.

DEFINICIÓN 3.5

Sea A una matriz cuadrada de orden n con coeficientes en K. El polinomio

$$\det (A - xI_n),$$

que se obtiene restando la variable x en la diagonal principal de la matriz A y hallando el determinante de la matriz que resulta, se denomina el *polinomio característico* de la matriz A.

El polinomio característico de una matriz de orden n tiene grado n, y sus raíces, que están en K, son los valores propios de A. La ecuación $\det(A-xI_n)=0$ se denomina *ecuación característica.*

Ejemplos

1. Consideremos la matriz de $M_3(\mathbb{R})$:

$$\begin{pmatrix} 3 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & 1 & 1 \end{pmatrix}$$

Su polinomio característico es

$$\begin{vmatrix} 3-x & 0 & 0 \\ 1 & -1-x & 0 \\ 2 & 1 & 1-x \end{vmatrix} \begin{aligned} &=(3-x)\ (-1-x)\ (1-x) \\ &=-x^3+3x^2+x-3 \end{aligned}$$

y sus raíces reales son los números 3, 1 y -1.

2. Consideremos la matriz de $M_2(\mathbb{R})$:

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}.$$

Su polinomio característico es

$$\begin{vmatrix} -x & 1 \\ -1 & -x \end{vmatrix} = x^2+1.$$

Este polinomio no tiene raíces reales; por tanto, la matriz no tiene valores propios reales.

3. Consideremos la misma matriz del ejemplo anterior, o sea,

$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

pero ahora como matriz en $M_2(\mathbb{C})$.
En este caso el polinomio característico sigue siendo x^2+1,
pero ahora sí tiene raíces: i, $-i$, que son los valores propios de la matriz en el campo de los números complejos.

Las matrices que hemos presentado como ejemplos definen endomorfismos de $\mathbb{R}^3$, $\mathbb{R}^2$ y $\mathbb{C}^2$, respectivamente. Sin embargo, en un espacio vectorial E de dimensión finita n, dada una base $(a_1,a_2,\ldots,a_n)$, todo endomorfismo T de E está representado por una matriz.

¿Podemos establecer alguna relación entre los valores propios de T y los valores propios de una de sus matrices? La respuesta es afirmativa, pero surge un inconveniente: ¿de cuál de las matrices que representan al endomorfismo se trata? pues sabemos que la representación de un endomorfismo por una matriz depende de la selección de una base en el espacio E.

PROPOSICIÓN 3.3

Dos matrices semejantes tienen el mismo polinomio característico.

Demostración

Sean A y B dos matrices de orden n, semejantes; es decir, existe una matriz P inversible tal que $B=P^{-1}AP$. Entonces, por la definición del polinomio característico y las propiedades de los determinantes, tenemos:

$$\begin{aligned}\det(B-tI_n) &= \det(P^{-1}AP-tI_n)\\ &= \det(P^{-1}AP-tP^{-1}I_n\,P)\\ &= \det(P^{-1}(A-tI_n)\,P)\\ &= (\det P^{-1})(\det(A-tI_n))\,(\det P)\\ &= \det(A-tI_n).\end{aligned}$$

Por tanto, el polinomio característico de la matriz B es igual al polinomio característico de la matriz A.

Esta propiedad es la base para establecer la relación entre los valores propios de un endomorfismo de un espacio vectorial de dimensión finita y los valores propios de una matriz que represente a este endomorfismo pues si T es un endomorfismo de un espacio vectorial E, de dimensión finita, todas las matrices que representan a T son semejantes entre sí y según la proposición 3.3 tienen el mismo polinomio característico y podemos tomar cualquier matriz que represente a T para calcular el polinomio característico y, por tanto, los valores propios. Entonces, si E es un espacio vectorial de dimensión n, $(e_1,e_2,\ldots,e_n)$ una base de E, T un endomorfismo de E, $A=M(T,(e_i))$ y $[v]$ las coordenadas de un vector v en la base (e_i), tenemos:

$$\begin{aligned}\lambda \text{ es valor propio de } T &\iff \exists v\in E \text{ no nulo tal que } T(v)=\lambda v\\ &\iff \exists\ v\in E \text{ no nulo tal que } A[v]=\lambda[v]\\ &\iff \exists v\in E \text{ no nulo tal que } (A-\lambda I_n)[v]=0\\ &\iff \text{la matriz } A-\lambda I_n \text{ no es inversible}\\ &\iff \det(A-\lambda I_n)=0\end{aligned}$$

$$\begin{aligned}v \text{ es vector propio del valor propio } \lambda \text{ de } T &\iff T(v)=\lambda_v\\ &\iff A\,[v]=\lambda[v]\\ &\iff (A-\lambda I_n)\,[v]=0\\ &\iff [v] \text{ es solución del sistema de ecuaciones lineales homogéneo:}\\ &\qquad (A-\lambda I_n)X=0.\end{aligned}$$

TEOREMA 3.2

Sean E un espacio vectorial de dimensión n sobre K, T un endomorfismo de E y A la matriz asociada a T en una base $(e_1,e_2,\ldots,e_n)$ de E. Entonces los valores propios de T son las raíces de la ecuación característica de la matriz A:

$$\det(A-tI_n)=0$$

Además, el subespacio propio $E(\lambda)$ asociado al valor propio $\lambda \in K$, es el subespacio formado por los vectores de E cuyas coordenadas en la base $(e_1, e_2, \ldots, e_n)$ satisfacen el sistema de ecuaciones lineales homogéneo

$(A - \lambda I_n) X = 0.$

Ejemplos

1. El endomorfismo de $\mathbb{R}^3$ dado por
$f(x, y, z) = (3x, x-y, 2x+y+z)$
tiene en la base canónica la matriz

$$A = \begin{pmatrix} 3 & 0 & 0 \\ 1 & -1 & 0 \\ 2 & 1 & 1 \end{pmatrix}$$

Ya sabemos que el polinomio característico de esta matriz es
$\det (A - tI_3) = (3-t)(-1-t)(1-t)$
y sus raíces en $\mathbb{R}$ son 3, 1 y -1.
El subespacio propio asociado al valor propio 3, es el espacio solución del sistema de ecuaciones.
$(A - 3I_3)\ X = 0.$

$$\begin{pmatrix} 0 & 0 & 0 \\ 1 & -4 & 0 \\ 2 & 1 & -2 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}$$

$$\begin{cases} x - 4y = 0 \\ 2x + y - 2z = 0 \end{cases}$$

La solución de este sistema es el subespacio
$E(3) = \{(8a, 2a, 9a) / a \in \mathbb{R}\}$,
o lo que es lo mismo, la recta generada por el vector $(8,2,9)$:
$E(3) = \mathbb{R}(8,2,9)$
Dejamos como ejercicio el cálculo de los subespacios propios asociados a 1 y -1.

2. Consideremos el endomorfismo T de $\mathbb{R}^4$ que en la base
$(e_1, e_2, e_3, e_4) = ((1,0,1,0), (0,1,1,0), (0,0,1,0), (0,0,0,1))$
tiene la matriz

$$A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 0 \end{pmatrix}$$

Primero calculamos el polinomio característico:
$\det (A - tI_4) = (t-2)^2 (t-1)^2$.
Por tanto, los valores propios son 2 y 1. Hallemos los subespacios propios $E(1)$ y $E(2)$.

Los vectores de $E(1)$ tienen coordenadas en la base (e_i) que satisfacen la ecuación
$(A-I_4)X=0.$

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 2 & 1 & 0 \\ -1 & -1 & 0 & 0 \\ -1 & -1 & -1 & -1 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}=0$$

Resolviendo el sistema encontramos que las relaciones que satisfacen las coordenadas son:

$x_1=0,\ x_3=-2x_2,\ x_4=x_2.$

Esto significa que las coordenadas de los vectores de $E(1)$ en la base (e_1, e_2, e_3, e_4) son de la forma
$(0, \alpha, -2\alpha, \alpha),\ \alpha\in \mathbb{R}.$
Los vectores de $E(2)$ satisfacen la ecuación $(A-2I_4)X=0.$

$$\begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ -1 & -1 & -1 & 1 \\ -1 & -1 & -1 & -2 \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{pmatrix}=0$$

Resolviendo el sistema encontramos que las relaciones que satisfacen las coordenadas son:

$x_3=-x_1-x_2,\quad x_4=0,$

y los vectores de $E(2)$ en la base (e_1, e_2, e_3, e_4) son de la forma
$(\alpha,\beta,-\alpha-\beta,0),\ \alpha,\beta\in \mathbb{R}.$
Por tanto, el endomorfismo de $\mathbb{R}^4$ dado, tiene por valores propios los números reales 1 y 2, y los subespacios propios asociados a estos valores propios son:

$E(1)=\{\alpha e_2-2\alpha e_3+\alpha e_4/\alpha\in \mathbb{R}\}$

$E(2)=\{\alpha e_1+\beta e_2-(\alpha+\beta)e_3/\alpha,\beta\in \mathbb{R}\}.$

Observemos que $\dim E_1=1,\ \dim E_2=2.$

El polinomio característico $P_T(x)$ de un endomorfismo T de un espacio vectorial E de dimensión finita n sobre K, es un polinomio de grado n y, por tanto, a lo sumo tiene n raíces, contando las raíces repetidas. Si λ es un valor propio de T que aparece con multiplicidad k como raíz del polinomio característico, decimos que λ es un valor propio de multiplicidad k.

Por los ejemplos que acabamos de presentar y lo que el estudiante sabe sobre polinomios en una indeterminada con coeficientes en K ($K=\mathbb{Q}, \mathbb{R}, \mathbb{C}$), vemos que las raíces de un polinomio dependen no solamente del polinomio en cuestión, sino también del campo de escalares que estemos considerando como coeficientes. Así, el polinomio x^2+1 no tiene raíces si lo consideramos sobre $\mathbb{Q}$ o $\mathbb{R}$ y tiene dos raíces distintas, $-i$, i, si lo consideramos sobre $\mathbb{C}$. Por ello, cuando se estén calculando los valores propios de un endomorfismo mediante su polinomio característico es importante tener en cuenta cuál es el campo de escalares K que estamos empleando.

3.3 Endomorfismo diagonalizable

Vamos a regresar al problema inicial: en un espacio E de dimensión finita sobre K, tenemos un endomorfismo T, y queremos saber cuándo existe una base en la que T se representa por una matriz diagonal.

DEFINICIÓN 3.6

Sea E un espacio vectorial de dimensión finita sobre K, se dice que un endomorfismo T de E es *diagonalizable* si existe una base de E en la cual la matriz que representa a T tiene forma diagonal.

A partir de este concepto podemos preguntarnos qué condición tiene que cumplir un endomorfismo T para ser diagonalizable.

Si T es un endomorfismo de E que en una base $(e_1,e_2,\ldots,e_n)$ se representa por una matriz diagonal

$$\begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

es evidente que los elementos de la base cumplen:

$$T(e_i) = \lambda_i \, e_i, \quad i=1,2,\ldots,n.$$

Inversamente, si en el espacio E existe una base B formada por vectores propios del endomorfismo T, $B=(e_1,e_2,\ldots,e_n)$, entonces

$$T(e_i) = \lambda_i e_i, \quad i=1,2,\ldots,n,$$

donde λ_i es el valor propio correspondiente al vector e_i.

Luego, por la definición de la matriz de un endomorfismo en una base,

$$M(T,\ (e_i)) = \begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

y T es diagonalizable.

Por tanto, podemos enunciar el teorema siguiente:

TEOREMA 3.3

En un espacio vectorial E de dimensión finita sobre K, un endomorfismo es diagonalizable si y solo si se puede formar una base de E con vectores propios del endomorfismo.

En otras palabras, la base en la cual T tiene una representación como una matriz diagonal, está formada por vectores propios de T.

El problema se transforma ahora en el siguiente: ¿Cuándo podemos formar una base del espacio E con vectores propios del endomorfismo T?

Vamos entonces a estudiar los problemas relacionados con la independencia lineal de los vectores propios y la dimensión de los subespacios propios.

TEOREMA 3.4

Sean E un espacio vectorial sobre K, T un endomorfismo de E. Si $S=(v_1,v_2,\ldots,v_k)$ es un sistema de vectores no nulos, formado por vectores propios asociados a valores propios diferentes, entonces el sistema S es linealmente independiente.

Demostración

Según las relaciones dadas por hipótesis, $\lambda_1,\lambda_2,\ldots,\lambda_k$ son valores propios diferentes y $v_1,v_2,\ldots,v_k$ son vectores propios asociados a cada uno de los valores propios

$$T(v_i)=\lambda_i\ v_i,\ \ i=1,2,\ldots,k$$
$$\lambda_i\neq\lambda_j,\ \ i\neq j.$$

Vamos a demostrar el teorema por inducción en el número de los valores propios.

Si $k=2$, tenemos λ_1,λ_2; v_1,v_2.

$$T(v_1)=\lambda_1 v_1,\ T(v_2)=\lambda_2 v_2;\ v_1\neq 0,\ v_2\neq 0,\ \lambda_1\neq\lambda_2.$$

Demostremos que v_1,v_2 son linealmente independientes. En este caso la independencia lineal significa que uno no es múltiplo del otro.

Supongamos que $v_1=av_2$. Entonces

$$T(v_1)=\lambda_1 v_1=T(av_2)=aT(v_2)=a(\lambda_2)\,v_2=\lambda_2(av_2)=\lambda_2 v_1.$$

Esto significa que $\lambda_1=\lambda_2$, lo cual contradice la hipótesis. Por tanto (v_1,v_2) es un sistema linealmente independiente.

Supongamos ahora, como hipótesis de inducción, que $(v_1,v_2,\ldots,v_{k-1})$ es linealmente independiente y probemos que $(v_1,v_2,\ldots,v_k)$ es también linealmente independiente. Para ello tomamos una combinación lineal y la igualamos a cero:

$$\sum_{i=1}^{k} a_i v_i=0$$

Aplicamos T y obtenemos:

$$T\left(\sum_{i=1}^{k} a_i v_i\right)=\sum_{i=1}^{k} T(a_i v_i)=\sum_{i=1}^{k} a_i T(v_i)=\sum_{i=1}^{k} a_i\lambda_i v_i$$
$$=a_1\lambda_1 v_1+a_2\lambda_2 v_2+\ldots+a_k\lambda_k v_k=0.$$

También, suponiendo que $\lambda_k \neq 0$ (si no, reordenamos los vectores adecuadamente.)

$$\lambda_k \left(\sum_{i=1}^{k} a_i v_i \right) = \sum_{i=1}^{k} \lambda_k a_i v_i = 0.$$

Restando $a_1\lambda_1 v_1 + a_2\lambda_2 v_2 + \ldots + a_k\lambda_k v_k = 0$ y $a_1\lambda_k v_1 + a_2\lambda_k v_2 + \ldots + a_k\lambda_k v_k = 0$, obtenemos:

$$a_1 (\lambda_1 - \lambda_k) v_1 + a_2 (\lambda_2 - \lambda_k) v_2 + \ldots + a_{k-1} (\lambda_{k-1} - \lambda_k) v_{k-1} = 0.$$

Pero como por hipótesis de inducción $(v_1, v_2, \ldots, v_{k-1})$ es linealmente independiente, resulta:

$$a_1(\lambda_1 - \lambda_k) = 0, \ a_2 (\lambda_2 - \lambda_k) = 0, \ldots, a_{k-1} (\lambda_{k-1} - \lambda_k) = 0,$$

y como $\lambda_i - \lambda_k \neq 0$, $i = 1, 2, \ldots, k-1$,

$$a_1 = a_2 = \ldots = a_{k-1} = 0.$$

Volviendo a la combinación lineal inicial tenemos:

$$\sum_{i=1}^{k} a_i v_i = a_1 v_1 + a_2 v_2 + \ldots + a_k v_k = a_k v_k = 0,$$

y como $v_k \neq 0$, entonces $a_k = 0$, lo cual demuestra que el sistema $(v_1, v_2, \ldots, v_k)$ es linealmente independiente.

Ya sabemos que si tenemos vectores propios de subespacios propios diferentes, obtenemos un sistema linealmente independiente. El problema consiste en investigar si podemos tomar de cada subespacio propio un número suficiente de vectores linealmente independientes de forma tal, que al unirlos obtengamos una base de todo el espacio. Debemos, por tanto, analizar el número máximo de vectores linealmente independientes que podemos tomar de un subespacio propio, es decir, la dimensión de un subespacio propio.

¿Cómo puede ser la dimensión del subespacio propio $E(\lambda)$ asociado a un valor propio λ?

El elemento fundamental que tenemos en este momento para obtener información, es el polinomio característico. Si E es de dimensión n y T es un endomorfismo de E, el polinomio característico $p_T(x)$ tiene grado n. Si $\lambda_1, \lambda_2, \ldots, \lambda_r$ son las raíces distintas de $P_T(x)$, de multiplicidad $k_1, k_2, \ldots, k_r$, respectivamente, y $p_T(x)$ tiene todas sus raíces en K, entonces la suma de las multiplicidades de las raíces es igual al grado del polinomio: $k_1 + k_2 + \ldots + k_r = n$, que es la dimensión del espacio.

El único número que aparece aquí asociado a un subespacio propio, es la multiplicidad del valor propio correspondiente. Además, si de cada subespacio propio podemos tomar tantos vectores linealmente independientes como la multiplicidad del valor propio, al unirlos obtendremos n vectores propios linealmente independientes, y nuestro problema estará resuelto.

¿ Existe alguna relación entre dim $E(\lambda)$ y la multiplicidad de λ como raíz del polinomio característico? Veamos qué indican los ejemplos siguientes.

Ejemplos

1. El endomorfismo de $\mathbb{R}^3$:

   $T(x, y, z) = (3x,\ x-y,\ 2x+y+z)$

   tiene tres valores propios simples y la dimensión de cada uno de los subespacios propios es 1. Por tanto, tenemos:

   dim $E(\lambda) =$ multiplicidad (λ).

2. En todo espacio vectorial E de dimensión finita n, el endomorfismo id_E tiene el polinomio característico $(1-x)^n$. Por tanto, tiene un solo valor propio, 1, de multiplicidad n. El subespacio propio $E(1)$ es todo el espacio, luego,

   dim $E(\lambda) = n =$ multiplicidad (λ).

3. El endomorfismo de $\mathbb{R}^3$ dado por

   $N(x, y, z,) = (x-z,\ x-z,\ y-z)$

   tiene el polinomio característico $-t^3$. N tiene un único valor propio triple que es el 0. Su subespacio propio está formado por los vectores de la forma

   $\{(a,\ a,\ a)/a \in \mathbb{R}\}$;

   por tanto, dim $E(0) = 1$. Tenemos entonces la relación

   dim $E(0) = 1 < 3 =$ multiplicidad (0).

4. Consideremos el endomorfismo T de $\mathbb{R}^4$ que en una base dada (e_1, e_2, e_3, e_4), tiene la matriz

$$A = \begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 0 \end{pmatrix}$$

   En este caso ya vimos que el polinomio característico es $(t-2)^2(t-1)^2$ y hay dos valores propios diferentes, cada uno de multiplicidad 2.

   Al calcular los subespacios propios tenemos

   $E(1) = \{(0, a, -2a, a)/a \in \mathbb{R}\}$,

   $E(2) = \{(a, \beta, -a-\beta, 0)/a,\ \beta \in \mathbb{R}\}$,

   donde las coordenadas que damos son las coordenadas en la base (e_1, e_2, e_3, e_4).

   De aquí obtenemos:

   dim $E(1) = 1 < 2 =$ multiplicidad (1),

   dim $E(2) = 2 =$ multiplicidad (2).

De estos ejemplos podemos intuir que la relación que existe entre *dimensión* y *multiplicidad* es que dim $E(\lambda) \leqslant$ multiplicidad (λ).

¿Podemos pensar en algún argumento que contradiga esta suposición?

Los elementos que tenemos a nuestra disposición son el polinomio característico y la definición de los vectores propios.

Si para un endomorfismo T podemos obtener k vectores propios de un mismo valor propio λ, linealmente independientes, la matriz de T en una base que contenga a estos vectores propios tendrá k columnas, en las cuales todos los elementos son cero, excepto los que aparecen en la diagonal principal que serán iguales a λ. Esto indica que el factor $(\lambda-x)$ va a aparecer en el polinomio característico por lo menos a la potencia k. Decimos *por lo menos* porque sabemos que esas k "columnas diagonales" aportan k veces el factor $(\lambda-x)$; no sabemos si el resto aportará más factores $(\lambda-x)$.

Este razonamiento no contradice lo que indicaban los ejemplos; por el contrario, fortalece nuestra suposición. Veamos la demostración de la relación entre dim $E(\lambda)$ y multiplicidad (λ), que sigue la idea de nuestro razonamiento.

TEOREMA 3.5

Sean E un espacio vectorial sobre K de dimensión finita n, T un endomorfismo de E, λ un valor propio de T de multiplicidad k. Entonces

$$\dim E(\lambda) \leqslant k.$$

Demostración

Que λ es un valor propio de multiplicidad k significa que el polinomio característico $p_T(x)$ admite a $(x-\lambda)^k$ como divisor, y que no es divisible por una potencia mayor de $(x-\lambda)$.

$$p_T(x)=(x-\lambda)^k\, q(x) \text{ y } (x-\lambda) \text{ no divide a } q(x)$$

Vamos a suponer que dim $E(\lambda)>k$ y vamos a llegar a una contradicción.

Sea dim $E(\lambda)=m>k$. Entonces podemos encontrar una base de $E(\lambda)$ con m vectores, $(e_1,e_2,\ldots,e_m)$. A partir de esta base de $E(\lambda)$, completamos una base del espacio $E(e_1,e_2,\ldots,e_m,e_{m+1},\ldots,e_n)$. Recordemos que los m primeros vectores de esta base cumplen:

$$T(e_i)=\lambda e_i,\quad i=1,2,\ldots,m.$$

Si calculamos la matriz del endomorfismo T en esta base, $M(T,(e_i))$ tiene la forma:

$$A=M(T,(e_i))=(a_{ij})$$

$$A=\begin{pmatrix}
\lambda & 0 & 0 & \ldots & 0 & a_{1\,m+1} & \ldots & a_{1n}\\
0 & \lambda & 0 & \ldots & 0 & a_{2\,m+1} & \ldots & a_{2n}\\
0 & 0 & \lambda & \ldots & 0 & a_{3\,m+1} & \ldots & a_{3n}\\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\
0 & 0 & 0 & \ldots & \lambda & a_{m\,m+1} & \ldots & a_{mn}\\
0 & 0 & 0 & \ldots & 0 & a_{m+1\,m+1} & \ldots & a_{m+1\,n}\\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots\\
0 & 0 & 0 & \ldots & 0 & a_{n\,m+1} & \ldots & a_{nn}
\end{pmatrix}$$

Podemos separar esta matriz en bloques:

$$A=\begin{pmatrix} \lambda i_m & A_{mr} \\ \cdots\cdots & \cdots\cdots \\ 0 & B_{rr} \end{pmatrix}$$

donde I_m es la matriz identidad de orden m, A_{mr} es una matriz de orden $m \times r$ y B_{rr} es una matriz de orden $r \times r$, $r=n-m$.

Como el polinomio característico de T se puede calcular a partir de cualquier matriz asociada a T podemos tomar A. Pero si calculamos $p_T(x)$ a partir de A, vemos que $p_T(x)$ tiene a $(x-\lambda)^m$ como factor, lo cual contradice que λ es valor propio de multiplicidad $k<m=\dim E(\lambda)$. Por tanto,

$$\dim E(\lambda) \leqslant \text{multiplicidad } (\lambda).$$

Observemos que $\dim E(\lambda) \geqslant 1$.

Si un valor propio es raíz simple, multiplicidad 1, del polinomio característico, entonces la dimensión del subespacio propio es 1.

A partir de este resultado, podemos obtener la condición necesaria y suficiente para que un endomorfismo de un espacio vectorial de dimensión finita sea diagonalizable.

TEOREMA 3.6

Sean E un espacio vectorial de dimensión finita n sobre K, T un endomorfismo de E. T es diagonalizable si y solo si se cumplen las condiciones siguientes:

(1) El polinomio característico de T tiene todas sus raíces en K.

(2) Para cada valor propio λ de T se cumple:
$\dim E(\lambda) = \text{multiplicidad } (\lambda)$.

Demostración

1) Supongamos que T es diagonalizable. Entonces existe una base de vectores propios de T, $B=(e_1,e_2,\ldots,e_n)$, y en esta base, T se representa por una matriz de la forma:

$$A=\begin{pmatrix} \lambda_1 & 0 & 0 & \ldots & 0 \\ 0 & \lambda_2 & 0 & \ldots & 0 \\ 0 & 0 & \lambda_3 & \ldots & 0 \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & 0 & \ldots & \lambda_n \end{pmatrix}$$

Luego el polinomio característico de T se puede calcular por esta matriz y es igual a
$\det(A-tI_n)=(\lambda_1-t)\ (\lambda_2-t)\ \ldots\ (\lambda_n-t)$.
Es evidente que como el polinomio se descompone en factores lineales, todos los valores propios, que son las raíces de este polinomio, están en K.
Sean $\lambda_1,\lambda_2,\ldots,\lambda_r$ los valores propios diferentes de T, $E(\lambda_1)$, $E(\lambda_2)$,...,$E(\lambda_r)$

los subespacios propios correspondientes, $k_1, k_2, \ldots, k_r$ la multiplicidad de cada valor propio. Entonces cada valor propio λ_i aparece k_i veces en la matriz A y, por tanto, en la base B aparecen k_i vectores de $E(\lambda_i)$.
Como $\dim E(\lambda_i) \leqslant k_i$ y hay k_i vectores de $E(\lambda_i)$ en una base de E, entonces
$\dim E(\lambda_i) = k_i =$ multiplicidad (λ_i)
y esto se cumple para todo valor propio.

2) Supongamos ahora que se cumplen las condiciones (1) y (2). Entonces, por (1), todas las raíces del polinomio característico de T están en K.
Sean $\lambda_1, \lambda_2, \ldots, \lambda_r$ las raíces distintas del polinomio característico y k_1, $k_2, \ldots, k_r$ la multiplicidad de cada una de ellas. Se cumple entonces, por (1), que $k_1 + k_2 + \ldots + k_r = n$.
Según (2), para cada $i = 1, 2, \ldots, r$ el subespacio propio $E(\lambda_i)$ tiene dimensión k_i. Esto significa que se pueden extraer k_i vectores linealmente independientes de $E(\lambda_i)$. El conjunto de todos estos vectores es linealmente independiente por el teorema 3.4. Así, tenemos un conjunto que tiene $k_1 + k_2 + \ldots + k_r = n$ vectores linealmente independientes y, por tanto, es una base de vectores propios de T.
Luego, T es diagonalizable.

Ejemplos

1. El endomorfismo de $\mathbb{R}^3$ dado por
$f(x, y, z) = (3x, x - y, 2x + y + z)$
tiene tres valores propios diferentes: 3, -1 y 1, y todos son de multiplicidad 1, o simples. Por tanto, en este caso, como
$\dim E(3) = 1 =$ multiplicidad (3),
$\dim E(-1) = 1 =$ multiplicidad (-1),
$\dim E(1) = 1 =$ multiplicidad (1),
se cumple la condición (2) del teorema 3.6 y el endomorfismo dado es diagonalizable.
Para encontrar una base en la cual la matriz del endomorfismo tiene la forma diagonal, escogemos un vector de cada subespacio propio. Una base de vectores propios es
$((8,2,9), (0,2,-1), (0,0,1))$.
En esta base la matriz de la aplicación f es
$$\begin{pmatrix} 3 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$
No es necesario realizar ningún cálculo para buscar la matriz, pues ya sabemos cuál tiene que ser el resultado: una matriz diagonal, y en la diagonal los valores propios en el orden apropiado.

2. El endomorfismo de $\mathbb{R}^2$ dado por la matriz
$$\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}$$

no es diagonalizable, pues el polinomio característico x^2+1 no tiene raíces en $\mathbb{R}$.

3. El endomorfismo de $\mathbb{C}^2$ dado por la matriz

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

es diagonalizable, pues el polinomio característico x^2+1 tiene dos raíces simples en $\mathbb{C}$, i y $-i$, y en este caso
dim $E(i)=1=$multiplicidad (i),
dim $E(-i)=1=$multiplicidad $(-i)$.
Una base de vectores propios es:
$((1, i), (1, -i))$.
En esta base la matriz del endomorfismo es

$$B=\begin{pmatrix} i & 0 \\ 0 & -i \end{pmatrix}.$$

Si deseamos comprobarlo, escribimos las matrices de cambio de coordenadas adecuadas y calculamos $P^{-1}AP$; el resultado será la matriz B. Pero no es necesario realizar los cálculos, pues ya sabemos cuál tiene que ser el resultado.

4. Sea T un endomorfismo de un espacio de dimensión 4 sobre $\mathbb{R}$, que en una base (e_1, e_2, e_3, e_4) se representa por la matriz

$$A=\begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 0 \end{pmatrix}$$

Ya vimos en un ejemplo anterior que el polinomio característico de este endomorfismo es $(x-2)^2(x-1)^2$; por tanto, tiene todas sus raíces en $\mathbb{R}$ y estas son 2 y 1, cada una de multiplicidad 2.
Cuando calculamos los subespacios propios vimos que
$E(1)=\{(0, \alpha, -2\alpha, \alpha)/\alpha\in\mathbb{R}\}$,
$E(2)=\{(\alpha, \beta, -\alpha-\beta, 0)/\alpha, \beta\in\mathbb{R}\}$.
De aquí obtenemos:
dim $E(1)=1\neq$multiplicidad (1).
Por tanto, el endomorfismo T no es diagonalizable.

Hay un caso particular en el cual el endomorfismo es diagonalizable, que aparece en los ejemplos 1 y 3, y que constituye una condición suficiente para que un endomorfismo sea diagonalizable.

PROPOSICIÓN 3.4

Sean E un espacio vectorial de dimensión n sobre K, T un endomorfismo de E. Si T tiene n valores propios distintos, entonces T es diagonalizable.

Demostración

El polinomio característico de T es de grado n, y si T tiene n valores propios diferentes, entonces el polinomio característico tiene todas sus raíces en K. Esto satisface la condición (1) de diagonalización.

Sean $\lambda_1, \lambda_2, \ldots, \lambda_n$ los n valores propios de T. Cada uno tiene multiplicidad 1. Por tanto,

$$1 \leqslant \dim E(\lambda_i) \leqslant \text{multiplicidad } (\lambda_i) = 1.$$

Se cumple:

$$\dim E\ (\lambda_i) = 1 = \text{multiplicidad } (\lambda_i),$$

lo cual satisface la condición (2) de diagonalización.

Luego T es diagonalizable.

Es importante destacar que en este caso no es necesario calcular los subespacios propios para saber si el endomorfismo es diagonalizable. También es importante resaltar que el endomorfismo tiene que tener n valores propios diferentes ya que, por ejemplo, si un endomorfismo de un espacio vectorial de dimensión 4 sobre $\mathbb{R}$ tiene por polinomio característico $(x^2-1)(x^2+1)$, entonces tiene sus valores propios diferentes, 1 y -1, pero no es diagonalizable porque el polinomio característico no cumple la condición de tener todas sus raíces en $\mathbb{R}$. Por ello, hay que prestar mucha atención al emplear la condición suficiente, y emplearla exactamente como está enunciada.

3.4 Suma directa y condición de diagonalización

Las condiciones planteadas en los teoremas 3.4 y 3.6 expresan propiedades de independencia lineal y de formación de bases en términos de los vectores propios. Podemos expresar estas condiciones de forma más "elegante" en términos de los subespacios propios y la suma de los subespacios propios.

Si estudiamos la demostración de la condición necesaria y suficiente de diagonalización, podemos observar que formamos una base del espacio completo a partir de la unión de bases de los subespacios propios; pero esta es la condición para que el espacio sea suma directa de los subespacios propios. Vamos, pues, a expresar los teoremas 3.4 y 3.6 en términos de suma directa.

TEOREMA 3.7

Sean E un espacio vectorial de dimensión finita sobre K, T un endomorfismo de E. Entonces la suma de los subespacios propios del endomorfismo T es una suma directa.

Demostración

Sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los valores propios diferentes de T y $E(\lambda_1)$, $E(\lambda_2), \ldots, E(\lambda_k)$ los subespacios propios asociados a cada uno de ellos. Tenemos que probar que la suma $E(\lambda_1) + E(\lambda_2) + \ldots + E(\lambda_k)$ es directa.

Recordemos que la suma de una familia de espacios es directa si y solo si todo vector del espacio suma se expresa en forma *única* como una suma de vectores de cada uno de los subespacios de la familia.

En nuestro caso, si $x \in [E(\lambda_1) + E(\lambda_2) + \ldots + E(\lambda_k)]$, entonces existen $v_1 \in E(\lambda_1)$, $v_2 \in E(\lambda_2), \ldots, v_k \in E(\lambda_k)$ *únicos* tales que

$$x = v_1 + v_2 + \ldots + v_k.$$

Sea $V = E(\lambda_1) + E(\lambda_2) + \ldots + E(\lambda_k)$.

Vamos a basarnos en la propiedad de independencia lineal de vectores propios de valores propios diferentes. Primero vamos a demostrar que el vector 0 se escribe en forma única.

$0 \in V$, por tanto, si existen $x_1 \in E(\lambda_1)$, $x_2 \in E(\lambda_2), \ldots, x_k \in E(\lambda k)$ no nulos tales que

$$0 = x_1 + x_2 + \ldots + x_k,$$

tenemos una contradicción, pues el sistema de vectores $x_1, x_2, \ldots, x_k$ es linealmente independiente, y esto sería una combinación lineal no trivial que tiene como resultado 0. Por tanto, el vector $0 \in V$ se escribe en forma única:

$$0 = 0 + 0 + \ldots + 0.$$

Sea ahora $v \in V$ y supongamos que existen dos formas de escribirlo como suma de elementos de los $E(\lambda_i)$, $i = 1, 2, \ldots, k$:

$$v = x_1 + x_2 + \ldots + x_k, \quad x_i \in E(\lambda_i)$$
$$v = v_1 + v_2 + \ldots + v_k, \quad v_i \in E(\lambda_i)$$

Restando las dos expresiones obtenemos:

$$0 = (x_1 - v_1) + (x_2 - v_2) + \ldots + (x_k - v_k),$$

donde $x_i - v_i \in E(\lambda_i)$, $i = 1, 2, \ldots, k$, por ser $E(\lambda_i)$, un subespacio. Pero como el vector 0 se expresa en forma única,

$$x_i - v_i = 0 \Longleftrightarrow x_i = v_i, \quad i = 1, 2, \ldots, k.$$

Luego, todo elemento del espacio suma se escribe en forma única como suma de los elementos de los subespacios y, por tanto, la suma de los subespacios propios de T es directa:

$$E(\lambda_1) \oplus E(\lambda_2) \oplus \ldots \oplus E(\lambda_k).$$

Esta propiedad equivale a la de la independencia lineal de los vectores propios de valores propios diferentes.

El próximo resultado es aparentemente mucho más fácil en la terminología de la suma directa, y es el equivalente al teorema que da la condición necesaria y suficiente de diagonalización.

El poder expresar un resultado en forma más "elegante" es algo que no debemos subestimar pues siempre es preciso tener en cuenta todo lo que se encuentra detrás de la expresión "elegante". Por lo general llegar a una expresión "elegante" implica tanto trabajo como la demostración "menos elegante". En el teorema siguiente la simplicidad del resultado y de su enunciado se justifican no solamente por esto, sino por su utilidad en presentar problemas más complejos de otras ramas de la matemática.

TEOREMA 3.8

Sean E un espacio vectorial de dimensión finita n sobre K, T un endomorfismo de E. T es diagonalizable si y solo si E es la suma directa de los subespacios propios de T.

Demostración

T es diagonalizable si y solo si existe una base de E formada por vectores propios. Sean $\lambda_1, \lambda_2, \ldots, \lambda_k$ los vectores propios diferentes de T.

Si

$$E = E(\lambda_1) \oplus E(\lambda_2) \oplus \ldots \oplus E(\lambda_k),$$

esto quiere decir que una base de E se forma por la unión de bases de las $E(\lambda_i)$ y, por tanto, obtenemos una base de vectores propios y T es diagonalizable.

Si T es diagonalizable, por la condición necesaria y suficiente de diagonalización,

$$\dim E(\lambda_i) = \text{multiplicidad } (\lambda_i) = n_i \quad \text{y} \quad n_1 + n_2 + \ldots + n_k = n.$$

Pero

$$\dim[E(\lambda_1) \oplus E(\lambda_2) \oplus \ldots \oplus E(\lambda_k)] = n_1 + n_2 + \ldots + n_k = n$$

y, por tanto,

$$E = E(\lambda_1) \oplus E(\lambda_2) \oplus \ldots \oplus E(\lambda_k).$$

Aunque este teorema expresa muy bien la condición necesaria y suficiente para que un endomorfismo sea diagonalizable, en la práctica emplearemos siempre las condiciones que establece el teorema 3.6.

3.5 Diagonalización de un endomorfismo

Hay problemas de índole muy variada relacionados con los vectores y los valores propios de un endomorfismo. Sin embargo, los dos problemas básicos que hemos presentado son:

1. Dado un endomorfismo T de un espacio E hallar sus valores propios y los subespacios propios asociados a estos.
2. Dado un endomorfismo T de un espacio E, decidir si es diagonalizable o no.

Por supuesto que para resolver el segundo problema hay que resolver el primero.

Aunque estos problemas están bien definidos y el estudiante llega a establecer un método de trabajo bastante sistemático, es necesario dominar los resultados del capítulo y realizar ejercicios de todo tipo para lograr asimilar el concepto y las propiedades de los vectores y los valores propios.

Sobre la forma de resolver los dos problemas que nos hemos planteado quisiéramos hacer algunas observaciones:

a) En un espacio de dimensión finita n, el endomorfismo puede estar dado de dos formas:

1ra. Se da el endomorfismo en forma explícita, y se busca una matriz que lo represente.

2da. Se da una matriz de orden n y se trabaja con la matriz como un endomorfismo de K^n.

Ejemplos

1. Dado el endomorfismo de K^3:
$f(x,y,z) = (2x+3y, 3y-2x+z,\ z-x+y)$
diga si es diagonalizable.

2. Dada la matriz $A \in M_3(K)$:

$$A = \begin{pmatrix} 2 & 0 & 2 \\ 3 & -1 & 1 \\ 1 & 2 & 1 \end{pmatrix}$$

diga si es diagonalizable.

Los dos ejercicios plantean la misma situación, pero desde dos puntos de vista: el de los endomorfismos y el de las matrices.

b) Una vez que tenemos una matriz del endomorfismo, hay que buscar el polinomio característico y sus raíces. Aquí se presentan dos posibilidades:

1ra. Todas las raíces del polinomio característico están en K.

2da. Hay raíces del polinomio que no están en K.

Ejemplos

1. En un espacio vectorial real el polinomio característico de un endomorfismo dado es $-(t-1)(t+2)(t-3)$.

2. En un espacio vectorial real el polinomio característico de un endomorfismo dado es $t^4-1=(t-1)\ (t+1)(t^2+1)$.

Si se presenta la segunda posibilidad, ya sabemos que en K el endomorfismo no es diagonalizable. Si en el primer caso todas las raíces del polinomio están en K y son simples, el endomorfismo es diagonalizable.

c) Una vez que tenemos los valores propios, procedemos a calcular los subespacios propios correspondientes resolviendo el sistema de ecuaciones

$(A-\lambda I_n)\ X=0.$

El espacio solución de este sistema es el subespacio propio asociado a λ.

d) Comprobamos si la dimensión de cada subespacio propio coincide con la multiplicidad del valor propio correspondiente y llegamos a una conclusión sobre la diagonalización del endomorfismo (o de la matriz).

e) Si el endomorfismo es diagonalizable y se pide una base en la cual la matriz tenga la forma diagonal, hay que construir una base de vectores propios. Para ello seleccionamos una base en cada subespacio propio y la unión de todas estas bases es una base del espacio, formada por vectores propios. En esa base el endomorfismo tiene una representación por una matriz diagonal.

f) Si A es la matriz original y D es la matriz diagonal encontrada, a veces se pide encontrar una matriz P inversible tal que

$D=P^{-1}AP$.

La matriz P es la matriz de cambio de coordenadas de la base de vectores propios a la base inicial en el espacio.

A veces esta pregunta varía en su forma, pero no en su esencia. Así, se puede pedir una matriz P inversible que satisfaga una de las relaciones siguientes:

$D=PAP^{-1}$

$A=P^{-1}\ DP$

$A=PDP^{-1}$

Por tanto, no se debe trabajar mecánicamente, sino pensar cuál matriz de cambio de coordenadas se pide y calcularla.

Como vemos, el procedimiento es bastante directo, y aparentemente, excepto la selección de la matriz P final, todo es simple y sin complicaciones. Sin embargo, a veces se cometen errores de cálculo y es en estos casos donde la importancia de conocer los resultados se hace más evidente.

Veamos algunos ejemplos de errores que se pueden evitar aplicando consecuentemente los resultados.

1. Al calcular el polinomio característico se comete un error y se calculan otros valores que no son valores propios del endomorfismo.
   Este error se detecta al tratar de calcular los subespacios propios, pues si un número λ no es valor propio del endomorfismo T (matriz A), entonces el sistema de ecuaciones
   $(A-\lambda I_n)X=0$
   tiene solución única trivial. Esto indica que λ no es valor propio de A.
2. Al calcular el subespacio propio de un valor propio λ a partir de la ecuación
   $(A-\lambda I_n)X=0$
   se cometen errores y se llega a la conclusión de que el subespacio propio es $\{0\}$.
   Esta conclusión es incorrecta porque si λ es valor propio, el sistema tiene que tener soluciones no triviales. Por tanto, hay que volver a calcular el sistema. Si de nuevo se obtiene que el sistema solamente tiene solución

trivial, entonces es que λ no es valor propio y el error está en el polinomio característico. En este caso hay que empezar de nuevo con el cálculo del polinomio característico y sus raíces.

3. Como la dimensión de un subespacio propio no puede ser mayor que la multiplicidad del valor propio, esto también es un índice para saber si se ha cometido algún error al calcular los subespacios propios.

Como vemos, todos los pasos del problema están relacionados entre sí y tenemos posibilidades de detectar los errores, pero es evidente que no es el trabajo mecánico sino la aplicación consciente de la teoría estudiada lo que nos permitirá realizar un cálculo correcto.

Para concluir vamos a plantear un ejercicio tipo de este capítulo.

Ejercicio

Sea T un endomorfismo de $\mathbb{R}^4$ cuya matriz A en la base canónica es

$$A=\begin{pmatrix} -4 & 6 & -6 & 3 \\ -6 & 11 & -12 & 6 \\ -6 & 12 & -16 & 9 \\ -6 & 12 & -18 & 11 \end{pmatrix}$$

a) Calcule el polinomio característico de T.

b) Halle los valores propios de T y los subespacios propios asociados a cada uno de ellos.

c) ¿Es T diagonalizable? En caso afirmativo, encuentre una base de $\mathbb{R}^4$ en la cual T tenga como representación una matriz diagonal D.

d) Encuentre una matriz inversible P tal que $D=PAP^{-1}$.

Solución

a) Comenzamos calculando el polinomio característico $\det(A-tI_4)$.

$$\begin{vmatrix} -4-t & 6 & -6 & 3 \\ -6 & 11-t & -12 & 6 \\ -6 & 12 & -16-t & 9 \\ -6 & 12 & -18 & 11-t \end{vmatrix} =(t-2)^2(t+1)^2.$$

b) Los valores propios son las raíces reales 2 y -1 del polinomio característico, cada una de multiplicidad 2.

Calculemos los subespacios propios asociados a cada uno de los valores propios.

$E(2)$ es Ker $(A-2I_4)$, y es el espacio solución del sistema:

$$\begin{pmatrix} -6 & 6 & -6 & 3 \\ -6 & 9 & -12 & 6 \\ -6 & 12 & -18 & 9 \\ -6 & 12 & -18 & 9 \end{pmatrix} \begin{matrix} x \\ y \\ z \\ t \end{matrix} \quad = \quad \begin{matrix} 0 \\ 0 \\ 0 \\ 0 \end{matrix}$$

que transformamos en el sistema equivalente:

$$\begin{cases} 2x-2y+2z-t=0 \\ \quad\quad y-2z+t=0 \end{cases}$$

el cual tiene por solución,
$x=\alpha$ $y=2\alpha$, $z=\beta$, $t=2\beta-2\alpha$.
Por tanto,
$E(2)=\{(\alpha,2\alpha,\beta,2\beta-2\alpha)/\alpha,\beta\in\mathbb{R}\}$.
Como el sistema tiene rango 2, entonces
dim $E(2)=2=$multiplicidad (2)
como valor propio.
Una base de $E(2)$ está dada por $(a_1=(1,2,0,-2),\ a_2=(0,0,1,2))$.
$E(-1)$ es Ker $(T+I_4)$, y es el subespacio solución del sistema:

$$\begin{pmatrix} -3 & 6 & -6 & 3 \\ -6 & 12 & -12 & 6 \\ -6 & 12 & -15 & 9 \\ -6 & 12 & -18 & 12 \end{pmatrix}\begin{pmatrix} x \\ y \\ z \\ t \end{pmatrix}=\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

que transformamos en el sistema equivalente:

$$\begin{cases} x-2y+2z-t=0 \\ \quad\quad\quad z-t=0 \end{cases}$$

el cual tiene rango 2; por tanto, dim $E(-1)=2$.
La solución del sistema está dada por
$z=t$, $x=2y-z$.
Por tanto, $E(-1)=\{(2\alpha-\beta,\alpha,\beta,\beta)/\alpha,\beta\in\mathbb{R}\}$.
Una base de $E(-1)$ está dada por
$(a_3=(2,1,0,0),\ a_4=(-1,0,1,1))$.

c) T es diagonalizable, porque cumple el criterio de diagonalización: la dimensión de cada subespacio propio es igual a la multiplicidad del valor propio correspondiente y todas las raíces del polinomio característico están en $\mathbb{R}$.
En una base formada por vectores propios (a_1,a_2,a_3,a_4), la matriz de T tiene la forma diagonal:

$$D=\begin{pmatrix} 2 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{pmatrix}.$$

d) Si escribimos la matriz

$$Q=\begin{pmatrix} 1 & 0 & 2 & -1 \\ 2 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ -2 & 2 & 0 & 1 \end{pmatrix}$$

esta se obtiene al escribir las coordenadas de los vectores de la base (a_1, a_2, a_3, a_4) en la base canónica y es, por tanto, la matriz de cambio de coordenadas de la base (a_i) a la base (e_i).

La matriz Q es inversible y su inversa es

$$Q^{-1} = \begin{pmatrix} -1 & 2 & -2 & 1 \\ -2 & 4 & -5 & 3 \\ 2 & -3 & 4 & -2 \\ 2 & -4 & 6 & -3 \end{pmatrix}$$

La relación que existe entre A y D está dada por

$D = Q^{-1}AQ$.

Tenemos que encontrar una matriz P tal que $D = PAP^{-1}$.

Es evidente que esta matriz es:

$P = Q^{-1}$, que es la matriz de cambio de coordenadas de la base (e_i) a la base (a_i)

Ejercicios

1. Para cada uno de los endomorfirmos de K^n que se dan, halle el polinomio característico, los valores propios y los subespacios propios correspondientes (sobre $\mathbb{R}$ y sobre $\mathbb{C}$). Determine si el endomorfismo es diagonalizable o no (sobre $\mathbb{R}$ y sobre $\mathbb{C}$). En caso de que sea diagonalizable, encuentre una base del espacio en la cual el endomorfismo se represente por una matriz diagonal.

   a) $f(x, y, z) = (x+2y+2z,\ x+2y-z,\ -x+y+4z)$
   b) $f(x, y, z) = (-x,\ -y,\ z)$
   c) $f(x, y, z) = (3x+y+z,\ 2x+4y+2z,\ x+y+3z)$
   d) $f(x, y, z) = (2x+5y-6z, 4x+6y-9z, 3x+6y-8z)$
   e) $f(x, y, z) = (-x, y, z)$
   f) $f(x, y, z) = (4x+y+z, x+4y+z, x+y+4z)$
   g) $f(x, y, z, t) = (x-3y+3t,\ -2x-6y+13t,\ -3y+z+3t,\ -x-4y+8t)$
   h) $f(x, y, z, t) = (3x-y+z-7t,\ 9x-3y-7z-t,\ 4z-8t,\ 2z-4t)$
   i) $f(x, y, z, t) = (3x-4y+2t,\ 4x-5y-2z+4t,\ 3z-2t,\ 2z-t)$
   j) $f(x, y, z, t) = (2z+3t,\ -2z-3t,\ 2x-2y-t,\ 3x-3y-z-3t)$

2. Para cada una de las matrices siguientes, que representan endomorfismos de K^n en la base canónica, halle el polinomio característico, los valores propios y los subespacios propios asociados a cada valor propio (sobre $\mathbb{R}$ y sobre $\mathbb{C}$). Determine si el endomorfismo representado por la matriz es diagonalizable (sobre $\mathbb{R}$ y sobre $\mathbb{C}$). Si A es la matriz dada, encuentre, en los casos diagonalizables, una matriz P inversible tal que la matriz $P^{-1}AP$ sea una matriz diagonal.

a) $\begin{pmatrix} 1 & 4 \\ 2 & 3 \end{pmatrix}$

b) $\begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$

c) $\begin{pmatrix} 0 & a \\ -a & 0 \end{pmatrix}$

d) $\begin{pmatrix} 1 & -3 & 3 \\ 3 & -5 & 3 \\ 6 & -6 & 4 \end{pmatrix}$

e) $\begin{pmatrix} -3 & 1 & -1 \\ -7 & 5 & -1 \\ -6 & 6 & -2 \end{pmatrix}$

f) $\begin{pmatrix} 2 & 0 & 4 \\ 3 & -4 & 12 \\ 1 & -2 & 5 \end{pmatrix}$

g) $\begin{pmatrix} 1 & 4 & 2 \\ 0 & -3 & -2 \\ 0 & 4 & 3 \end{pmatrix}$

h) $\begin{pmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{pmatrix}$

i) $\begin{pmatrix} -2 & -2 & 1 \\ -2 & 1 & -2 \\ 1 & -2 & -2 \end{pmatrix}$

j) $\begin{pmatrix} 3 & -1 & -1 \\ -6 & -1 & 2 \\ 2 & 1 & 0 \end{pmatrix}$

k) $\begin{pmatrix} 2 & -2 & 1 \\ 1 & 3 & 1 \\ 0 & 1 & 2 \end{pmatrix}$

l) $\begin{pmatrix} 5 & 6 & -3 \\ -1 & 0 & 1 \\ 1 & 2 & -1 \end{pmatrix}$

ll) $\begin{pmatrix} 2 & -1 & 2 \\ 5 & -3 & 3 \\ -1 & 0 & -2 \end{pmatrix}$

m) $\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}$

n) $\begin{pmatrix} 3 & 1 & 0 \\ -4 & -1 & 0 \\ 4 & -8 & -2 \end{pmatrix}$

ñ) $\begin{pmatrix} 5 & 3 & 2 \\ 1 & 1 & 1 \\ 3 & 3 & 1 \end{pmatrix}$

o) $\begin{pmatrix} 5 & -1 & 9 \\ 3 & 4 & 0 \\ 1 & 1 & 1 \end{pmatrix}$

p) $\begin{pmatrix} -9 & 4 & 4 \\ -8 & 3 & 4 \\ -16 & 8 & 7 \end{pmatrix}$

q) $\begin{pmatrix} 3 & 1 & 1 \\ 2 & 4 & 2 \\ 1 & 1 & 3 \end{pmatrix}$

r) $\begin{pmatrix} 1 & 2 & 2 \\ 1 & 2 & -1 \\ -1 & 11 & 4 \end{pmatrix}$

s) $\begin{pmatrix} 1 & 2 & 3 \\ 0 & 1 & 2 \\ 0 & 0 & 1 \end{pmatrix}$

t) $\begin{pmatrix} 1 & 3 & 0 & 0 \\ 4 & 2 & 0 & 0 \\ 1 & -1 & 5 & -3 \\ 2 & 0 & 4 & -2 \end{pmatrix}$

u) $\begin{pmatrix} 3 & 1 & 0 & 0 \\ -4 & -1 & 0 & 0 \\ 7 & 1 & 2 & 1 \\ -17 & -6 & -1 & 0 \end{pmatrix}$

3. Para cada una de las matrices siguientes determine los valores propios y los subespacios propios (sobre $\mathbb{R}$ y sobre $\mathbb{C}$). Determine para qué valores de los parámetros la matriz es diagonalizable.

a) $\begin{pmatrix} 1 & a & 1 \\ 0 & 1 & b \\ 0 & 0 & c \end{pmatrix}$

b) $\begin{pmatrix} c & 2a & 0 \\ b & 0 & a \\ 0 & 2b & -c \end{pmatrix}$

c) $\begin{pmatrix} 0 & 0 & 0 & 0 \\ a & 0 & 0 & 0 \\ 0 & b & 0 & 0 \\ 0 & 0 & c & 0 \end{pmatrix}$

d) $\begin{pmatrix} a & -b & -c & -d \\ b & a & d & -c \\ c & -d & a & b \\ d & c & -b & a \end{pmatrix}$

e) $\begin{pmatrix} 0 & a & a^2 \\ \dfrac{1}{a} & 0 & a \\ \dfrac{1}{a^2} & \dfrac{1}{a} & 0 \end{pmatrix}$

4) Sea A una matriz de orden n, inversible, con coeficientes en K. Determine los valores propios de A^{-1} a partir de los valores propios de A.

5. Sean E un espacio vectorial sobre K, f un proyector de E, $f^2=f$. Determine los valores propios de f.

6. Sean E un espacio vectorial sobre K, f un endomorfismo nilpotente de E de orden p, $f^p=0$. Calcule los valores propios de f.

7. Sean E un espacio vectorial de dimensión finita sobre K, f un endomorfismo de E tal que para un entero $k>0$ se tiene $f^k=\mathrm{id}_E$. Calcule los valores propios de f.

8. Sea A una matriz de orden n con coeficientes en K. Demuestre que A y ${}^tA$ tienen los mismos valores propios.

9. Sea A una matriz de orden n con coeficientes en K. Demuestre que los valores propios son las raíces de un polinomio que tiene por término independiente a det A.

10. En el espacio $K_n[x]$ se define el endomorfismo
$f(p(x))=(x-a)(p'(x)+p'(a))-2(p(x)-p(a))$,
donde $p(x)\in K_n[x]$, $a\in K$.
Halle los valores propios y los subespacios propios de f.

11. Sea E el espacio de los polinomios con coeficientes complejos de grado menor o igual que $2n$. Considere la aplicación f de E en $\mathbb{C}[x]$ definida por
$f(P)=(x-a)\ (x-b)P'-(2nx-n(a+b)+c)P$,

donde $a, b, c\in\mathbb{C}, P\in E$, y tales que $\dfrac{c}{(a-b)}$ no es un entero.

a) Demuestre que f es un endomorfismo de E.
b) Halle los valores propios y los vectores propios de f y deduzca que f es un automorfismo de E.
c) Halle el polinomio P de E tal que $f(P)=1$.

Sugerencia: Analice la forma que tienen los polinomios de E que son vectores propios de f, a partir de sus raíces. Para el último inciso emplee una base de E formada por vectores propios de f.

12. Sean E un espacio vectorial de dimensión finita sobre K, f y g dos endomorfismos de E tales que $f\circ g=g\circ f$.

a) Demuestre que si f y g son diagonalizables, existe una base (a_i) de E tal que $M(f,(a_i))$ y $M(g,(a_i))$ son, ambas, diagonales.
b) Aplique el resultado del inciso a) a los endomorfismos de $\mathbf{R}^2$ dados en la base canónica por las matrices:

$$A=\begin{pmatrix}1 & 2\\ 0 & 2\end{pmatrix},\ B=\begin{pmatrix}3 & -8\\ 0 & -1\end{pmatrix}.$$

13. Sean f y g dos endomorfismos de un espacio vectorial E de dimensión finita n sobre K. Cada uno de los endomorfismos tiene n valores propios

diferentes. Demuestre que las dos propiedades siguientes son equivalentes:

$f \circ g = g \circ f$.

f y g tienen los mismos valores propios.

14. Sean E el espacio de las funciones continuas con valores reales de variable real, T el endomorfismo de E definido por

$$T(f)(x) = \int_0^x f(t)\,dt, \quad f(x) \in E.$$

Calcule los posibles valores propios de T.

15. Sea ε una raíz primitiva de la unidad de orden n (n impar). Halle los valores propios de la matriz:

$$\begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & \varepsilon & \varepsilon^2 & \dots & \varepsilon^{n-1} \\ 1 & \varepsilon^2 & \varepsilon^4 & \dots & \varepsilon^{2n-1} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & \varepsilon^{n-1} & \varepsilon^{2(n-1)} & \dots & \varepsilon^{(n-1)^2} \end{pmatrix}$$

16. Sean E un espacio vectorial de dimensión n sobre K, T un endomorfismo de E tal que todas las raíces de su polinomio característico están en K. Demuestre que existen dos endomorfismos T_0 y T_D tales que $T = T_0 + T_D$, donde T_D es un endomorfismo diagonalizable y T_0 es un endomorfismo nilpotente.

17. Sea U un endomorfismo de K^4 que en la base canónica del espacio tiene la matriz:

$$\begin{pmatrix} 0 & 0 & 0 & a_0 \\ 1 & 0 & 0 & -a_1 \\ 0 & 1 & 0 & -a_2 \\ 0 & 0 & 1 & -a_3 \end{pmatrix}$$

a) Calcule el polinomio característico de U.

b) Demuestre que $a_0 I_4 + a_1 u + a_2 u^2 + a_3 u^3 + u^4 = 0$.

18. Sea A una matriz cuadrada de orden 2 con coeficientes reales:

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

a) Demuestre que si b y c son del mismo signo, entonces el polinomio característico de A tiene todos sus valores propios reales.

b) Demuestre que si A es simétrica, entonces A es diagonalizable.

19. Sea A una matriz cuadrada de orden 3 con coeficientes reales. Demuestre que si el polinomio característico de A no tiene todos sus valores propios reales, entonces A, como matriz compleja, es diagonalizable.

20. Sean E un espacio vectorial de dimensión finita, n, sobre K, $(e_1, e_2, \ldots, e_n)$ una base de E, f un endomorfismo de E. Sea x_1 un vector de E y $x_2 = f(x_1)$, $x_3 = f(x_2) = f^2(x_1), \ldots, x_k = f(x_{k-1}) = f^{k-1}(x_1)$.

a) Demuestre que el subespacio F generado por los vectores $x_1, x_2, \ldots, x_k, \ldots$ es estable por el endomorfismo f.

b) Sea p el mayor entero tal que $(x_1, x_2, \ldots, x_p)$ sea linealmente independiente. Demuestre que $(x_1, x_2, \ldots, x_p)$ es una base de F. Si $F = E$ se dice que x_1 engendra a E.

c) Suponga que $A = M(f, (e_i))$ es diagonal. Demuestre que para que exista un vector x_1 que genere a E, es necesario y suficiente que los valores propios de f sean distintos dos a dos. Dé un ejemplo de una matriz de orden 2 no diagonalizable y que, sin embargo, exista un vector que genere a E.

d) Sea x_1 un vector que genera a E. ¿Qué forma tiene la matriz $B = M(f, (x_i))$ ¿Cuál es su polinomio característico?

21. Sea E un espacio vectorial de dimensión $n+1$ sobre $\mathbb{C}$, y sean f, g y h tres endomorfismos de E tales que
$h \circ f - f \circ h = 2f$, $h \circ g - g \circ h = -2g$, $f \circ g - g \circ f = -h$.
Suponga que los únicos subespacios estables para los tres endomorfismos son E y $\{0\}$.

a) Sean $x \in E$ y $\lambda \in \mathbb{C}$ tal que $h(x) = \lambda\, x$. Demuestre que el vector $y = f(x)$ satisface $h(x) = (\lambda + 2)y$ y el vector $z = g(x)$ satisface $h(z) = (\lambda - 2)z$.

b) Demuestre que existe un vector no nulo x y un escalar λ tal que $h(x) = \lambda x$, $f(x) = 0$.

c) Sea x el vector del inciso b) y sea $x_k = \dfrac{g^k(x)}{k!}$, $k \geq 0$.

Demuestre que
$h(x_k) = (\lambda - 2k)\, x_k$,
$f(x_k) = (k - 1 - \lambda)\, x_{k-1}$,
$g(x_k) = (k+1)\, x_{k+1}$.

d) A partir de los subespacios estables supuestos y de las relaciones demostradas en c), demuestre que $\lambda = n$ y que los $n+1$ vectores, $x_0, x_1, \ldots, x_n$, forman una base de E. ¿Cómo son las matrices de f, g y h en esta base? Considere en particular los casos $n = 2$ y $n = 3$.

e) Sea E el espacio de los polinomios de grado menor o igual que n. Compruebe que las condiciones de los incisos anteriores se cumplen para los endomorfismos:

$f(p) = n \times p(x) - x^2 p'(x)$,
$g(p) = p'(x)$,
$h(p) = -np(x) + 2xp'(x)$,
para $p(x) \in E$.

4 λ- matrices y matriz de Jordán

Introducción

En el capítulo anterior planteamos el problema de cómo buscar la matriz "más simple" (que tuviese más ceros) que represente a un endomorfismo dado. Como modelo de matriz más simple planteamos las matrices diagonales, y vimos que no todos los endomorfismos se pueden representar por matrices diagonales. Por tanto, queda por resolver el problema siguiente:

¿Cuál es la matriz más simple que representa a un endomorfismo dado?

Sea T un endomorfismo de un espacio vectorial E de dimensión finita, n, sobre K. Dada una base $(e_1, e_2, \ldots, e_n)$ de E, T se representa por una matriz:

$$A = M(T, (e_i)).$$

Cualquier otra matriz que represente también al endomorfismo T, es semejante a A. Es decir, si $(e'_1, e'_2, \ldots, e'_n)$ es otra base de E y $B = M(T, (e'_i))$, entonces existe una matriz P de orden n, inversible, tal que

$$B = P^{-1}AP.$$

Por tanto, la matriz más simple que represente a T se tiene que buscar entre las matrices semejantes a A. Este es un problema bastante difícil; consideremos el caso particular de dos matrices A y B de orden 2 con coeficientes reales. Queremos saber si son semejantes.
Si

$$A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}, \quad B = \begin{pmatrix} a' & b' \\ c' & d' \end{pmatrix},$$

hay que encontrar una matriz P inversible,

$$P = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix},$$

tal que $B = P^{-1}AP$, o lo que es equivalente, $PB = AP$. Efectuando el producto,

$$PB = \begin{pmatrix} \alpha a' + \beta c' & \alpha b' + \beta d' \\ \gamma a' + \delta c' & \gamma b' + \delta d' \end{pmatrix} = \begin{pmatrix} a\alpha + \gamma b & \beta a + \delta b \\ \alpha c + \gamma d & \beta c + \delta d \end{pmatrix} = AP.$$

obtenemos un sistema de ecuaciones lineales homogéneo con las incógnitas $\alpha, \beta, \gamma, \delta$:

$$\begin{cases} (a'-a)\,\alpha+c'\beta-b\gamma=0 \\ b'\alpha+(d'-a)\,\beta-\delta b=0 \\ -c\alpha+(a'-d)\,\gamma+c'\delta=0 \\ -c\beta+b'\gamma+(d'-d)\;\;\delta=0 \end{cases}$$

Tenemos que resolver este sistema de ecuaciones con la condición de que P sea inversible; por tanto, det $P\neq 0$.

Supongamos que

$\alpha\delta-\beta\gamma=1.$

Entonces, el sistema que hay que resolver es:

$$\begin{cases} (a'-a)\,\alpha+c'\beta-b\gamma=0 \\ b'\alpha+(d'-a)\,\beta-\delta b=0 \\ -c\alpha+(a'-d)\,\gamma+c'\delta=0 \\ -c\beta+b'\gamma+(d'-d)\,\delta=0 \\ \alpha\delta-\beta\gamma=1, \end{cases}$$

que es un sistema de ecuaciones no lineales con las incógnitas $\alpha, \beta, \gamma, \delta$. Por la experiencia en el trabajo con estos sistemas, sabemos que son difíciles de resolver y, en general, no tienen solución. Por ejemplo, para el caso particular

$$A=\begin{pmatrix} a & b \\ c & d \end{pmatrix},\; B=\begin{pmatrix} a' & 0 \\ 0 & d' \end{pmatrix},$$

el sistema que se obtiene:

$$\begin{cases} (a'-a)\,\alpha-b\gamma=0 \\ (d'-a)\,\beta-\delta b=0 \\ -c\alpha+(a'-d)\;\;\gamma=0 \\ -c\beta+(d'-d)\;\;\delta=0 \\ \alpha\delta\neq 0 \end{cases}$$

no siempre tiene solución, cualquiera que sea la matriz A, pues si tiene solución, esto significa que A es diagonalizable, y ya sabemos que esto no siempre es posible.

La conclusión que sacamos es que plantearnos el problema de buscar una matriz más simple semejante a una matriz dada A es muy difícil. Es por eso que vamos a desarrollar una teoría especial denominada *teoría de las λ-matrices*, que nos permitirá definir la forma más simple de una matriz, y cómo calcularla.

El desarrollo de la misma puede parecer laborioso y poco intuitivo, y esto se debe a que, en realidad, se basa en estructuras algebraicas más generales que la de espacio vectorial, y se utiliza para resolver problemas de tipo más general, de los cuales las matrices son casos particulares.

Si recordamos los conceptos estudiados en el capítulo anterior: valores propios, subespacios propios, ecuación característica, etc., vemos que todos están relacionados con la matriz

$$(A - tI_n),$$

donde t es una indeterminada. Esta matriz no tiene coeficientes en K, sino en el conjunto de los polinomios en una indeterminada con coeficientes en K. En otras palabras, es una matriz cuyos elementos son polinomios. Además, este tipo de matriz, o su evaluación para un caso particular de t, aparece en todos los resultados hallados en el capítulo anterior.

Vamos, pues, a comenzar por el estudio más general de este tipo de matrices.

4.1 Matrices polinomiales o λ-matrices

DEFINICIÓN 4.1

Sea K un campo de escalares. Se denomina *matriz polinomial* o *λ-matriz de orden $m \times n$* sobre K, a una matriz cuyos elementos son polinomios en una indeterminada con coeficientes en K. Se representa en la forma:

$$A(\lambda) = \begin{pmatrix} p_{11}(\lambda) & p_{12}(\lambda) & \dots & p_{1n}(\lambda) \\ p_{21}(\lambda) & p_{22}(\lambda) & \dots & p_{2n}(\lambda) \\ \dots & \dots & \dots & \dots \\ p_{m1}(\lambda) & p_{m2}(\lambda) & \dots & p_{mn}(\lambda) \end{pmatrix}$$

Por tradición, la indeterminada se representa por la letra griega λ(de ahí el nombre λ-matriz), pero podemos emplear también las letras x, y, t, indistintamente.

Aunque el desarrollo de la teoría puede realizarse para λ-matrices de orden $m \times n$, solamente nos ocuparemos de las λ-matrices cuadradas, es decir, de aquellas que tienen igual número de filas que de columnas.

Al igual que para las matrices que ya conocemos, para las λ-matrices definimos las operaciones de suma, producto por un escalar y producto de matrices en la forma usual; lo único que cambia en este caso es que los elementos son polinomios. También definiremos lo que llamamos transformaciones elementales, que son semejantes a las del método de Gauss para resolver sistemas de ecuaciones lineales.

DEFINICIÓN 4.2

Sea $A(\lambda)$ una λ-matriz de orden n sobre K. Se denominan *transformaciones elementales* de $A(\lambda)$ las transformaciones siguientes:

1) Multiplicar una fila cualquiera de $A(\lambda)$ por un escalar no nulo.
2) Multiplicar una columna cualquiera de $A(\lambda)$ por un escalar no nulo.
3) Sumar a una fila de la matriz $A(\lambda)$ cualquier otra fila multiplicada por un polinomio arbitrario con coeficientes en K.

(4) Sumar a una columna de la matriz $A(\lambda)$ cualquier otra columna multiplicada por un polinomio arbitrario con coeficientes en K.

Estas transformaciones se pueden representar como sigue:

Para $a \in K$, $a \neq 0$ $p(\lambda) \in K[\lambda]$,

$$(i) \to (ai), \quad \begin{pmatrix} i \\ j \end{pmatrix} \to \begin{pmatrix} i+p(\lambda)j \\ j \end{pmatrix}, \quad (i\ j) \to (i+p(\lambda)j\ j).$$

La primera es multiplicar la fila i o la columna i por el escalar a, la segunda es sumarle a la fila i la fila j multiplicada por el polinomio $p(\lambda)$ y la tercera es sumarle a la columna i la columna j multiplicada por el polinomio $p(\lambda)$.

Aunque a estas cuatro transformaciones las llamamos *elementales*, hay otras transformaciones entre λ-matrices que se pueden obtener como combinaciones de estas; el caso más significativo es el intercambio de dos filas, o de dos columnas de la λ-matriz. Por ejemplo, si queremos intercambiar la fila i y la fila j, procedemos de la manera siguiente:

1ro. Sumamos a la fila i la fila j.
2do. Restamos de la fila j la nueva fila i.
3ro. Sumamos a la fila i la nueva fila j.
4to. Cambiamos el signo de la fila j.

Si representamos esquemáticamente estas transformaciones, obtenemos la secuencia:

$$\begin{pmatrix} i \\ j \end{pmatrix} \overset{1.}{\rightsquigarrow} \begin{pmatrix} i+j \\ j \end{pmatrix} \overset{2.}{\rightsquigarrow} \begin{pmatrix} i+j \\ -i \end{pmatrix} \overset{3.}{\rightsquigarrow} \begin{pmatrix} j \\ -i \end{pmatrix} \overset{4.}{\rightsquigarrow} \begin{pmatrix} j \\ i \end{pmatrix}.$$

De manera similar, podemos intercambiar dos columnas.

Es evidente que para cada transformación elemental existe una transformación inversa que lleva la matriz a su forma original. Esto se puede representar como sigue:

Para $a \in K$, $a \neq 0$, $p(\lambda) \in K[\lambda]$,

1) Para $(i) \to (ai)$, la inversa es $(i) \to (a^{-1}i)$.

2) Para $\begin{pmatrix} i \\ j \end{pmatrix} \to \begin{pmatrix} i+p(\lambda)j \\ j \end{pmatrix}$, la inversa es $\begin{pmatrix} i \\ j \end{pmatrix} \to \begin{pmatrix} i-p(\lambda)j \\ j \end{pmatrix}$.

3) Para $(i\ j) \to (i+p(\lambda)j\ j)$, la inversa es $(i\ j) \to (i-p(\lambda)j\ j)$.

Si realizamos una transformación elemental y su inversa, una a continuación de la otra, la matriz se mantiene sin alteración.

Las notaciones que emplearemos son las siguientes:

Una transformación elemental se denota por las letras griegas σ, τ, y cuando se aplica a una matriz $A(\lambda)$ se escribe:

$\sigma A(\lambda)$.

Si se aplica primero τ y después σ, se escribe:

$\sigma\tau A(\lambda)$.

La transformación inversa de σ se denota por σ^{-1}.
Si $B(\lambda)$ es la matriz que resulta de aplicarle la transformación elemental σ a la matriz $A(\lambda)$, se escribe:

$B(\lambda)=\sigma A(\lambda)$.

La importancia de estas transformaciones elementales radica en que permiten clasificar las λ-matrices.

DEFINICIÓN 4.3

Dos λ-matrices sobre K, $A(\lambda)$ y $B(\lambda)$ son *equivalentes* si $B(\lambda)$ se puede obtener de $A(\lambda)$ mediante un número finito de transformaciones elementales. Se denota por

$A(\lambda)\sim B(\lambda)$.

Esta relación entre dos λ-matrices posee las propiedades que se dan en la proposición siguiente:

PROPOSICIÓN 4.1

(1) $A(\lambda)\sim A(\lambda)$ para toda λ-matriz $A(\lambda)$
(2) $A(\lambda)\sim B(\lambda)\Rightarrow B(\lambda)\sim A(\lambda)$
(3) $A(\lambda)\sim B(\lambda)$ y $B(\lambda)\sim C(\lambda)\Rightarrow A(\lambda)\sim C(\lambda)$

Demostración

Para (1):

Si en $A(\lambda)$ realizamos una transformación elemental y a continuación su inversa, entonces el resultado es $A(\lambda)$. Esto significa que

$A(\lambda)\sim A(\lambda)$.

Para (2):

Si $A(\lambda)\sim B(\lambda)$, entonces existen un número finito de transformaciones elementales, que denotamos por $\sigma_1, \sigma_2, \ldots, \sigma_k$, tales que

$$B(\lambda)=\sigma_k\ldots\sigma_2\sigma_1 A(\lambda).$$

Realizando en $B(\lambda)$ las transformaciones inversas de $\sigma_1, \sigma_2, \ldots, \sigma_k$ en el orden inverso, obtenemos $A(\lambda)$:

$$\begin{aligned}\sigma_1^{-1}\sigma_2^{-1}\ldots\ \sigma_k^{-1}B(\lambda) &= \sigma_1^{-1}\sigma_2^{-1}\ \ldots\ \sigma_k^{-1}\cdot\sigma_k\ \ldots\ \sigma_2\ \sigma_1 A(\lambda)\\ &= A(\lambda).\end{aligned}$$

Por tanto.

$B(\lambda) \sim A(\lambda)$.

Para (3):

Si $A(\lambda) \sim B(\lambda)$ y $B(\lambda) \sim C(\lambda)$, esto quiere decir que $C(\lambda)$ se obtiene de $B(\lambda)$ por un número finito de transformaciones elementales y que $B(\lambda)$ se obtiene de $A(\lambda)$ por un número finito de transformaciones elementales. Por tanto, realizando todas estas transformaciones, se obtiene $C(\lambda)$ a partir de $A(\lambda)$ por un número finito de transformaciones. Luego,

$A(\lambda) \sim C(\lambda)$.

De estas propiedades resulta que la relación $A(\lambda) \sim B(\lambda)$ divide a las λ-matrices en subconjuntos disjuntos o clases. En cada una de estas clases están todas las λ-matrices que son equivalentes entre sí.

Ahora planteamos el problema siguiente:

Entre todas las λ-matrices de una misma clase, ¿cuál es la λ-matriz que tiene la forma más "simple"?

Este problema tiene una respuesta positiva para todos los casos; esa matriz "más simple" es la denominada *λ-matriz canónica.*

DEFINICIÓN 4.4

Se dice que una λ-matriz es *canónica* si cumple las condiciones siguientes:

(1) Es una λ-matriz diagonal:

$$\begin{pmatrix} e_1(\lambda) & 0 & \dots & 0 \\ 0 & e_2(\lambda) & \dots & 0 \\ \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & e_n(\lambda) \end{pmatrix}$$

(2) Todos los polinomios no nulos que aparecen en la diagonal son mónicos, es decir, tienen coeficiente principal 1.

(3) Todo polinomio de la diagonal es divisible por el polinomio de la diagonal en la fila anterior, es decir, todo $e_i(\lambda)$, $i=2,3,\dots,n$, es divisible por $e_{i-1}(\lambda)$.

De esta definición de λ-matriz canónica, resultan algunas conclusiones elementales.

1ra. Si hay ceros en la diagonal, estos ocupan las posiciones de las últimas filas de la matriz.

2da. El grado de los polinomios no decrece, o sea,

grad $(e_i(\lambda)) \geqslant$ grad $(e_{i-1}(\lambda))$, $i=2,3,\dots,n$.

3ra. Si hay polinomios de grado 0, y que por ser mónicos son todos iguales a 1, estos aparecen en los primeros lugares de la diagonal.

En resumen, si hay polinomios constantes en una λ-matriz canónica, esta es de la forma:

$$\begin{pmatrix} 1 & & & & & & & & & & \\ & 1 & & & & & & & & 0 & \\ & & \ddots & & & & & & & & \\ & & & 1 & & & & & & & \\ & & & & p_1(\lambda) & & & & & & \\ & & & & & \ddots & & & & & \\ & & & & & & p_k(\lambda) & & & & \\ & & & & & & & 0 & & & \\ & & & & & & & & \ddots & & \\ 0 & & & & & & & & & 0 \end{pmatrix}$$

Hecho este análisis inicial, enunciaremos el teorema fundamental de equivalencia de λ-matrices.

TEOREMA 4.1

Toda λ-matriz es equivalente a una λ-matriz canónica.

Demostración

Vamos a hacer una demostración por inducción en el orden de la matriz.

Supongamos que $A(\lambda)$ es una λ-matriz no nula. Para matrices de orden 1 el teorema es válido, pues una λ-matriz de orden 1 es de la forma:

$(p(\lambda))$, $p(\lambda) \in K[\lambda]$,

y es diagonal.

Si $p(\lambda)$ no es un polinomio mónico, sea a el coeficiente principal de $p(\lambda)$, entonces $a^{-1}p(\lambda)$ es un polinomio mónico y la matriz $(a^{-1}p(\lambda))$ es equivalente a la matriz $(p(\lambda))$. En este caso $(a^{-1}p(\lambda))$ es una matriz canónica.

Supongamos que toda λ-matriz de orden $(n-1)$ es equivalente (se puede reducir por transformaciones elementales) a una λ-matriz canónica.

Probemos la propiedad para una λ-matriz de orden n, o sea,

$$A(\lambda) = \begin{pmatrix} p_{11}(\lambda) & p_{12}(\lambda) & \dots & p_{1n}(\lambda) \\ p_{21}(\lambda) & p_{22}(\lambda) & \dots & p_{2n}(\lambda) \\ \dots & \dots & \dots & \dots \\ p_{n1}(\lambda) & p_{n2}(\lambda) & \dots & p_{nn}(\lambda) \end{pmatrix}$$

Consideremos todas las matrices equivalentes a $A(\lambda)$ tales que el primer elemento de la primera fila sea un polinomio no nulo. Los polinomios que aparecen en esa posición en cada una de las matrices, pueden formar un conjunto infinito, y sus grados también pueden formar, en general, un conjunto infinito de números naturales. Pero todo conjunto de números naturales, tiene un elemento que es menor que todos los demás; de modo que entre todos los polinomios no nulos que ocupan el lugar (1,1) en una λ-matriz equivalente a $A(\lambda)$, podemos seleccionar uno de grado mínimo.

Sea $B(\lambda)$ una matriz obtenida de esta forma. El polinomio que está en el primer lugar de la primera fila de $B(\lambda)$, tiene grado mínimo con respecto a un polinomio que ocupe la misma posición en una matriz semejante a $A(\lambda)$. Sea $e_1(\lambda)$ este polinomio. Si $e_1(\lambda)$ no es mónico, sea a su coeficiente principal; multiplicando la primera fila de $B(\lambda)$ por el escalar a^{-1}, la matriz que resulta tiene en la primera posición de la primera fila, un polinomio mónico que cumple la misma condición de grado mínimo que $e_1(\lambda)$. Esto significa que podemos asumir que la matriz seleccionada es

$$B(\lambda) = \begin{pmatrix} e_1(\lambda) & b_{12}(\lambda) & \dots & b_{1n}(\lambda) \\ b_{21}(\lambda) & b_{22}(\lambda) & \dots & b_{2n}(\lambda) \\ \dots & \dots & \dots & \dots \\ b_{n1}(\lambda) & b_{n2}(\lambda) & \dots & b_{nn}(\lambda) \end{pmatrix}$$

donde $B(\lambda) \sim A(\lambda)$ y $e_1(\lambda)$ es mónico y de grado mínimo, en relación con todo otro polinomio que ocupe esa posición en cualquier matriz que sea equivalente a $A(\lambda)$.

Vamos a probar que $e_1(\lambda)$ divide a todos los elementos de la primera fila y la primera columna de $B(\lambda)$. Lo haremos solamente para los elementos de la primera columna; de forma similar se realiza para los de la primera fila.

Tomemos el elemento $b_{i1}(\lambda)$. Por el algoritmo de la división con resto, existen dos polinomios $q(\lambda)$, $r(\lambda)$ *únicos* tales que

$$b_{i1}(\lambda) = q(\lambda)\, e_1(\lambda) + r(\lambda),$$

donde $r(\lambda) = 0$ o grado $r(\lambda) <$ grado $e_1(\lambda)$.

Supongamos que $r(\lambda) \neq 0$. Entonces hacemos la transformación elemental siguiente: a la i-ésima fila de la matriz $B(\lambda)$ le restamos la primera fila multiplicada por el polinomio $q(\lambda)$ y a continuación intercambiamos la primera fila y la i-ésima fila resultante. Así obtenemos una matriz $B'(\lambda)$ semejante a $B(\lambda)$ y, por tanto, semejante a $A(\lambda)$.

El primer elemento de la primera fila de $B'(\lambda)$ será el polinomio

$$b_{i1}(\lambda) - q(\lambda)\, e_1(\lambda) = r(\lambda)$$

y grado $r(\lambda) <$ grado $e_1(\lambda)$, lo cual entra en contradicción con la manera en que seleccionamos la matriz $B(\lambda)$, pues el polinomio $e_1(\lambda)$ es de grado mínimo entre los polinomios que ocupan el primer lugar de la primera fila de las matrices equivalentes a $A(\lambda)$. Por tanto, $r(\lambda) = 0$. Luego,

$$b_{i1}(\lambda) = q(\lambda)\, e_1(\lambda),$$

y esto quiere decir que $e_1(\lambda)$ divide a $b_{i1}(\lambda)$.

Entonces, para cada $b_{i1}(\lambda)$, $i = 2,3,\dots,n$, existe un polinomio $q_i(\lambda)$ tal que

$$b_{i1}(\lambda) = q_i(\lambda)\ e_1(\lambda),$$

y para cada $b_{1i}(\lambda)$, existe un polinomio $q'_i(\lambda)$ tal que

$$b_{1i}(\lambda) = q'_i(\lambda)\, e_1(\lambda).$$

Transformemos ahora la matriz $B(\lambda)$ de la manera siguiente: a la i-ésima fila le restamos la primera multiplicada por el polinomio $q_i(\lambda)$; y a la i-ésima columna le restamos la primera, multiplicada por el polinomio $q'_i(\lambda)$.

Al hacer estas transformaciones, obtenemos una nueva matriz, $C(\lambda)$, equivalente a la matriz $A(\lambda)$ y que tiene la forma:

$$C(\lambda) = \begin{pmatrix} e_1(\lambda) & 0 & \ldots & 0 \\ 0 & c_{22}(\lambda) & \ldots & c_{2n}(\lambda) \\ \ldots & \ldots & \ldots & \ldots \\ 0 & c_{n2}(\lambda) & \ldots & c_{nn}(\lambda) \end{pmatrix}$$

Denotemos por

$$C'(\lambda) = \begin{pmatrix} c_{22}(\lambda) & \ldots & c_{2n}(\lambda) \\ \ldots & \ldots & \ldots \\ c_{n2}(\lambda) & \ldots & c_{nn}(\lambda) \end{pmatrix}$$

a la matriz de orden $(n-1)$, y observemos que cualquier transformación elemental que realicemos en ella no afecta a la primera fila ni a la primera columna de $C(\lambda)$. Como $C'(\lambda)$ es de orden $(n-1)$, por hipótesis de inducción se puede reducir a una matriz canónica:

$$C'(\lambda) = \begin{matrix} e_2(\lambda) & 0 & \ldots & 0 \\ 0 & e_3(\lambda) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & e_n(\lambda) \end{matrix}$$

y, por tanto, $C(\lambda)$ se transforma en una matriz de la forma:

$$D(\lambda) = \begin{pmatrix} e_1(\lambda) & 0 & 0 & \ldots & 0 \\ 0 & e_2(\lambda) & 0 & \ldots & 0 \\ 0 & 0 & e_3(\lambda) & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & e_n(\lambda) \end{pmatrix} \sim A(\lambda).$$

Falta demostrar que esta matriz es canónica.

Sabemos que $e_i(\lambda)$ divide a $e_{i+1}(\lambda)$ para $i=2,3,\ldots,n-1$; es preciso probar que e_1 divide a e_2

El razonamiento es muy similar al que ya vimos. Por el algoritmo de la división, existen polinomios $q(\lambda)$ y $r(\lambda)$ *únicos* tales que

$$e_2(\lambda) = q(\lambda)\; e_1(\lambda) + r(\lambda),$$

donde $r(\lambda)=0$ o grado $r(\lambda) <$ grado $e_1(\lambda)$.

Si $r(\lambda) \neq 0$, entonces grado $r(\lambda) <$ grado $e_1(\lambda)$ y hacemos las transformaciones siguientes: sumamos la primera fila a la se unda fila de la matriz

$D(\lambda)$ y restamos de la segunda columna la primera multiplicada por el polinomio $q(\lambda)$. Aparece, entonces, en lugar de $e_2(\lambda)$ el polinomio

$$e_2(\lambda) - q(\lambda)\ e_1(\lambda) = r(\lambda).$$

Por último, intercambiamos la primera y la segunda filas, y la primera y la segunda columnas.

Hagamos el esquema de las transformaciones:

1ro. $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \to \begin{pmatrix} 1 \\ 1+2 \end{pmatrix}.$

2do. $(1,\ 2) \to (1,\ 2\ - q(\lambda)\,1).$

3ro. $\begin{pmatrix} 1 \\ 2 \end{pmatrix} \to \begin{pmatrix} 2 \\ 1 \end{pmatrix}.$

4to. $(1,\ 2) \to (2,\ 1).$

El resultado es una matriz equivalente a $A(\lambda)$ y que tiene al polinomio $r(\lambda)$, de grado estrictamente menor que el grado de $e_1(\lambda)$, en la primera posición de la primera fila. Esto contradice la forma en que seleccionamos a $e_1(\lambda)$ y, por tanto, $r(\lambda) = 0$. Luego, $e_1(\lambda)$ divide a $e_2(\lambda)$ y la λ-matriz $D(\lambda)$ es una λ-matriz canónica equivalente a la matriz $A(\lambda)$.

El teorema que hemos demostrado prueba la existencia de una λ-matriz canónica para *cada* clase de λ-matrices equivalentes, pero no demuestra que esta λ-matriz canónica sea *única*, lo cual es muy importante. En el próximo epígrafe demostraremos la unicidad de la λ-matriz canónica y daremos un método para calcularla.

4.2 Menores de una λ-matriz

Si analizamos la demostración del teorema 4.1, podemos observar que aunque es una demostración constructiva, el método que plantea no es fácil de llevar a la práctica. Por ese motivo no dimos ningún ejemplo de cálculo de la λ-matriz canónica de una λ-matriz dada. Ahora, al plantear la unicidad de la λ-matriz canónica, vamos a desarrollar nociones que permitirán calcular la λ-matriz canónica de una λ-matriz dada. Estas nociones se utilizan para demostrar lo que constituye el resultado fundamental de este epígrafe:

Toda λ-matriz es equivalente a una única λ-matriz canónica.

Para ello, vamos a introducir la noción de factor invariante.

Sea $A(\lambda)$ una λ-matriz de orden n. Para todo número natural k, $1 \leqslant k \leqslant n$, consideremos todos los menores de orden k de la matriz $A(\lambda)$; estos menores son polinomios en λ. Por ejemplo, los menores de orden 1 son todos los elementos de la matriz $A(\lambda)$, y el menor de orden n es el determinante de la matriz $A(\lambda)$.

El número de los menores de orden k de una matriz de orden n es finito (proponemos al estudiante que lo calcule) y, por tanto, para cada número k obtenemos un conjunto finito de polinomios en λ. Denotemos por $d_k(\lambda)$ el máximo común divisor de los menores de orden k, y este polinomio está unívocamente determinado por la matriz $A(\lambda)$. Así obtenemos un sistema de n polinomios:

$$d_1(\lambda),\ d_2(\lambda),\ \ldots, d_n(\lambda),$$

asociados a la λ-matriz $A(\lambda)$.

Recordemos que estos n polinomios, por ser máximos comunes divisores, son polinomios mónicos, o sea, tienen coeficiente principal 1. Es evidente que si la matriz $A(\lambda)$ tiene rango r, entonces todos los menores de orden mayor que r son nulos y, por tanto,

$$d_{r+1}(\lambda) = d_{r+2}(\lambda) = \ldots = d_n(\lambda) = 0.$$

Además, todos los otros polinomios del sistema son no nulos.

Los polinomios $d_i(\lambda)$ se denominan *divisores elementales.*

Ejemplos

1. Consideremos la λ-matriz de orden 2:

$$A(\lambda) = \begin{pmatrix} \lambda^2+1 & 4 \\ \lambda-1 & \lambda \end{pmatrix}$$

En este caso $d_1(\lambda) = 1$, pues como hay un elemento que es una constante, polinomio de grado 0, entonces el máximo común divisor es 1 y $d_2(\lambda) = \det A(\lambda) = \lambda^2 - 3\lambda + 4$.

2. Consideremos la matriz de orden 3:

$$A(x) = \begin{pmatrix} 13-x & 0 & 22 \\ 4 & 2-x & 8 \\ -6 & 0 & -10-x \end{pmatrix}$$

En este caso, como hay elementos de la matriz de grado 0, $d_1(x) = 1$.

Al calcular los menores de orden 2 obtenemos:

$$-(x-2)(x+10), -4(x-2), 6(x-2), (x-1)(x-2),$$
$$-22(x-2), -8(x-2), (x-2)(x-13), 0.$$

Por tanto, el máximo común divisor es:

$d_2(x) = (x-2)$

$\det(A(x)) = (x-2)^2(1-x)$,

y tenemos:

$d_3(x) = (x-1)(x-2)^2.$

3. Consideremos la λ-matriz de orden 3:

$$A(\lambda)=\begin{pmatrix}3-\lambda & -1 & -1\\ 1 & 1-\lambda & -1\\ 0 & 0 & 1-\lambda\end{pmatrix}$$

Aquí también tenemos:

$d_1(\lambda)=1.$

Calculamos los menores de orden 2 y estos son iguales a los polinomios $(\lambda-1)^2$, $(\lambda-2)^2$, $(\lambda-1)$, $(\lambda-2)$, $(\lambda-1)(\lambda-3)$, 0, o a múltiplos de estos polinomios. Por tanto,

$d_2(\lambda)=1$

$\det A(\lambda)=(1-\lambda)(\lambda-2)^2.$

Luego,

$d_3(\lambda)=(\lambda-1)(\lambda-2)^2.$

Vamos a demostrar que los polinomios

$$d_1(\lambda), d_2(\lambda), \ldots, d_n(\lambda)$$

que se obtienen de la matriz $A(\lambda)$, son los mismos para toda otra matriz equivalente a $A(\lambda)$.

PROPOSICIÓN 4.2

Los polinomios $d_1(\lambda), d_2(\lambda), \ldots, d_n(\lambda)$ son invariantes por transformaciones elementales.

Demostración

Recordemos dos propiedades fundamentales de los determinantes.

(1) Si los elementos de una fila o de una columna de un determinante se multiplican por un escalar, el valor del determinante se multiplica por el mismo escalar.

(2) Si a una fila o a una columna de un determinante se le suma otra fila u otra columna multiplicada por un escalar, el valor del determinante no se altera.

Si tenemos en cuenta que ahora estamos trabajando con determinantes de matrices cuyos elementos son polinomios, las propiedades (1) y (2) son válidas si sustituimos *escalar*, por *polinomio*.

Considerando estas propiedades, vamos a demostrar que los polinomios $d_k(\lambda)$ no se alteran al efectuar transformaciones elementales en la matriz $A(\lambda)$. Trabajaremos solamente con las columnas, pues para las filas se utiliza el mismo razonamiento.

Sea $d_k(\lambda)$ el máximo común divisor de los menores de orden k de $A(\lambda)$.

1) Si multiplicamos la i-ésima columna de $A(\lambda)$ por un escalar $a\in K$ no nulo, los menores de orden k de la nueva matriz en los que no aparece la i-ésima columna, no cambian, y aquellos en los que sí aparece, se multiplican por el escalar $a\in K$. Pero multiplicar polinomios por un escalar

no nulo, no altera el máximo común divisor de estos polinomios; por tanto, $d_k(\lambda)$ no se altera por esta transformación elemental.

2) A la j-ésima columna de $A(\lambda)$ le sumamos la columna i-ésima multiplicada por un polinomio $p(\lambda) \in K[\lambda]$. Así obtenemos una matriz $A'(\lambda)$ semejante a $A(\lambda)$.

Analicemos cómo son los menores de orden k de $A'(\lambda)$. Hay tres casos:

1ro. Si el menor que tomamos en $A'(\lambda)$ no incluye a la j-ésima columna, el menor es igual al menor correspondiente en la matriz $A(\lambda)$.

2do. Si el menor que tomamos en $A'(\lambda)$ incluye las columnas i-ésima y j-ésima, entonces como los elementos de la j-ésima columna de $A'(\lambda)$ se obtienen sumando a la j-ésima columna de $A(\lambda)$ los elementos de la i-ésima columna multiplicados por el polinomio $p(\lambda)$, el menor considerado en $A(\lambda)$ es igual a un menor de orden k en $A(\lambda)$.

3ro. Si el menor que tomamos en $A'(\lambda)$ contiene a la j-ésima columna, pero no a la i-ésima, entonces como todo elemento de esa columna es de la forma

$a_{nj}(\lambda) + p(\lambda) a_{ni}(\lambda)$,

El menor se puede expresar como la suma de dos determinantes, el primero es un menor de $A(\lambda)$ donde aparece la i-ésima columna, llamémoslo M; el segundo es también un menor de $A(\lambda)$, pero en el cual aparece la j-ésima columna multiplicada por el polinomio $p(\lambda)$ y, por tanto, se puede escribir como un menor de $A(\lambda)$ en el que aparece la j-ésima columna, M', multiplicada por el polinomio $p(\lambda)$. En este caso el menor es de la forma

$M + p(\lambda) M'$,

donde M y M' son menores de $A(\lambda)$.

Veamos que $d_k(\lambda)$, máximo común divisor de los menores de orden k, es el mismo para $A(\lambda)$ y $A'(\lambda)$.

Para los menores de $A'(\lambda)$ del primero y segundo tipos, es evidente que no se altera el máximo común divisor. Para los menores del tercer tipo, $d_k(\lambda)$ divide a M y M' y, por tanto, divide a $M + p(\lambda)M'$. Esto quiere decir que $d_k(\lambda)$ divide al máximo común divisor de los menores de orden k de $A'(\lambda)$.

Pero podemos razonar a la inversa. $A(\lambda)$ se obtiene de $A'(\lambda)$ por la transformación elemental inversa a la realizada, y por el mismo razonamiento, el máximo común divisor de los menores de orden k de la matriz $A'(\lambda)$, sea este $d'_k(\lambda)$, divide al máximo común divisor de los menores de orden k de la matriz $A(\lambda)$, $d_k(\lambda)$.

Por tanto, $d_k(\lambda)$ divide a $d'_k(\lambda)$ y $d'_k(\lambda)$ divide a $d_k(\lambda)$, pero como $d_k(\lambda)$ y $d'_k(\lambda)$ son polinomios mónicos, por ser máximos comunes divisores, entonces

$$d_k(\lambda) = d'_k(\lambda),$$

y esto completa la demostración.

De la proposición 4.2 resulta que todas las λ-matrices de una misma clase son equivalentes entre sí. Ya hemos logrado un elemento que las caracteriza: los polinomios

$$d_1(\lambda),\ d_2(\lambda),\dots,d_n(\lambda).$$

Además, sabemos que en cada clase hay al menos una λ-matriz canónica. Vamos, entonces, a estudiar cómo son los polinomios $d_k(\lambda)$ para una λ-matriz canónica.

PROPOSICIÓN 4.3

Sea $C(\lambda)$ una λ-matriz canónica:

$$C(\lambda)=\begin{pmatrix} e_i(\lambda) & 0 & \dots & 0 \\ 0 & e_2(\lambda) & \dots & 0 \\ . & . & . & . \\ 0 & 0 & \dots & e_n(\lambda) \end{pmatrix}$$

Entonces,

$$d_k(\lambda)=e_1(\lambda)\,e_2(\lambda),\dots,e_k(\lambda),\quad k=1,2,\dots,n.$$

Demostración

Calculemos los menores de orden k de la matriz $C(\lambda)$.

Los menores de orden k se forman tomando los elementos de k filas y k columnas de $C(\lambda)$ y calculando el valor del determinante que se obtiene.

Veamos las posibilidades que tenemos para formar los menores.

Si formamos un menor escogiendo la i-ésima fila y no la i-ésima columna, como la matriz $C(\lambda)$ es diagonal, ese menor tendrá una fila entera de ceros y, por tanto, su valor será cero. Estos menores no aportan nada para el máximo común divisor. Veamos, entonces, los menores que se obtienen seleccionando las filas $i_1<i_2<\dots<i_k$ y las columnas $i_1<i_2<\dots<i_k$.

Este tipo de menor tiene todos sus elementos nulos, excepto los que están en la diagonal, que son $e_{i_1}, e_{i_2},\dots,e_{i_k}$. Por tanto, el valor del menor es $e_{i_1}e_{i_2}\dots e_{i_k}$.

Como $1\leqslant i_1<i_2<\dots<i_k\leqslant n$ y por ser $C(\lambda)$ una λ-matriz canónica se cumple que e_i divide a e_{i+1} y, por tanto, divide a todo e_j tal que $i<j$, tenemos:

e_1 divide a e_{i_1}

e_2 divide a e_{i_2}

e_k divide a e_{i_k}

y el producto $e_1e_2\dots e_k$ divide a $e_{i_1}e_{i_2}\dots e_{i_k}$.

Si tenemos en cuenta que el menor que contiene a las k primeras filas y las k primeras columnas tiene el valor

$$e_1e_2\dots e_k,$$

llegamos a la conclusión de que

$$d_k=e_1e_2\dots e_k,\quad k=1,2,\dots,\ n.$$

A partir de estos resultados, podemos demostrar que la λ-matriz canónica asociada a una λ-matriz dada es única, pues como veremos, los elementos $e_k(\lambda)$ están determinados por los $d_k(\lambda)$.

TEOREMA 4.2

La λ-matriz canónica equivalente a una λ-matriz dada es *única*.

Demostración

Sea $A(\lambda)$ una λ-matriz y sean

$$d_1(\lambda), d_2(\lambda), \ldots, d_n(\lambda)$$

los máximos comunes divisores de sus menores.

Sea $C(\lambda)$ una λ-matriz canónica equivalente a $A(\lambda)$ y sean

$$e_1(\lambda), e_2(\lambda), \ldots, e_n(\lambda)$$

los polinomios de la diagonal principal de $C(\lambda)$.

Por la proposición 4.3 sabemos que

$$d_k(\lambda) = e_1(\lambda)\, e_2(\lambda) \ldots e_k(\lambda).$$

Si $A(\lambda)$ es de rango r, entonces $d_r(\lambda) \neq 0$. Pero $d_{r+1} = 0$; por tanto, $e_1(\lambda), e_2(\lambda), \ldots, e_r(\lambda)$ son no nulos y

$$e_{r+1}(\lambda) = 0.$$

Por las propiedades de la λ-matriz canónica $C(\lambda)$, tenemos:

$$e_{r+1}(\lambda) = e_{r+2}(\lambda) = \ldots = e_n(\lambda) = 0.$$

Además, $d_1(\lambda), \ldots, d_r(\lambda)$ son no nulos y se cumplen las relaciones:

$$d_k(\lambda) = e_1(\lambda)\, e_2(\lambda) \ldots e_{k-1}(\lambda)\, e_k(\lambda) \neq 0$$
$$d_{r-1}(\lambda) = e_1(\lambda)\, e_2(\lambda) \ldots e_{r-1}(\lambda) \neq 0$$

para $1 \leqslant k \leqslant r$.

Obtenemos así:

$$e_k(\lambda) = \frac{d_k(\lambda)}{d_{k-1}(\lambda)}, \quad k = 2, 3, \ldots, r$$

$$e_1(\lambda) = d_1(\lambda).$$

Como los polinomios $d_1(\lambda)$, $d_2(\lambda), \ldots, d_n(\lambda)$ son únicos para cada clase de λ-matrices equivalentes, las relaciones

$$e_1(\lambda) = d_1(\lambda)$$

$$e_k(\lambda) = \frac{d_k(\lambda)}{d_{k-1}(\lambda)}, \quad k = 2, 3, \ldots, r.$$

$$e_{r+1}(\lambda) = e_n(\lambda) = 0,$$

donde r es el rango de $A(\lambda)$, determinan en forma única la λ-matriz canónica en cada clase.

Como los polinomios $e_1(\lambda), e_2(\lambda), \ldots, e_n(\lambda)$ caracterizan a la λ-matriz canónica, se denominan los *factores invariantes* de la λ-matriz $A(\lambda)$.

Ejemplos

1. Para la λ-matriz de orden 2:

$$A(\lambda) = \begin{pmatrix} \lambda_2+1 & 4 \\ \lambda-1 & \lambda \end{pmatrix}$$

teníamos que $d_1(\lambda) = 1$, $d_2(\lambda) = \lambda^2 - 3\lambda + 4$.
Por tanto, la λ-matriz canónica asociada a ella es

$$\begin{pmatrix} 1 & 0 \\ 0 & \lambda^2-3\lambda+4 \end{pmatrix}.$$

2. Para la matriz de orden 3:

$$A(x) = \begin{pmatrix} 13-x & 0 & 22 \\ 4 & 2-x & 8 \\ -6 & 0 & -10-x \end{pmatrix}$$

teníamos que
$d_1(x) = 1$, $d_2(x) = (x-2)$, $d_3(x) = (x-1)(x-2)^2$.
Por tanto,

$$e_1(x) = d_1(x) = 1, \; e_2(x) = \frac{d_2(x)}{d_1(x)} = (x-2),$$

$$e_3(x) = \frac{d_3(x)}{d_2(x)} = (x-1)(x-2),$$

y la λ-matriz canónica equivalente es

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & (x-2) & 0 \\ 0 & 0 & (x-1)(x-2) \end{pmatrix}$$

3. Consideremos la λ-matriz de orden 3:

$$A(\lambda) = \begin{pmatrix} 3-\lambda & -1 & -1 \\ 1 & 1-\lambda & -1 \\ 0 & 0 & 1-\lambda \end{pmatrix}$$

Aquí teníamos:
$d_1(\lambda) = 1$, $d_2(\lambda) = 1$, $d_3(\lambda) = (\lambda-1)(\lambda-2)^2$.
Por tanto,
$e_1(\lambda) = 1$, $e_2(\lambda) = 1$, $e_3(\lambda) = (\lambda-1)(\lambda-2)^2$

y la λ-matriz canónica equivalente es

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & (\lambda-1)(\lambda-2)^2 \end{pmatrix}$$

Le planteamos al estudiante como ejercicio, que trate en cada uno de estos ejemplos de pasar de la matriz $A(\lambda)$ a la matriz canónica mediante transformaciones elementales.

A partir de la unicidad de la λ-matriz canónica asociada a una λ-matriz, podemos dar un nuevo criterio para saber cuándo dos λ-matrices son equivalentes.

TEOREMA 4.3

Sean $A(\lambda)$ y $B(\lambda)$ dos matrices de orden n. $A(\lambda)$ y $B(\lambda)$ son equivalentes si y solo si el máximo común divisor de los menores de orden k de la matriz $A(\lambda)$, coincide con el máximo común divisor de los menores de orden k de la matriz $B(\lambda)$, para todos los órdenes $k=1, 2,..., n$.

Demostración

Si $A(\lambda)$ y $B(\lambda)$ son equivalentes, por la proposición 4.2 sabemos que los polinomios $d_1(\lambda), d_2(\lambda),...,d_r(\lambda)$ son los mismos para las dos, pues son invariantes por transformaciones elementales y, por tanto, por equivalencia.

Recíprocamente si $A(\lambda)$ y $B(\lambda)$ tienen los mismos máximos comunes divisores de los menores de orden k, para todos los órdenes $k=1,2,...,n$, entonces por el teorema 4.2, las dos tienen la misma λ-matriz canónica $C(\lambda)$, o sea, $A(\lambda)\sim C(\lambda)$ y $B(\lambda)\sim C(\lambda)$, por las propiedades de la equivalencia de matrices,

$$A(\lambda)\sim C(\lambda),\ C(\lambda)\sim B(\lambda) \Rightarrow A(\lambda)\sim B(\lambda),$$

lo cual demuestra el criterio dado.

Este criterio, muy práctico para la equivalencia de λ-matrices, es aún insuficiente para nuestros requerimientos.

Probablemente, el estudiante se preguntará qué tiene esto que ver con endomorfismos y matrices más simples, aunque si analiza bien los ejemplos 2 y 3 observará que hay relaciones conocidas. Aún tenemos que trabajar un poco más para llegar a nuestro objetivo; por el momento, vamos a buscar otra forma de expresar la equivalencia de λ-matrices.

4.3 Otro criterio de equivalencia de *λ-matrices*

Hasta ahora tenemos las siguientes formas de expresar la equivalencia de λ-matrices:

1) Dos λ-matrices, $A(\lambda)$ y $B(\lambda)$, son equivalentes si y solo si una se puede obtener de la otra mediante transformaciones elementales.

2) Dos λ-matrices, $A(\lambda)$ y $B(\lambda)$, son equivalentes si y solo si tienen la misma λ-matriz canónica.

3) Dos λ-matrices, $A(\lambda)$ y $B(\lambda)$, son equivalentes si y solo si los divisores elementales de las dos matrices, $d_k(\lambda)$, son iguales para todo los órdenes $k=1, 2, \ldots, n$.

Sin embargo, aunque las formas 2) y 3) brindan criterios muy prácticos para determinar si dos λ-matrices dadas son equivalentes, no tenemos ningún índice para relacionar esto con el problema planteado: dada una matriz, encontrar otra semejante a ella, más simple. Es por esto que trataremos de expresar la equivalencia de λ-matrices en una nueva forma, que va a permitir relacionar la equivalencia de λ-matrices con la semejanza de matrices.

Vamos a expresar las transformaciones elementales en términos de producto de matrices. Analicemos el caso de λ matrices de orden 2.

Sea

$$A(\lambda) = \begin{pmatrix} a(\lambda) & b(\lambda) \\ c(\lambda) & d(\lambda) \end{pmatrix}.$$

Supongamos que queremos multiplicar la primera fila, o la primera columna por el escalar α. Si tomamos la matriz

$$B = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix}$$

y efectuamos los productos $BA(\lambda)$ y $A(\lambda)B$ obtenemos:

$$BA(\lambda) = \begin{pmatrix} \alpha a(\lambda) & \alpha b(\lambda) \\ c(\lambda) & d(\lambda) \end{pmatrix}.$$

$$A(\lambda)B = \begin{pmatrix} \alpha a(\lambda) & b(\lambda) \\ \alpha c(\lambda) & d(\lambda) \end{pmatrix}.$$

La matriz B sirve para realizar la transformación elemental de multiplicar la primera fila o la primera columna de $A(\lambda)$ por el escalar α.

De manera similar, la matriz

$$B' = \begin{pmatrix} 1 & 0 \\ 0 & \alpha \end{pmatrix}$$

se utiliza para realizar el mismo tipo de transformación elemental, pero con la segunda fila y la segunda columna.

Supongamos que queremos transformar $A(\lambda)$ sumándole a la segunda fila la primera multiplicada por el polinomio $p(\lambda)$. Para esto tomamos la matriz

$$P(\lambda) = \begin{pmatrix} 1 & 0 \\ p(\lambda) & 0 \end{pmatrix}$$

y efectuamos el producto:

$$P(\lambda)\ A(\lambda) = \begin{pmatrix} a(\lambda) & b(\lambda) \\ a(\lambda)\ p(\lambda) + c(\lambda) & b(\lambda)\ p(\lambda) + d(\lambda) \end{pmatrix}$$

Dejamos al estudiante el cálculo de $P(\lambda)A(\lambda)$.

Con la matriz $P(\lambda)$ se realiza la transformación elemental de sumar a la segunda fila la primera multiplicada por $p(\lambda)$, o la transformación de sumar a la primera columna la segunda columna multiplicada por $p(\lambda)$, en dependencia del orden en que se tome el producto.

El análisis de este caso nos lleva a considerar las transformaciones elementales de λ-matrices como productos de matrices.

Nota. Al estudiar el método de Gauss para resolver sistemas de ecuaciones lineales, definíamos también transformaciones elementales; en ese caso solamente para las filas de la matriz. Mediante esas transformaciones elementales, convertíamos la matriz del sistema en una matriz escalonada, que pertenecía a un sistema equivalente. Esas transformaciones también se pueden expresar mediante productos de matrices.

Consideremos el caso de orden n.

El primer tipo de matrices es el de la forma:

$$U_i(a) = \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & a & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 \end{pmatrix} \begin{matrix} \\ \\ \\ \leftarrow i\text{-ésima fila} \\ \\ \\ \end{matrix}$$

la cual difiere de la matriz identidad solamente en el elemento de la i-ésima fila y la i-ésima columna; este elemento es un escalar: $a \neq 0$.

Este tipo de matrices sirve para realizar las transformaciones elementales de multiplicar la i-ésima fila o la i-ésima columna por el escalar $a \neq 0$.

$U_i(a)\ A(\lambda)$ es la matriz que se obtiene de $A(\lambda)$ multiplicando su i-ésima fila por a.

$A(\lambda)\ U_i(a)$ es la matriz que se obtiene de $A(\lambda)$ multiplicando su i-ésima columna por a.

El segundo tipo de matriz es de la forma:

$$\begin{array}{cc} V_{ij}\ (p(\lambda)) = \begin{pmatrix} 1 \dots 0 \dots 0 & \dots 0 \\ \dots\dots\dots\dots & \dots \\ 0 \dots 1 \dots p(\lambda) & \dots 0 \\ \dots\dots\dots\dots & \dots \\ 0 \dots 0 \dots 1 & \dots 0 \\ \dots\dots\dots\dots & \dots \\ 0 \dots 0 \dots 0 & \dots 1 \end{pmatrix} & \begin{matrix} \\ \\ \leftarrow i\text{-ésima fila} \\ \\ \\ \\ \\ \end{matrix} \\ \uparrow & \\ j\text{-ésima columna} & \end{array}$$

la cual se obtiene a partir de la matriz identidad, cambiando el cero que está en la posición (i, j) por el polinomio $p(\lambda)$.

Este tipo de matriz sirve para realizar las otras dos transformaciones elementales.

$V_{ij}(p(\lambda))\ A(\lambda)$ es la matriz que se obtiene de $A(\lambda)$ sumando a la i-ésima fila, la j-ésima fila multiplicada por el polinomio $p(\lambda)$.

$A(\lambda)\ V_{ij}(p(\lambda))$ es la matriz que se obtiene de $A(\lambda)$ sumando a la j-ésima columna, la i-ésima columna multiplicada por el polinomio $p(\lambda)$.

Las matrices $U_i(a)$ y $V_{ij}(p(\lambda))$ se denominan *λ-matrices elementales* y tienen una característica muy interesante: sus determinantes son escalares:

$\det [U_i(a)] = a$, $\det [V_{ij}\ (p(\lambda)] = 1$.

Las λ-matrices que cumplen esta propiedad reciben el nombre de λ-matrices *unimodulares*.

DEFINICIÓN 4.5

Se dice que una λ-matriz es *unimodular* si su determinante es un escalar.

Ejemplos

1. $U_i(\lambda)$ y $V_{ij}\ (p(\lambda))$ son unimodulares.
2. Las matrices escalares son unimodulares.

Es evidente a partir de la definición y de las propiedades de los determinantes, que *el producto de dos λ-matrices unimodulares es una λ-matriz unimodular*. Además, en general, una λ-matriz de determinante no nulo no es inversible, pues su determinante es un polinomio y al dividir los cofactores por el determinante, el resultado, en general, será una fracción racional y no un polinomio y, por tanto, la "inversa" no es una λ-matriz. Sin embargo, si la matriz es unimodular, como el determinante es un número (no nulo), es posible calcular su inversa, y esta será otra λ-matriz unimodular. Por tanto, podemos enunciar la proposición siguiente:

PROPOSICIÓN 4.4

Si $A(\lambda)$ es una λ-matriz unimodular de determinante no nulo, entonces existe una λ-matriz unimodular que es la inversa de $A(\lambda)$.

Las matrices del tipo $U_i(\lambda)$ y $V_{ij}\ (p(\lambda))$ son unimodulares e inversibles; por tanto, cualquier producto de ellas será también unimodular e inversible. Podemos enunciar ahora la nueva caracterización de la equivalencia de λ- matrices.

TEOREMA 4.4

Dos λ-matrices, $A(\lambda)$ y $B(\lambda)$, son equivalentes si y solo si existen dos matrices unimodulares P y Q de determinante no nulo, tales que

$B(\lambda) = PA(\lambda)Q$.

Demostración

Si $A(\lambda)$ y $B(\lambda)$ son equivalentes, esto quiere decir que $B(\lambda)$ se puede obtener a partir de $A(\lambda)$ por transformaciones elementales. Por tanto, existen matrices del tipo $U_i(a)$ y $V_{ij}(p(\lambda))$ tales que premultiplicando a $A(\lambda)$ y postmultiplicando a $A(\lambda)$ por dichas matrices, obtenemos $B(\lambda)$.

Sean $X_1(\lambda), X_2(\lambda), \ldots, X_k(\lambda)$ las matrices que premultiplican a $A(\lambda)$, y $Y_1(\lambda), Y_2(\lambda), \ldots, Y_m(\lambda)$ las matrices que postmultiplican a $A(\lambda)$.

Cada $X_i(\lambda)$, $i=1, 2, \ldots, k$, y cada $Y_j(\lambda)$, $j=1, 2, \ldots, m$, es de uno de los tipos U o V, por tanto, son unimodulares de determinante no nulo. Así:

$$B(\lambda) = X_1(\lambda) \ldots X_k(\lambda) \cdot A(\lambda) \cdot Y_1(\lambda) \ldots Y_m(\lambda)$$

Si denotamos

$$P = X_1(\lambda) X_2(\lambda) \ldots X_k(\lambda),$$
$$Q = Y_1(\lambda) Y_2(\lambda) \ldots Y_m(\lambda),$$

P y Q son unimodulares por ser producto de matrices unimodulares y son inversibles por ser producto de matrices inversibles, y tenemos:

$$B(\lambda) = PA(\lambda)Q.$$

Supongamos ahora que $A(\lambda)$ y $B(\lambda)$ son dos λ-matrices tales que existen dos λ-matrices unimodulares inversibles P y Q tales que $B(\lambda) = PA(\lambda)Q$. Demostremos que $A(\lambda)$ y $B(\lambda)$ son equivalentes.

Este resultado lo vamos a obtener a partir del lema siguiente:

LEMA 4.1

Toda λ-matriz unimodular inversible de orden n es equivalente a la matriz unidad I_n.

Demostración

Sea P una λ-matriz unimodular inversible de orden n. Entonces

$$\det P = a \neq 0, \quad a \in K.$$

Como $\det P = a$, el polinomio $d_n(\lambda)$, máximo común divisor de los menores de orden n (hay uno solo), es igual a 1, o sea,

$$d_n(\lambda) = 1.$$

Pero

$$d_n(\lambda) = e_1(\lambda) \cdot e_2(\lambda) \ldots e_n(\lambda) = 1,$$

donde $e_1(\lambda), e_2(\lambda), \ldots, e_n(\lambda)$ son los factores invariantes de P. Como su producto es la unidad, entonces

$$e_1(\lambda) = e_2(\lambda) = \ldots = e_n(\lambda) = 1,$$

y la matriz canónica de la λ-matriz unimodular inversible P, es la matriz unidad.

Por tanto, P es equivalente a la matriz unidad.

Sobre la base de este resultado podemos hacer el razonamiento siguiente:
P es una λ-matriz unimodular inversible, por tanto, es equivalente a la matriz unidad. Esto significa que podemos realizar transformaciones elementales en la matriz unidad, hasta obtener la matriz P. Estas transformaciones elementales las podemos expresar como el producto de las matrices $X_1(\lambda),\dots,X_k(\lambda)$ y $Y_1(\lambda),\dots,Y_m(\lambda)$ del tipo elemental, en la forma:

$$
\begin{aligned}
P &= X_1(\lambda)\dots X_k(\lambda)\, I_n\, Y_1(\lambda)\dots Y_m(\lambda)\\
&= X_1(\lambda)\dots X_k(\lambda)\;\; Y_1(\lambda)\dots Y_m(\lambda).
\end{aligned}
$$

Podemos formular esta propiedad como otro lema.

LEMA 4.2

Toda λ-matriz unimodular inversible se puede expresar como un producto de λ-matrices elementales.

Para terminar la demostración de que $A(\lambda) \sim B(\lambda)$, si

$$B(\lambda) = PA(\lambda)Q,$$

entonces P se descompone como producto de λ-matrices elementales y Q también; por tanto, estamos premultiplicando y postmultiplicando a $A(\lambda)$ por λ-matrices elementales, y el resultado es $B(\lambda)$. Esto significa que realizando transformaciones elementales en $A(\lambda)$, se obtiene $B(\lambda)$ o, lo que es lo mismo, $A(\lambda)$ y $B(\lambda)$ son equivalentes.

Con esta forma de interpretar la equivalencia de λ-matrices, el concepto de λ-matriz canónica, los factores invariantes y la forma de calcularlos mediante los máximos comunes divisores de los menores, estamos en condiciones de resolver el problema de encontrar la forma más simple de representar un endomorfismo.

Lo que hemos hecho hasta ahora para resolver el problema planteado puede parecer muy trabajoso y poco intuitivo, pero insistimos en que forma parte de una teoría algebraica más general, que resuelve también otros problemas.

Veamos primero la relación entre las matrices semejantes y las λ-matrices equivalentes.

4.4 Matrices semejantes y *λ-matrices equivalentes*

Con los elementos dados hasta aquí, vamos a resolver el problema siguiente:

Dadas dos matrices A y B de $M_n(K)$, ¿estas son semejantes o no son semejantes?

En otras palabras, ¿existe una matriz inversible P tal que $B = P^{-1}AP$?

Al inicio del capítulo dijimos que este era un problema en general muy difícil, pero con los elementos que tenemos, ya lo podemos resolver.

¿Cómo utilizaremos las λ-matrices para resolverlo? Como sabemos, a cada matriz cuadrada A está asociada una λ-matriz que ya hemos empleado, la λ-matriz cuyo determinante es su polinomio característico, esta es

$$A-\lambda I_n.$$

Vamos, entonces, a enunciar y demostrar el teorema fundamental para la semejanza de matrices.

TEOREMA 4.5

Sean A y B dos matrices de orden n con coeficientes en K. A y B son semejantes si y solo si sus λ-matrices características $(A-\lambda I_n)$ y $(B-\lambda I_n)$ son equivalentes.

Demostración

1) Supongamos que A y B son semejantes. Esto significa que existe una matriz inversible P tal que
   $B=P^{-1}AP.$
   Entonces, para las matrices $B-\lambda I_n$ y $A-\lambda I_n$, tenemos:
   $$\begin{aligned} B-\lambda I_n &= P^{-1}AP-\lambda I_n = P^{-1}AP-\lambda P^{-1}\,P \\ &= P^{-1}AP-P^{-1}(\lambda I_n)P \\ &= P^{-1}(A-\lambda I_n)P \end{aligned}$$
   y P es una matriz escalar inversible, por tanto, es un caso particular de λ-matriz unimodular inversible. Luego, por el teorema 4.4, las λ-matrices $(B-\lambda I_n)$ y $(A-\lambda I_n)$ son equivalentes.
2) Supongamos que $(B-\lambda I_n)$ y $(A-\lambda I_n)$ son equivalentes. Tenemos que encontrar una matriz escalar inversible tal que
   $P^{-1}AP=B.$

Por la equivalencia de $(B-\lambda I_n)$ y $(A-\lambda I_n)$, sabemos que existen dos λ-matrices unimodulares inversibles P y Q tales que

$$\begin{aligned} B-\lambda I_n &= P(A-\lambda I_n)Q \\ P^{-1}(B-\lambda I_n) &= (A-\lambda I_n)\,Q \\ (B-\lambda I_n)Q^{-1} &= P(A-\lambda I_n). \end{aligned}$$

Las expresiones $A-\lambda I_n$ y $B-\lambda I_n$ son polinomios en la indeterminada λ, cuyos coeficientes son matrices. Con ellos podemos operar en forma análoga a como lo hacemos con los polinomios aunque hay que prestar atención a la conmutatividad y a la división por matrices no inversibles. En particular, nos interesa aplicar el algoritmo de la división con resto.

Vamos a aplicar este algoritmo al par P y $B-\lambda I_n$ y al par Q y $B-\lambda I_n$. Así obtenemos matrices T_1, T_2 y R_1, R_2 tales que

$$\begin{aligned} P &= (B-\lambda I_n)\,T_1+R_1, \\ Q &= T_2(B-\lambda I_n)+R_2, \end{aligned}$$

donde R_1 y R_2 o son matrices nulas o son λ-matrices que tienen elementos de grado menor que el grado de $(B-\lambda I_n)$ y $(A-\lambda I_n)$, respectivamente. Estas tienen grado 1: por tanto, R_1 y R_2 son matrices escalares (en este caso no pueden ser nulas).

Empleando las relaciones establecidas, obtenemos:

$$
\begin{aligned}
&R_1=P-(B-\lambda I_n)\ T_1\\
&R_2=Q-T_2\ (B-\lambda I_n)\\
&R_1(A-\lambda I_n)R_2=[P-(B-\lambda I_n)\ T_1]\ (A-\lambda I_n)\ [Q-T_2\ (B-\lambda I_n)]\\
&\qquad =P(A-\lambda I_n)Q-(B-\lambda I_n)\ T_1(A-\lambda I_n)\ Q\\
&\qquad -P(A-\lambda I_n)T_2\ (B-\lambda I_n)+\\
&\qquad +(B-\lambda I_n)\ T_1\ (A-\lambda I_n)\ T_2\ (B-\lambda I_n)\\
&\qquad =(B-\lambda I_n)-(B-\lambda I_n)\ T_1\ P^{-1}\ (B-\lambda I_n)\\
&\qquad -(B-\lambda I_n)\ Q^{-1}\ T_2\ (B-\lambda I_n)+\\
&\qquad +(B-\lambda I_n)T_1\ (A-\lambda I_n)\ T_2\ (B-\lambda I_n),
\end{aligned}
$$

de donde

$$R_1\ (A-\lambda I_n)R_2=(B-\lambda I_n)\ \{I_n-[T_1P^{-1}+Q^{-1}\ T_2+T_1(A-\lambda I_n)T_2](B-\lambda I_n)\}.$$

La expresión del primer miembro de la igualdad tiene grado 1, o es igual a 0; si la expresión entre corchetes del segundo miembro es no nula, la expresión total de este miembro tiene por lo menos grado 2. Esto no es posible; por tanto, la expresión entre corchetes es nula y tenemos:

$$
\begin{aligned}
&R_1(A-\lambda I_n)\ R_2=(B-\lambda I_n)\\
&R_1AR_2-\lambda R_1R_2=B-\lambda I_n,
\end{aligned}
$$

de donde, igualando coeficientes según las potencias de λ:

$$
\begin{aligned}
&R_1AR_2=B,\\
&R_1R_2=I_n.
\end{aligned}
$$

R_2 es una matriz escalar inversible y R_1 es su inversa; por tanto,

$$B=R_2^{-1}\ A\ R_2.$$

Luego, A y B son semejantes.

Ejemplo

Las matrices

$$A=\begin{pmatrix}13 & 0 & 22\\ 4 & 2 & 8\\ -6 & 0 & -10\end{pmatrix} \quad \text{y} \quad B=\begin{pmatrix}3 & -1 & -1\\ 1 & 1 & -1\\ 0 & 0 & 1\end{pmatrix}$$

tienen los mismos polinomios característicos; sin embargo, no son semejantes pues, como vimos en ejemplos anteriores,

$$(A-\lambda I_n)\sim\begin{pmatrix}1 & 0 & 0\\ 0 & (\lambda-2) & 0\\ 0 & 0 & (\lambda-1)(\lambda-2)\end{pmatrix}$$

$$(B-\lambda I_n)\sim\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & (\lambda-1)(\lambda-2)^2\end{pmatrix}$$

Esto significa que las λ-matrices canónicas de cada una de ellas son diferentes. Por tanto, $(A-\lambda I_n)$ y $(B-\lambda I_n)$ no son λ-matrices equivalentes y esto significa que A y B no son semejantes.

Ya podemos determinar cuándo dos matrices son semejantes, pero aún falta buscar la matriz más simple semejante a una matriz dada A. Este problema quedará resuelto en el próximo epígrafe.

4.5 Matriz de Jordán

Vamos a analizar de nuevo el problema siguiente:

Dado un endomorfismo T de un espacio vectorial E, hay que encontrar una base de E en la cual la representación matricial de T sea la más simple.

Como ya hemos dicho, por "más simple" entendemos con la mayor cantidad de ceros. Vamos, pues, a introducir un tipo de matriz, que es muy simple y que denominamos *matriz de Jordán.*

DEFINICIÓN 4.5

Sea $a \in K$. Se denomina *celda de Jordán* de orden n asociada al escalar a, a una matriz de orden n de la forma:

$$\begin{pmatrix} a & 1 & & & 0 \\ & a & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & a \end{pmatrix}$$

Esta es una matriz de orden n en cuya diagonal principal aparece el escalar a, en las posiciones inmediatamente sobre la diagonal principal aparece el número 1, y el resto de los elementos de la matriz son cero.

Ejemplo

$$(\lambda), \quad \begin{pmatrix} \lambda & 1 \\ 0 & \lambda \end{pmatrix} \begin{pmatrix} \lambda & 1 & 0 \\ 0 & \lambda & 1 \\ 0 & 0 & \lambda \end{pmatrix}$$

son celdas de Jordán asociadas al escalar λ, de orden 1, 2 y 3, respectivamente.

DEFINICIÓN 4.7

Una *matriz de Jordán* de orden n, es una matriz que tiene la forma:

$$\begin{pmatrix} J_1 & & & 0 \\ & J_2 & & \\ & & \ddots & \\ 0 & & & J_k \end{pmatrix}$$

Esta es una matriz diagonal por bloques, en la cual los bloques de la diagonal, $J_1, J_2, \ldots, J_k$, son celdas de Jordán correspondientes a distintos escalares, no necesariamente todas diferentes entre sí.

Ejemplos

1. Toda matriz diagonal es de la forma de Jordán. Si la matriz es de orden n, tiene n celdas de Jordán de orden 1.

2. $\begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

   es una matriz de Jordán, formada por tres celdas de Jordán:

   $\begin{pmatrix} 2 & 1 \\ 0 & 2 \end{pmatrix}$, (1), (1).

Como cada celda de Jordán es una matriz triangular y la matriz de Jordán es triangular por bloques, dicha matriz es triangular, y por tanto, en la diagonal principal aparecen los valores propios de la matriz; podemos, entonces, enunciar el lema siguiente:

LEMA 4.3

En la diagonal principal de una matriz de Jordán aparecen los valores propios de la matriz.

A partir de este resultado podemos obtener la conclusión siguiente:

Para que una matriz A sea semejante a una matriz de Jordán J, es necesario que su polinomio característico tenga todas las raíces en K.

A continuación veremos que esta es una condición no solo necesaria, sino también suficiente.

Para ello es de esperar, aunque no sepamos realmente por qué, la aplicación de lo que hemos desarrollado sobre λ-matrices. Vamos a comenzar estudiando los divisores elementales y los factores invariantes asociados a una matriz de Jordán. Primero lo haremos para las celdas de Jordán.

LEMA 4.4

Sea J una celda de Jordán de orden k, asociada al escalar $a \in K$:

$$J=\begin{pmatrix} a & 1 & & & 0 \\ & a & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & a \end{pmatrix}$$

Entonces para $J-\lambda I_k$ tenemos:

$$1=d_1(\lambda)=d_2(\lambda)=\ldots=d_{k-1}(\lambda);\ d_k(\lambda)=(\lambda-a)^k$$
$$1=e_1(\lambda)=e_2(\lambda)=\ldots=e_{k-1}(\lambda);\ e_k(\lambda)=(\lambda-a)^k$$

Demostración

Como

$$J-\lambda I_k=\begin{pmatrix} (a-\lambda) & 1 & & & 0 \\ & (a-\lambda) & 1 & & \\ & & \ddots & \ddots & \\ & & & \ddots & 1 \\ 0 & & & & (a-\lambda) \end{pmatrix}$$

tiene forma triangular, es evidente que

$$(-1)^k \det(J-\lambda I_k)=d_k(\lambda)=(\lambda-a)^k.$$

Al analizar los menores de orden $k-1$, si consideramos el menor que se obtiene al suprimir la primera columna y la última fila, vemos que tiene la forma triangular inferior:

$$\begin{pmatrix} 1 & & & 0 \\ (a-\lambda) & 1 & & \\ & \ddots & \ddots & \\ 0 & & (a-\lambda) & 1 \end{pmatrix}$$

y, por tanto, es igual a 1, lo cual implica que

$$d_{k-1}(\lambda)=1.$$

También obtenemos que

$$d_1(\lambda)=d_2(\lambda)=\ldots=d_{k-1}(\lambda)=1.$$

Como conclusión resulta que la forma canónica de la λ-matriz $J-\lambda I_k$ es

$$\begin{pmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & (\lambda-a)^k \end{pmatrix}$$

Vamos a ilustrar mediante un ejemplo, el procedimiento general de trabajo con una matriz de Jordán.

Ejemplo

Sea J una matriz de Jordán de orden 10 de la forma

$$J=\begin{pmatrix}
\boxed{\begin{matrix}3&1&0\\0&3&1\\0&0&3\end{matrix}} & & & & & \text{0}\\
& \boxed{\begin{matrix}3&1\\0&3\end{matrix}} & & & & \\
& & \boxed{\begin{matrix}1&1\\0&1\end{matrix}} & & & \\
& & & \boxed{\begin{matrix}1&1\\0&1\end{matrix}} & & \\
\text{0} & & & & \boxed{3} &
\end{pmatrix}$$

Podemos realizar transformaciones elementales en $(J-\lambda I_{10})$ de forma tal, que cada bloque se reduzca a su forma canónica:

$$J-\lambda I_{10}\sim\begin{pmatrix}
1 & & & & & & & & & 0\\
& 1 & & & & & & & & \\
& & (\lambda-3)^3 & & & & & & & \\
& & & 1 & & & & & & \\
& & & & (\lambda-3)^2 & & & & & \\
& & & & & 1 & & & & \\
& & & & & & (\lambda-1)^2 & & & \\
& & & & & & & 1 & & \\
& & & & & & & & (\lambda-1)^2 & \\
0 & & & & & & & & & (\lambda-3)
\end{pmatrix}$$

Reordenando las filas y las columnas obtenemos:

$$J-\lambda I_{10}\sim\begin{pmatrix}
1 & & & & & & & & & 0\\
& 1 & & & & & & & & \\
& & 1 & & & & & & & \\
& & & 1 & & & & & & \\
& & & & 1 & & & & & \\
& & & & & (\lambda-3) & & & & \\
& & & & & & (\lambda-3)^2 & & & \\
& & & & & & & (\lambda-3)^3 & & \\
& & & & & & & & (\lambda-1)^2 & \\
0 & & & & & & & & & (\lambda-1)^2
\end{pmatrix}$$

A partir de aquí se pueden calcular los divisores elementales.

$$d_1(\lambda)=d_2(\lambda)=d_3(\lambda)=d_4(\lambda)=d_5(\lambda)=1,$$

porque en todo menor de orden menor o igual que 5, hay uno que es constante e igual a 1.

Al buscar los menores de orden 6, en cada uno de ellos, no nulo, aparece uno de los elementos de la diagonal; por tanto, $d_6(\lambda)$ es el máximo común divisor.

$$d_6(\lambda)=1.$$

Para calcular $d_7(\lambda)$, tomamos menores de orden 7 no nulos, que tienen que formarse con los elementos que están en la diagonal, y los de menor grado serán aquellos que incluyen solamente dos de los elementos no constantes de la diagonal. En este caso hay uno igual a $(\lambda-1)^2(\lambda-1)^2=(\lambda-1)^4$ y otro igual a $(\lambda-3)(\lambda-3)^2=(\lambda-3)^3$, que son primos entre sí y, por tanto,

$$d_7(\lambda)=1.$$

Los menores de orden 8 tienen que incluir al menos a tres de los factores constantes; en este caso todos tienen que incluir un factor $(\lambda-3)$ y, por tanto,

$$d_8(\lambda)=(\lambda-3).$$

Calculando $d_9(\lambda)$ observamos que los menores de orden 9, no nulos, tienen que incluir al menos cuatro factores no constantes de la diagonal y, en este caso, el máximo común divisor es

$$d_9(\lambda)=(\lambda-3)^3(\lambda-1)^2.$$

Por último, $d_{10}(\lambda)=(\lambda-3)^6(\lambda-1)^4$.

De estos divisores, obtenemos los factores invariantes:

$$e_1(\lambda)=e_2(\lambda)=\dots=e_7(\lambda)=1,$$
$$e_8(\lambda)=\lambda-3,\ e_9(\lambda)=(\lambda-3)^2(\lambda-1)^2,\ e_{10}(\lambda)=(\lambda-3)^3(\lambda-1)^2.$$

La forma canónica de $J-\lambda I_{10}$ es:

$$\begin{pmatrix}
1 & & & & & & & & & \\
 & 1 & & & & & & & 0 & \\
 & & 1 & & & & & & & \\
 & & & 1 & & & & & & \\
 & & & & 1 & & & & & \\
 & & & & & 1 & & & & \\
 & & & & & & 1 & & & \\
 & & & & & & & (\lambda-3) & & \\
0 & & & & & & & & (\lambda-1)^2(\lambda-3)^2 & \\
 & & & & & & & & & (\lambda-1)^2(\lambda-3)^3
\end{pmatrix}$$

Para organizar este método utilizaremos los polinomios que aparecen al hacer la primera simplificación. Vamos a ordenarlos en filas por orden decreciente de exponente y según el factor. En este ejemplo tenemos:

$(\lambda-3)^3, (\lambda-3)^2, (\lambda-3)$
$(\lambda-1)^2, (\lambda-1)^2$

Si tomamos el producto de estos polinomios por columnas de este arreglo (considerando un 1 donde no hay polinomio), obtenemos los factores invariantes de la matriz $J-\lambda I_{10}$:

$(\lambda-3)^3(\lambda-1)^2, (\lambda-3)^2(\lambda-1)^2, (\lambda-3).$

Observemos, además, que por cada factor de los factores invariantes, aparece en la matriz J una celda de Jordán asociada al escalar y de orden igual al exponente al que está elevado el factor $(\lambda-a)$.

Vamos a tratar de estructurar el método de la demostración, también a partir del ejemplo.

En la matriz $(J-\lambda I_{10})$ realizamos transformaciones elementales y llegamos a la matriz:

$$\begin{pmatrix} 1 & & & & & & & & & 0 \\ & 1 & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 1 & & & & & & \\ & & & & 1 & & & & & \\ & & & & & (\lambda-3) & & & & \\ & & & & & & (\lambda-3)^2 & & & \\ & & & & & & & (\lambda-1)^2 & & \\ & & & & & & & & (\lambda-1)^2 & \\ 0 & & & & & & & & & (\lambda-3)^3 \end{pmatrix}$$

Aquí hemos ordenado los factores partiendo del último elemento de la diagonal y, subiendo por la diagonal, hemos situado primero las potencias de mayor orden de cada factor $(\lambda-\lambda_i)$, después las potencias de orden inmediatamente inferior en cada caso, etc. Así analizamos los bloques siguientes:

$$\begin{pmatrix} (\lambda-1)^2 & 0 \\ 0 & (\lambda-3)^3 \end{pmatrix}, \begin{pmatrix} (\lambda-1)^2 & 0 \\ 0 & (\lambda-3)^2 \end{pmatrix}, (\lambda-3).$$

En el primer bloque están las potencias de mayor orden de $(\lambda-1)$ y $(\lambda-3)$; en el segundo, las potencias que siguen en orden, y al final, el tercer bloque, que es solamente $(\lambda-3)$; el resto de los factores son unidades. Entonces, por transformaciones elementales, cada uno de estos bloques es equivalente a:

$$\begin{pmatrix} 1 & 0 \\ 0 & (\lambda-1)^2(\lambda-3)^3 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & (\lambda-1)^2(\lambda-3)^2 \end{pmatrix}, (\lambda-3),$$

respectivamente, y transformadas filas y columnas, llegamos a:

$$J-\lambda I_{10} \sim \begin{pmatrix} 1 & & & & & & & & & 0 \\ & 1 & & & & & & & & \\ & & 1 & & & & & & & \\ & & & 1 & & & & & & \\ & & & & 1 & & & & & \\ & & & & & 1 & & & & \\ & & & & & & 1 & & & \\ & & & & & & & (\lambda-3) & & \\ & & & & & & & & (\lambda-1)^2(\lambda-3)^2 & \\ 0 & & & & & & & & & (\lambda-1)^2(\lambda-3)^3 \end{pmatrix}$$

y esta matriz está en la forma diagonal canónica.

Observemos ahora la relación entre J y sus factores invariantes. J tiene:

una celda de orden 3 asociada al escalar 3,
una celda de orden 2 asociada al escalar 3,
una celda de orden 1 asociada al escalar 3,
dos celdas de orden 2 asociadas al escalar 1,

y en los factores invariantes aparecen los factores siguientes:

$(\lambda-3)^3$, $(\lambda-3)^2$, $(\lambda-3)$, $(\lambda-1)^2$, $(\lambda-1)^2$.

Podemos, entonces, intuir que a cada factor $(\lambda-\lambda_0)^k$ que aparezca en uno de los factores invariantes de la matriz $(J-\lambda I_n)$, corresponde una celda de Jordán de orden k asociada al escalar λ_0.

Antes de comenzar el análisis de una matriz de Jordán en general, es necesario demostrar el lema siguiente:

LEMA 4.5

Si una λ-matriz tiene la forma:

$$\begin{pmatrix} p_1(\lambda) & & & 0 \\ & p_2(\lambda) & & \\ & & \ddots & \\ 0 & & & p_n(\lambda) \end{pmatrix}$$

donde los polinomios $p_i(\lambda)$ son primos entre sí, entonces la matriz $A(\lambda)$ es equivalente a la matriz:

$$\begin{pmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 0 & & & & p_1(\lambda)p_2(\lambda)\ldots p_n(\lambda) \end{pmatrix}$$

Demostración

Como la matriz $A(\lambda)$ tiene forma diagonal, basta demostrarlo para el caso $n=2$, pues si es cierto para el caso $n=2$, podemos hacer transformaciones elementales considerando pares de filas y columnas de la forma siguiente:

$$A(\lambda) = \begin{pmatrix} p_1(\lambda) & 0 & 0 & \dots & 0 \\ 0 & p_2(\lambda) & 0 & \dots & 0 \\ 0 & 0 & p_3(\lambda) & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & p_n(\lambda) \end{pmatrix} \sim \begin{pmatrix} 1 & 0 & \dots & 0 \\ 0 & p_1(\lambda)p_2(\lambda) & \dots & 0 \\ \vdots & \vdots & & \\ 0 & 0 & \dots & p_n(\lambda) \end{pmatrix}$$

Como $p_1(\lambda)$, $p_2(\lambda)$ y $p_3(\lambda)$ son primos relativos podemos repetir el proceso con la segunda y la tercera filas y la segunda y tercera columnas, y tenemos:

$$A(\lambda) \sim \begin{pmatrix} 1 & 0 & 0 & \dots & 0 \\ 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & p_1(\lambda)p_2(\lambda)p_3(\lambda) & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & p_n(\lambda) \end{pmatrix}$$

y así continuar el proceso.

Demostremos el lema para el caso $n=2$.
Si $p_1(\lambda)$, $p_2(\lambda)$ son dos polinomios primos entre sí, por la identidad de Bezout existen polinomios $u_1(\lambda)$, $u_2(\lambda) \in K[\lambda]$ tales que

$$u_1(\lambda)p_1(\lambda) + u_2(\lambda)\ p_2(\lambda) = 1.$$

Podemos, entonces, realizar las transformaciones elementales siguientes:

$$(1\quad 2) \rightsquigarrow (1\quad u_1(\lambda)\,1+2) \rightsquigarrow \begin{pmatrix} 1+u_2(\lambda)\,2 \\ 2 \end{pmatrix}$$

$$\rightsquigarrow (2\quad 1) \rightsquigarrow \begin{pmatrix} 1 \\ 2-p_2(\lambda)\,1 \end{pmatrix}$$

$$\rightsquigarrow (1\quad 2-p_1(\lambda)\,1) \rightsquigarrow (1\quad -2).$$

Estas transformaciones nos dan la siguiente secuencia de equivalencias:

$$\begin{pmatrix} p_1(\lambda) & 0 \\ 0 & p_2(\lambda) \end{pmatrix} \approx \begin{pmatrix} p_1(\lambda) & u_1(\lambda)p_1(\lambda) \\ 0 & p_2(\lambda) \end{pmatrix} \approx$$

$$\sim \begin{pmatrix} p_1(\lambda) & u_1(\lambda)p_1(\lambda)+u_2(\lambda)p_2(\lambda) \\ 0 & p_2(\lambda) \end{pmatrix}.$$

Empleando la identidad de Bezout:

$$= \begin{pmatrix} p_1(\lambda) & 1 \\ 0 & p_2(\lambda) \end{pmatrix} \sim \begin{pmatrix} 1 & p_1(\lambda) \\ p_2(\lambda) & 0 \end{pmatrix} \sim \begin{pmatrix} 1 & p_1(\lambda) \\ 0 & -p_1(\lambda)p_2(\lambda) \end{pmatrix}$$

$$\sim \begin{pmatrix} 1 & 0 \\ 0 & -p_1(\lambda)\, p_2(\lambda) \end{pmatrix} \sim \begin{pmatrix} 1 & 0 \\ 0 & p_1(\lambda)\, p_2(\lambda) \end{pmatrix},$$

lo cual completa la demostración del lema.

Sea J una matriz de Jordán, diagonal por bloques, con bloques formados por celdas de Jordán:

$$J_{11}, \ldots, J_{1j,}, J_2, \ldots, J_{2j,}, \ldots, J_{k,}, \ldots, J_{kj_k},$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_k$ son escalares distintos y la celda J_{ij} es una celda de Jordán correspondiente al escalar λ_i de orden n_{ij}.

$$J_{ij} = \begin{pmatrix} \lambda_i & 1 & & 0 \\ & \lambda_i & 1 & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix} \quad \text{y es de orden } n_{ij}$$

$$J = \begin{pmatrix} J_{11} & & & & 0 \\ & \ddots & & & \\ & & J_{ij} & & \\ & & & \ddots & \\ 0 & & & & J_{kj_k} \end{pmatrix}$$

Realizando transformaciones elementales, en la matriz $J-\lambda I_n$ cada bloque $I_{ij}-\lambda I_{n_{ij}}$ se puede reducir a su forma canónica:

$$\begin{vmatrix} 1 & & & 0 \\ & 1 & & \\ & & \ddots & \\ 0 & & & (\lambda-\lambda_i)^{n_{ij}} \end{vmatrix}$$

Así tenemos:

$$J-\lambda I_n \sim \begin{pmatrix} 1 & & & & & 0 \\ & \ddots & & & & \\ & & (\lambda-\lambda_1)^{n_{11}} & & & \\ & & & \ddots & & \\ & & & & (\lambda-\lambda_i)^{n_{ij}} & \\ & & & & & \ddots \\ 0 & & & & & (\lambda-\lambda_k)^{n_{kj_k}} \end{pmatrix}$$

Consideremos la familia Λ de todas las potencias $(\lambda-\lambda_i)^{n_{ij}}$ que aparecen en la matriz que obtuvimos. Vamos a ordenar los elementos de la diagonal, intercambiando filas y columnas de la manera siguiente:
Comenzando por el final de la diagonal principal, sacamos de la familia Λ las potencias de mayor orden de cada factor $(\lambda-\lambda_i)$ y los ponemos en los últimos lugares de la diagonal principal. Queda una familia Λ_1. A continuación repetimos este proceso con Λ_1, o sea, sacamos de Λ_1 las potencias de mayor orden de cada factor $(\lambda-\lambda_i)$ y los ponemos en las últimas posiciones que quedan sin ocupar en la diagonal principal. Queda una nueva familia Λ_2 de potencias de $(\lambda-\lambda_i)$. Repetimos el proceso hasta quedarnos sin ningún elemento en la familia de potencias de $(\lambda-\lambda_i)$ y completamos con 1 la parte superior de la diagonal principal que está sin cubrir.

Ahora consideramos cada subconjunto de elementos de la diagonal principal que se formó con cada paso, como un bloque:

$$A_r(\lambda) = \begin{pmatrix} (\lambda-\lambda_1)^{h_1} & & & 0 \\ & (\lambda-\lambda_2)^{h_2} & & \\ & & \ddots & \\ 0 & & & (\lambda-\lambda_k)^{h_k} \end{pmatrix}$$

Observemos que realmente este no tiene que ser un bloque de orden k, pues puede ser de un paso del proceso en que ya se terminaron las potencias de uno de los $(\lambda-\lambda_i)$, pero por simplificar la notación vamos a suponer que tiene esta forma, y es el bloque que se obtuvo después de efectuar r veces el proceso descrito anteriormente. Por el lema 4.5, como los $(\lambda-\lambda_i)$ son primos entre sí, tenemos:

$$A_r(\lambda) \sim A'_r(\lambda) = \begin{pmatrix} 1 & & & & 0 \\ & 1 & & & \\ & & \ddots & & \\ & & & 1 & \\ 0 & & & & (\lambda-\lambda_1)^{h_1} \dots (\lambda-\lambda_k)^{h_k} \end{pmatrix}$$

Supongamos que el proceso de selección de las potencias tuvo m pasos. Entonces de los m bloques que obtenemos, $A_1(\lambda)$, $A_2(\lambda)$, ..., $A_m(\lambda)$, resultan m polinomios, $p_1(\lambda), p_2(\lambda), \dots, p_m(\lambda)$, que son productos de potencias de $(\lambda-\lambda_i)$, $i=1,2,\dots,k$. Pero en $p_2(\lambda)$ aparecen los factores $(\lambda-\lambda_i)$ con exponentes menores que en $p_1(\lambda)$; luego $p_2(\lambda)$ divide a $p_1(\lambda)$. En general, en $p_j(\lambda)$ las potencias de $(\lambda-\lambda_i)$ tienen menor grado que en $p_{j-1}(\lambda)$ y, por tanto,

$p_j(\lambda)$ divide a $p_{j-1}(\lambda)$.

Por transformaciones elementales tenemos:

$$\begin{pmatrix} 1 & & & & & & 0 \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & A_m(\lambda) & & \\ & & & & & \ddots & \\ 0 & & & & & & A_1(\lambda) \end{pmatrix} \begin{pmatrix} 1 & & & & & & 0 \\ & 1 & & & & & \\ & & \ddots & & & & \\ & & & 1 & & & \\ & & & & p_m(\lambda) & & \\ & & & & & \ddots & \\ 0 & & & & & & p_1(\lambda) \end{pmatrix}$$

En este caso la última matriz está en forma canónica y

$$e_n(\lambda)=p_1(\lambda),\ e_{n-1}(\lambda)=p_2(\lambda),\dots,e_{n-m+1}(\lambda)=p_m(\lambda),$$

y el resto de los factores invariantes son iguales a 1.

Entonces podemos afirmar que en cada polinomio $e_i(\lambda)$, distinto de 1, aparecen factores $(\lambda-\lambda_i)^{h_i}$, que corresponden a la existencia en J de una celda de Jordán asociada al escalar λ_i de orden h_i.

TEOREMA 4.6

Sea J una matriz de Jordán. Entonces los factores invariantes de la matriz $(J-\lambda I_n)$ se escriben como productos de potencias de la forma $(\lambda-\lambda_i)^{h_i}$, donde por cada factor de esta forma que aparece en la forma canónica de $(J-\lambda I_n)$, hay en la matriz J una celda de Jordán asociada al escalar λ_i, de orden h_i.

La demostración de este teorema es la que hemos desarrollado tan laboriosamente. En realidad la dificultad mayor radica en la notación pues el método práctico que se obtiene del teorema es muy fácil de aprender. Veamos otro ejemplo.

Ejemplo

Consideremos la matriz de Jordán:

$$J=\begin{pmatrix} 2 & 1 & & & & & & & \text{\Large 0} \\ 0 & 2 & & & & & & & \\ & & 3 & & & & & & \\ & & & 3 & 1 & & & & \\ & & & 0 & 3 & & & & \\ & & & & & 2 & 1 & & \\ & & & & & 0 & 2 & & \\ & & & & & & & 3 & \\ \text{\Large 0} & & & & & & & & 2 \end{pmatrix}$$

Para buscar los factores invariantes de $J-\lambda I_{10}$, cada celda de Jordán aporta un factor $(\lambda-\lambda_i)^{h_i}$, donde h_i es el orden de la celda; en este caso los factores a considerar son:

$$(\lambda-2)^2,\ (\lambda-3),\ (\lambda-3)^2,\ (\lambda-2)^2,\ (\lambda-3),\ (\lambda-2).$$

Escogemos primero las potencias de mayor orden y las multiplicamos:

$$e_{10}(\lambda)=(\lambda-2)^2\,(\lambda-3)^2.$$

Quedan

$$(\lambda-3),\ (\lambda-2)^2,\ (\lambda-3),\ (\lambda-2).$$

De estas potencias volvemos a tomar las de mayor orden, las multiplicamos y obtenemos $e_9(\lambda)$. En este caso,

$$e_9(\lambda)=(\lambda-2)^2\,(\lambda-3).$$

Quedan

$$(\lambda-3),\ (\lambda-2).$$

El próximo paso consiste en tomar nuevamente las potencias de mayor orden y tenemos:

$$e_8(\lambda)=(\lambda-3)(\lambda-2).$$

El resto de los factores invariantes son iguales a 1.

$$e_1(\lambda)=\ldots=e_7(\lambda)=1,\ e_8(\lambda)=(\lambda-2)(\lambda-3),$$
$$e_9(\lambda)=(\lambda-2)^2\,(\lambda-3),\ e_{10}(\lambda)=(\lambda-2)^2\,(\lambda-3).$$

Observemos que hay dos celdas de orden 2 para el escalar 2, por tanto, tiene que aparecer dos veces $(\lambda-2)^2$. Hay una celda de orden 1 para el escalar 2; tiene que aparecer una sola vez el factor $(\lambda-2)$. Hay una celda de orden 2 asociada al escalar 3; aparece una vez el factor $(\lambda-3)^2$. Hay dos celdas de orden 1 asociadas al escalar 3; aparece dos veces el factor $(\lambda-3)$.

Si analizamos detenidamente la demostración del teorema, podemos observar que el orden en que aparecen las celdas de Jordán en la matriz J no influye en el resultado final; por tanto, podemos enunciar el teorema siguiente:

TEOREMA 4.7

Dos matrices de Jordán son semejantes si y solo si están formadas por las mismas celdas de Jordán, las cuales pueden estar en un orden diferente en la diagonal.

Demostración

Sean J y J' dos matrices de Jordán. Si J y J' son semejantes, esto quiere decir que $J-\lambda I_n$ y $J'-\lambda I_n$ son equivalentes y, por tanto, tienen la misma forma canónica. Según vimos en los factores invariantes asociados en una matriz de Jordán, cada factor $(\lambda-\lambda_i)^{h_i}$ que aparece, corresponde a una celda de Jordán del escalar λ_i de orden h_i. Como $J-\lambda I_n$ y $J'-\lambda I_n$ tienen los mismos factores invariantes, entonces J y J' están formados por las mismas celdas de Jordán, lo único que puede variar es el orden.

Si J y J' están formadas por las mismas celdas de Jordán, entonces $J-\lambda I_n$ y $J'-\lambda I_n$ tienen los mismos factores invariantes y, por tanto, son equivalentes. Luego, por el teorema 4.5 J y J' son semejantes.

Ahora plantearemos el problema siguiente:

¿Cuándo una matriz A es semejante a una matriz de Jordán J?

Por el teorema 4.5, las dos matrices son semejantes si las λ-matrices asociadas $(A-\lambda I_n)$ y $(J-\lambda I_n)$ son equivalentes, es decir, si las matrices $(A-\lambda I_n)$ y $(J-\lambda I_n)$ tienen la misma forma canónica, o lo que es equivalente, si tienen los mismos factores invariantes.

Ya conocemos la forma canónica de la matriz $J-\lambda I_n$. Si J es una matriz de Jordán, ¿qué tiene de especial su forma canónica? Que todos los factores invariantes se descomponen como potencias de factores lineales. Esta es la única condición que necesitamos.

TEOREMA 4.8 (teorema de Jordán)

Sea f un endomorfismo de un K-espacio E de dimensión n tal, que su polinomio característico tiene todas sus raíces en K. Entonces existe una base B de E y una matriz de Jordán J que representa a f en la base B.

$$J=M(f,\ B).$$

La matriz J es única, salvo el orden en que aparecen sus celdas.

Demostración

Sea B' una base de E y A la matriz que representa a f en esa base.

$$A=M(f,\ B').$$

Como el polinomio característico tiene todas sus raíces en K, entonces $\det(A-\lambda I_n)$ se descompone completamente como potencias de factores lineales. Pero

$$\det(A-\lambda I_n)=d_n(\lambda)=e_1(\lambda)\,e_2(\lambda)\ldots e_n(\lambda)$$

y, por tanto, los factores invariantes $e_i(\lambda)$ se descomponen también completamente como potencias de factores lineales.

A partir de la descomposición de los polinomios $e_i(\lambda)$, existe una matriz de Jordán J formada por celdas de Jordán de la manera siguiente:

Por cada factor $(\lambda-\lambda_i)^h$ que aparezca en uno de los $e_i(\lambda)$, J tendrá una celda de Jordán de orden h asociada al escalar λ_i.

La matriz de Jordán así construida será tal, que su λ-matriz asociada $J-\lambda I_n$ tendrá los mismos factores invariantes que $A-\lambda I_n$ y, por tanto, serán equivalentes.

Por el teorema 4.5 las matrices A y J son semejantes. Esto quiere decir que existe una matriz inversible P de orden n tal que

$$J=P^{-1}AP.$$

Entonces, por la proposición 1.11, existe una base B de E tal, que la matriz P es la matriz de cambio de coordenadas de la base B a la base B' y, por tanto,

$$J=M(f,B).$$

La unicidad de la matriz de Jordán la garantiza el teorema 4.8. Esto completa la demostración del teorema.

Si estamos trabajando con $K=\mathbb{C}$, es decir, con espacios vectoriales complejos, podemos afirmar que todo endomorfismo se puede representar por una matriz de Jordán, porque por el teorema fundamental del Álgebra, todo polinomio con coeficientes complejos se descompone en factores lineales. Además, si un endomorfismo es diagonalizable, su matriz de Jordán es la propia matriz diagonal, que es también de Jordán.

Lo importante del método que hemos seguido es que las demostraciones dan los métodos de trabajo, y que estos son simples. Las demostraciones parecen muy complicadas por las dificultades en la notación.

El método para hallar la matriz de Jordan es el siguiente:

1ro. Tenemos un endomorfismo f de E. Si lo dan en forma de matriz, trabajamos con la matriz, de lo contrario, buscamos la matriz asociada a f en una base que escogemos de E.

2do. Con la matriz obtenida, A, formamos $A-\lambda I_n$ y procedemos a calcular su polinomio característico y verificar que todas sus raíces están en el cuerpo de escalares en que estamos trabajando.

3ro. Procedemos a calcular los divisores elementales: $d_1(\lambda), d_2(\lambda), \ldots, d_n(\lambda)$.

4to. A partir de los divisores elementales, calculamos los factores invariantes:

$$e_i(\lambda) = \frac{d_i(\lambda)}{d_{i-1}(\lambda)}.$$

5to. Por cada factor $(\lambda-\lambda_i)^h$ que aparezca en un factor invariante, tenemos una celda de Jordán de orden h asociada al escalar λi.

6to. Para encontrar, con los elementos que tenemos, la base en la cual f se representa por la matriz de Jordán obtenida, hay que recorrer toda la cadena de transformaciones elementales que indican los teoremas, y de esta forma lograr una matriz de cambio de coordenadas.
El estudiante debe leer la demostración del Teorema 4.5 para que comprenda que aquí hay un "pequeño" problema.

Exceptuando este último punto, el de buscar la base, el resto del método para buscar la matriz de Jordán no presenta dificultades. Vamos a mostrar un ejemplo, en el cual resolveremos por otra vía la dificultad de hallar la base.

Ejemplo

Consideremos el endomorfismo f de $\mathbb{R}^4$ dado por la matriz

$$A=\begin{pmatrix} 2 & 0 & 0 & 0 \\ 1 & 3 & 1 & 0 \\ -1 & -1 & 1 & 1 \\ -1 & -1 & -1 & 0 \end{pmatrix}$$

de la cual sabemos, por ejemplos anteriores, que tiene el polinomio característico:

$$(t-2)^2(t-1)^2$$

y no es diagonalizable, pues

$$E(1)=\{(0,\alpha,-2\alpha,\ A)/\ \alpha\in\mathbb{R}\},$$
$$E(2)=\{(\alpha,\beta,-\alpha-\beta,0)/\alpha,\beta\in\mathbb{R}\}.$$

Tenemos dim $E(1)=1$, dim $E(2)=2$ y no se satisface la condición necesaria y suficiente de diagonalización.

Busquemos la matriz de Jordán asociada a A.

La matriz de Jordán existe porque el polinomio característico de A tiene todas sus raíces en $\mathbb{R}$.

Construimos la matriz $A-\lambda I_4$:

$$A-\lambda I_4=\begin{pmatrix} 2-\lambda & 0 & 0 & 0 \\ 1 & 3-\lambda & 1 & 0 \\ -1 & -1 & 1-\lambda & 1 \\ -1 & -1 & -1 & -\lambda \end{pmatrix}$$

y calculamos sus divisores elementales:

$$d_1(\lambda)=1,\ d_2(\lambda)=1,\ d_3(\lambda)=(\lambda,-2),\ d_4(\lambda)=(\lambda-1)^2(\lambda-2)^2.$$

A partir de estos divisores calculamos los factores invariantes:

$$e_1(\lambda)=1,\ e_2(\lambda)=1,\ e_3(\lambda)=(\lambda-2),\ e_4(\lambda)=(\lambda-1)^2(\lambda-2).$$

De estos factores invariantes resulta que la matriz de Jordán asociada a A tiene:

a) dos celdas de Jordán de orden 1 asociadas al valor propio 2;
b) una celda de Jordán de orden 2 asociada al valor propio 1.

Por tanto, podemos escribirla como:

$$J=\begin{pmatrix}2&0&0&0\\0&2&0&0\\0&0&1&1\\0&0&0&1\end{pmatrix}$$

Veamos ahora el procedimiento práctico que vamos a emplear para calcular la base en la que el endomorfismo se representa por J.

Supongamos que (a_1,a_2,a_3,a_4) es una base en la que el endomorfismo f se representa por J. Entonces, por la definición de matriz asociada a un endomorfismo en una base, tenemos:

$$\begin{aligned}
f(a_1)&=2a_1\\
f(a_2)&=2a_2\\
f(a_3)&=\ a_3\\
f(a_4)&=a_3+a_4
\end{aligned}$$

De aquí resulta que a_1 y a_2 son vectores propios, independientes, asociados al valor propio 2, y a_3 es un vector propio asociado al valor propio 1. Por tanto, podemos plantear:

$$a_1=(1,0,-1,0),\ a_2=(0,1,-1,0),\ a_3=(0,1,-2,1).$$

Para calcular a_4 tenemos que transformar la última relación:

$$\begin{aligned}
f(a_4)&=a_3+a_4\\
f(a_4)-a_4&=a_3\\
(f-Id)(a_4)&=a_3.
\end{aligned}$$

El vector a_3 es conocido; por tanto, hay que buscar un vector $a_4=(x,y,z,t)$ que satisfaga el sistema de ecuaciones lineales no homogéneo:

$$(A-I_4)(a_4)=a_3$$

$$\begin{pmatrix}1&0&0&0\\1&2&1&0\\-1&-1&0&1\\-1&-1&-1&-1\end{pmatrix}\begin{pmatrix}x\\y\\z\\t\end{pmatrix}=\begin{pmatrix}0\\1\\-2\\1\end{pmatrix}.$$

Este se transforma en el sistema equivalente:

$$\begin{cases}x=0\\-y+t=-2\\z+2t=-3\end{cases}$$

y obtenemos una solución particular de este sistema para $t=0$:

$$x=0,\ y=2,\ z=-3,\ t=0,$$

de donde resulta el vector

$$a_4=(0,2,-3,\ 0),$$

La matriz P de cambio de coordenadas de la base (a_1,a_2,a_3,a_4) a la base canónica de $\mathbf{R}^4$ es

$$P=\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 2 \\ -1 & -1 & -2 & -3 \\ 0 & 0 & 1 & 0 \end{pmatrix}$$

y podemos comprobar que

$$J=P^{-1}AP.$$

Como podemos observar, el método para encontrar la base consiste en buscar cada uno de los vectores a partir del anterior. Los vectores de la base están relacionados entre sí por las igualdades:

$$f(a_1)=\lambda_1 a_1$$
$$f(a_i)=\lambda_i a_i \text{ o } f(a_i)=\lambda_i a_i+\ a_{i-1},$$

las cuales se transforman en

$$(f-\lambda_i I_n)(a_i)=0 \text{ o } (f-\lambda_i I_n)(a_i)=\ a_{i-1}.$$

Por tanto, cada vector a_i, o es un vector propio de un cierto valor propio, o satisface un sistema de ecuaciones lineales no homogéneo.

Aquí se presenta un posible problema: que el sistema planteado no sea compatible, es decir, que el vector seleccionado en un paso anterior, a_{i-1}, no haga compatible el sistema que se forma. Esto significa que hay que seleccionar otro vector en el paso anterior.

La justificación de por qué mediante este método se obtiene una base, se verá más ampliamente en el próximo capítulo.

Ejercicios

1. Reduzca las siguientes λ-matrices a su forma canónica mediante transformaciones elementales.

a) $\begin{pmatrix} \lambda-2 & -1 & 0 \\ 0 & \lambda-2 & -1 \\ 0 & 0 & \lambda-2 \end{pmatrix}$

c) $\begin{pmatrix} \lambda(\lambda+1) & 0 & 0 \\ 0 & \lambda & 0 \\ 0 & 0 & \lambda(\lambda+1)^2 \end{pmatrix}$

b) $\begin{pmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda & \lambda^2 & -\lambda^2 \end{pmatrix}$

d) $\begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 5 & 4 & 3 & \lambda+2 \end{pmatrix}$

2. Para las λ-matrices dadas, calcule los divisores elementales, los vectores invariantes y la forma canónica de la λ-matriz.

a) $\begin{pmatrix} 1-\lambda & \lambda^2 & \lambda \\ \lambda & \lambda & -\lambda \\ 1+\lambda^2 & \lambda^2 & -\lambda^2 \end{pmatrix}$ c) $\begin{pmatrix} \lambda^2+2 & \lambda^2+1 & \lambda^2+1 \\ 3 & \lambda^2+1 & 3 \\ \lambda^2+1 & \lambda^2+1 & \lambda^2+1 \end{pmatrix}$

b) $\begin{pmatrix} \lambda & 1 & 0 & 0 \\ 0 & \lambda & 1 & 0 \\ 0 & 0 & \lambda & 1 \\ 5 & 4 & 3 & \lambda+2 \end{pmatrix}$ d) $\begin{pmatrix} \lambda(\lambda+1) & 0 & 0 & 0 \\ 0 & \lambda^2 & 0 & 0 \\ 0 & 0 & (\lambda+1)^2 & 0 \\ 0 & 0 & 0 & \lambda(\lambda+1) \end{pmatrix}$

3. Para los endomorfismos dados en el ejercicio 1 del capítulo 3, determine si se pueden representar por una matriz de Jordán y encuentre una base en la cual el endomorfismo se represente por una matriz de Jordán.

4. Para las matrices dadas en el ejercicio 2 del capítulo 3, determine si existe una matriz de Jordán semejante, y en caso de que sea posible, encuentre una matriz inversible P tal que $J=P^{-1}AP$.

5. Para las matrices dadas, determine si existe una matriz de Jordán real J semejante y en caso de que sea posible, encuentre una matriz inversible P tal que $J=P^{-1}AP$.

a) $\begin{pmatrix} 1 & 2 & 0 \\ 0 & 2 & 0 \\ -2 & -2 & -1 \end{pmatrix}$ g) $\begin{pmatrix} 3 & 1 & 0 & 0 \\ -4 & -1 & 0 & 0 \\ 7 & 1 & 2 & 1 \\ -17 & -6 & -1 & 0 \end{pmatrix}$

b) $\begin{pmatrix} 13 & 16 & 16 \\ -5 & -7 & -6 \\ -6 & -8 & -7 \end{pmatrix}$ h) $\begin{pmatrix} 4 & 6 & 0 \\ -3 & -5 & 0 \\ -3 & -6 & 1 \end{pmatrix}$

c) $\begin{pmatrix} 3 & 3 & 3 \\ -6 & -6 & -2 \\ 6 & 5 & -1 \end{pmatrix}$ i) $\begin{pmatrix} 3 & 0 & 8 \\ 3 & -1 & 6 \\ -2 & 0 & 5 \end{pmatrix}$

d) $\begin{pmatrix} 3 & -2 & -2 \\ 5 & 4 & 5 \\ 3 & 1 & 0 \end{pmatrix}$ j) $\begin{pmatrix} 0 & 2 & 1 \\ 1 & -4 & 0 \\ 3 & 0 & 0 \end{pmatrix}$

e) $\begin{pmatrix} 3 & 2 & -3 \\ 4 & 10 & -12 \\ 3 & 6 & -7 \end{pmatrix}$ k) $\begin{pmatrix} 0 & 3 & 3 \\ -1 & 8 & 6 \\ 2 & -14 & -10 \end{pmatrix}$

f) $\begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & 1 & 1 \end{pmatrix}$

l) $\begin{pmatrix} 1 & 1 & -1 & 2 & -1 \\ 2 & 0 & 1 & -4 & -1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & -3 & 3 & -1 \end{pmatrix}$ m) $\begin{pmatrix} 1 & 2 & 3 & 4 \\ 0 & 1 & 2 & 3 \\ 0 & 0 & 1 & 2 \\ 0 & 0 & 0 & 1 \end{pmatrix}$

6. Determine la forma canónica de una λ-matriz de orden 5 y rango 4, cuyo sistema de divisores elementales tiene por factores a λ, λ, λ^2, $(\lambda+1)$, $(\lambda+1)^2$, $(\lambda-1)$.

7. Demuestre que toda matriz compleja A de orden n es semejante a su traspuesta.

8. Demuestre que todas las matrices complejas de orden n que cumplen $A^n = I_n$ son semejantes.

9. Sea A una matriz compleja de orden n que tiene todos sus valores propios reales. Demuestre que A es semejante a una matriz real.

10. Demuestre que toda matriz A de orden n con coeficientes en K, cuyo polinomio característico se descompone totalmente en K, se puede descomponer como la suma de dos matrices, una diagonalizable y la otra nilpotente.

11. Diga cuáles de las matrices de los ejercicios 4 y 5 son semejantes entre sí.

$A - \lambda I$

$\begin{vmatrix} 1-\lambda & 2 & 0 \\ 0 & 2-\lambda & 0 \\ -2 & -2 & -1-\lambda \end{vmatrix} = (-1-\lambda)\begin{vmatrix} 1-\lambda & 2 \\ 0 & 2-\lambda \end{vmatrix} = (-1-\lambda)(1-\lambda)(2-\lambda)$

$\lambda_1 = -1 \quad \lambda = 1 \quad \lambda = 2$

$\begin{pmatrix} -1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix} \qquad \begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & \boxed{2} \end{pmatrix}$

5 Los endomorfismos nilpotentes y el teorema de Jordán

Introducción

En los capítulos anteriores estudiamos un importante problema del Álgebra lineal: la posibilidad de representar los endomorfismos de un espacio vectorial de dimensión finita mediante matrices de la forma más simple. En relación con esto planteamos las condiciones para que un endomorfismo pudiese ser representado mediante una matriz diagonal, y en caso de que esto no fuera posible, explicamos cómo representar el endomorfismo mediante una matriz de Jordán. Este tratamiento matricial, aunque laborioso, tiene entre otras ventajas la de dar un criterio para la semejanza de matrices, sin tener que calcular los valores propios, y por tanto, sin hallar raíces de polinomios, pues es conocido que la búsqueda de las raíces de un polinomio de grado mayor que 4 presenta dificultades. Sin embargo, el tratamiento dado a la búsqueda de la matriz de Jordán de un endomorfismo se aleja bastante del método seguido en el caso de los endomorfismos diagonalizables, que consiste en poder encontrar una base del espacio mediante la unión de bases de subespacios invariantes por el endomorfismo. En el caso de los endomorfismos diagonalizables, vimos que dichos subespacios invariantes son los subespacios propios.

Es posible dar a la búsqueda de la matriz de Jordán un tratamiento análogo al dado a los endomorfismos diagonalizables, el cual tiene la ventaja de su posible generalización al estudio de los endomorfismos de espacios vectoriales de dimensión infinita, de gran importancia en el Análisis funcional, las Ecuaciones diferenciales, la Física, etcétera. Es por ello que dedicamos este capítulo al tratamiento de la reducción de la matriz de un endomorfismo a la forma de Jordán por la vía funcional mediante los llamados *endomorfismos nilpotentes.* De este modo se completa el estudio de las dos vías fundamentales para el tratamiento de este problema.

La vía de demostración del teorema de Jordán mediante los endomorfismos nilpotentes necesita de muchos recursos técnicos, por lo que no siempre será posible seguir el método de exposición usado en los capítulos 1, 2 y 3. Muchas definiciones y resultados tendrán que ser aceptados por el estudiante sin que previamente tenga una idea de su utilidad; solamente podrá comprenderlos cuando sean aplicados.

5.1 Subespacios característicos de un endomorfismo

El hecho de que un endomorfismo f de E sea diagonalizable está determinado por la posibilidad de encontrar una base de E formada por la unión de las bases de los subespacios propios de f; en otras palabras, por la posibilidad de expresar E como

$$E = \text{Ker } (f-\lambda_1\ id_E) \oplus \text{Ker } (f-\lambda_2 id_E) \oplus \ldots \oplus \text{Ker } (f-\lambda_k id_E),$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los valores propios distintos de f.

Sabemos que no para todos los endomorfismos, los subespacios propios aportan el número suficiente de vectores linealmente independientes que permita formar una base del espacio; es decir, que no para todo endomorfismo f se cumple que la suma directa de sus subespacios propios sea el espacio E completo. Este es el caso del endomorfismo f de $\mathbb{R}^5$ representado en la base canónica de dicho espacio por

$$A = \begin{pmatrix} 1 & 1 & -1 & 2 & -1 \\ 2 & 0 & 1 & -4 & -1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & -3 & 3 & -1 \end{pmatrix}$$

cuyo polinomio característico es $p_f(\lambda) = -(\lambda+1)^2(\lambda-1)^3$. Sus subespacios propios tienen dimensiones

$$\dim E(-1) = \dim \text{Ker } (f-id_E) = 2,$$
$$\dim E(1) = \dim \text{Ker } (f+id_E) = 2,$$

por lo que $\mathbb{R}^5 \neq E(-1) \oplus E(1)$.

La característica de los subespacios propios que determina la obtención de matrices simples para representar el endomorfismo, es la *invarianza* de dichos subespacios por el endomorfismo, lo que da lugar a la aparición de ceros en la matriz.

Si en el ejemplo anterior, en lugar de $\text{Ker}(f-id_E)$ y $\text{Ker}(f+id_E)$ tomamos

$$V_1 = \text{Ker}(f-id_E)^3 \text{ y } V_2 = \text{Ker}(f+id_E)^2,$$

observaremos que estos subespacios son estables por f, lo que permite obtener para una base de E formada por la unión de una base de V_1 y otra de V_2 una matriz que represente a f de la forma

$$\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix}.$$

En efecto, si $x \in \text{Ker } (f-id_E)^3$, entonces $f(x) \in \text{Ker } (f-id_E)^3$, pues $(f-id_E)^3(f(x)) = f(f^3 - id_E(x)) = f(0) = 0$.

Análogamente si $x \in \text{Ker}(f+id_E)^2$, entonces $f(x) \in \text{Ker } (f+id_E)^2$.

Esto resulta de una propiedad planteada en el ejercicio 49 del capítulo de aplicaciones lineales, que establece que *si dos endomorfismos conmutan, el núcleo de cada uno de ellos es estable por el otro.*

No es difícil darse cuenta de que si en el ejemplo anterior hubiéramos tomado un endomorfismo cualquiera y los subespacios $\text{Ker } (f-\lambda_i\ id_E)^{k_i}$, donde k_i es la multiplicidad de λ_i como raíz del polinomio característico de

f, podríamos demostrar que estos subespacios son estables por f. Daremos, pues, la definición siguiente:

DEFINICIÓN 5.1

Se denomina *subespacio característico* de f asociado al valor propio λ de f, al núcleo del endomorfismo $(f-\lambda id_E)^k$, donde k es la multiplicidad de λ.

Como ya analizamos, estos subespacios son estables por f y serán la generalización de los subespacios propios que permitirá hallar una base del espacio en la cual la matriz de f tiene la forma de Jordán. Realmente todo subespacio de la forma Ker $(f-\lambda id_E)^r$, donde λ es un valor propio de f y r un número natural, es estable por f por las mismas razones que los subespacios característicos.

El camino a seguir para obtener el teorema de Jordán es largo y es necesario dar algunas definiciones y obtener algunos resultados previos, así como recordar ciertos resultados conocidos sobre polinomios.

5.2 Polinomios que se anulan en un endomorfismo o en una matriz

En los razonamientos que hemos realizado desde el inicio de este capítulo, hemos visto cómo la búsqueda de subespacios estables por un endomorfismo f está relacionada con expresiones del tipo $(f-\lambda id_E)^k$ y esto no es más que un polinomio de $K[x]$, $(x-\lambda)^k$, evaluado en f. Por tanto, vamos a estudiar expresiones más generales del tipo $p(f)$, donde p es un polinomio de $K[x]$.

Más adelante veremos que tienen especial interés los polinomios que se anulan en un endomorfismo, esto es, $p\in K[x]$ tal que $p(f)=0$.

La primera pregunta que lógicamente debemos hacer es la siguiente:

Dados un espacio vectorial de dimensión n sobre K y un endomorfismo f de E, ¿existirá un polinomio de $K[x]$ tal que $p(f)=0$?

Demos respuesta a esta pregunta.

PROPOSICIÓN 5.1

Sean E un espacio vectorial de dimensión n sobre K, f un endomorfismo de E. Entonces existe un polinomio no nulo p de $K[x]$ tal que $p(f)=0$.

Demostración

Sabemos que $L_K(E)$ es un espacio vectorial sobre K de dimensión n^2. Consideremos los n^2+1 endomorfismos:

$$id_E,\ f,\ f^2, \dots,\ f^{n^2}.$$

Estos tienen que ser linealmente dependientes, o sea, existen a_0, $a_1, \dots$, a_{n^2}, no todos nulos, tales que

$$a_0 id_E + a_1 f + a_2 f^2 + \dots + a_{n^2} f^{n^2} = 0.$$

De aquí que el polinomio

$$p(x) = a_0 + a_1 x + \ldots + a_{n^2} x^{n^2} \text{ de } K[x]$$

sea tal, que $p(f) = 0$. Con esto queda probada la proposición.

El estudiante puede verificar fácilmente que si $p(f) = 0$ y $q(x) \in K[x]$, entonces $(q \cdot p)(f) = 0$. Además, si $p(f) = 0$ y $q(f) = 0$ se tiene que $(p \pm q)(f) = 0$.

Queremos destacar que $(q \cdot p)(f) = q(f) \circ p(f)$ y que $q(f) \circ p(f) = p(f) \circ q(f)$, pues en los polinomios en f solo aparecen potencias de f y la identidad, y por tanto, conmutan.

La otra cuestión importante que es preciso esclarecer es la siguiente:

Dado un endormorfismo f de un espacio vectorial E de dimensión n sobre K, ¿qué relación existe entre los diferentes polinomios que se anulan en f? Veamos.

PROPOSICIÓN 5.2

Sean f un endomorfismo de un espacio vectorial de dimensión n sobre Ke,

$$I = \{p(x) \in K[x] / p(f) = 0\}.$$

Entonces I es de la forma $I = \{q(x) \cdot m(x)\}$, donde $m(x)$ es el polinomio mónico de grado mínimo tal que $m(f) = 0$.

Demostración

Anteriormente planteamos, para que fuera comprobado por el estudiante, que si $p(x)$ y $q(x)$ pertenecen a I, entonces $p(x) \pm q(x)$ pertenece a I y que $p(x) \cdot q(x) \in I$ para todo $p(x) \in I$ y $q(x) \in K[x]$. Ahora tenemos que probar que efectivamente existe en I un polinomio mónico de grado mínimo tal que para todo $p(x) \in K[x]$, $p(x) = q(x) \cdot m(x)$ para algún $q(x) \in K[x]$.

Sabemos que $I \neq \{0\}$ por la proposición 5.1. Como el grado de un polinomio es un número natural, está claro que entre los elementos de I tienen que existir polinomios de grado mínimo y que entre ellos podemos tomar uno que sea mónico, llamémosle $m(x)$.

Debemos probar que todo $p(x) \in I$ es de la forma $p(x) = m(x) \cdot q(x)$ para algún $q(x) \in K[x]$.

Empleando el algoritmo de la división con resto, en $K[x]$ existen $q(x)$, $r(x) \in K[x]$ únicos tales que

$$p(x) = m(x) \cdot q(x) + r(x),$$

donde $r(x) = 0$ o grado $r(x) <$ grado $m(x)$, por tanto $r(x) = p(x) - m(x) \cdot q(x)$. Como $p(x) \in I$, $m(x) \in I$, entonces $m(x) \cdot q(x) \in I$ y la diferencia $p(x) - m(x) \cdot q(x) = r(x) \in I$.

De lo anterior resulta que $r(x) = 0$, pues tomamos $m(x)$ de grado mínimo entre los elementos de I.

Así queda probado que $p(x) = m(x) \cdot q(x)$ para un cierto $q(x) \in K[x]$. Falta demostrar que el polinomio $m(x)$ es único.

Supongamos que hay otro polinomio $m'(x)$, mónico, que tenga el mismo grado que $m(x)$, y por tanto, sea de grado mínimo en I. Entonces todo polinomio $p(x) \in I$ se puede expresar también en la forma

$$p(x) = m'(x) \cdot q(x) \text{ para } q(x) \in K[x].$$

En particular, podemos hacer esto con $m(x)$ y $m'(x)$ y obtenemos que existe un polinomio $q(x) \in K[x]$ tal que

$$m(x) = m'(x) \cdot q(x)$$

Pero $m(x)$ y $m'(x)$ tienen el mismo grado, de aquí que $q(x)$ sea una constante, y como $m(x)$ y $m'(x)$ son polinomios mónicos, entonces $q(x) = 1$ y

$$m(x) = m'(x)$$

Con esto se demuestra la unicidad del polinomio $m(x)$.

DEFINICIÓN 5.2

Se denomina *polinomio minimal* de f y se denota por w_f, el polinomio mónico de grado mínimo que se anula en un endomorfismo f de un espacio vectorial de dimensión n sobre K.

Obsérvese que, según esta definición, el polinomio minimal de un endomorfismo es único.

De la relación de isomorfismo que existe entre el espacio de los endomorfismos de un espacio E de dimensión n y el espacio de las matrices cuadradas de orden n resulta que podemos definir el polinomio minimal de una matriz A como el polinomio mónico de menor grado que se anula en A(se denotará por $w_A(x)$) y que el polinomio minimal de un endomorfismo f es el de cualquiera de sus matrices asociadas. Obsérvese que para demostrar esto se debe utilizar el hecho de que si $p(x) \in K[x]$ es tal que $p(A) = 0$ y $B = P^{-1}AP$ con P inversible es semejante a A, entonces $p(B) = 0$.

Una vez definido el polinomio minimal de un endomorfismo, debemos tratar de hallar un método para calcularlo, pero los datos que conocemos hasta ahora sobre el polinomio minimal son pocos y de carácter teórico.

Hay otro polinomio notable asociado a cada endomorfismo f de un espacio vectorial de dimensión finita: el polinomio característico p_f, para el cual sí conocemos un método de cálculo. En algunos ejercicios se pidió al estudiante que comprobara que al evaluar el polinomio característico de un endomorfismo, se obtenía el endomorfismo nulo; esto puede comprobarse en general mediante el teorema siguiente:

TEOREMA 5.1 (teorema de Hamilton-Cayley)

Sean f un endomorfismo de un espacio vectorial E de dimensión finita sobre K y p_f su polinomio característico. Entonces $p_f(f) = 0$.

Demostración

Para la demostración de este teorema admitiremos un resultado que demostraremos posteriormente y que será de gran utilidad: la posibilidad de encontrar para todo endomorfismo de un espacio vectorial de dimensión finita sobre $\mathbb{C}$, una base en la cual la matriz del endomorfismo tome la forma triangular superior. Por supuesto que la utilización de este resultado limitará nuestra demostración a endomorfismo de espacios vectoriales de dimensión finita sobre $\mathbb{C}$, pero luego lo generalizaremos para el caso $K=\mathbb{R}$.

Supongamos que E es un espacio vectorial de dimensión n sobre $\mathbb{C}$, f un endomorfismo de E y p_f su polinomio característico. Probemos que $p_f(f)=0$. Esto equivale a $p_f(f)(x)=0 \ \ \forall x \in E$.

Para ello basta probar que $p_f(f)(a_i)=0$ para todo a_i de una base de E.

Apliquemos el resultado que mencionamos al principio de la demostración y consideremos una base de E, $(a_1,a_2,\ldots,a_n)$, en la cual la matriz de f toma la forma triangular:

$$B=M(f,(a_i))=\begin{pmatrix} \lambda_1 & \beta_{12} & \cdots & \beta_{1n} \\ 0 & \lambda_2 & \cdots & \beta_{2n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & 0 & \cdots & \lambda_n \end{pmatrix},$$

es decir,

$$f(b_1)=\lambda_1 b_1$$
$$f(b_2)=\beta_{12}b_1+\lambda_2 b_2$$
$$\vdots$$
$$f(b_i)=\beta_{1i}b_1+\ldots+\lambda_i b_i$$
$$\vdots$$
$$f(b_n)=\beta_{1n}b_1+\ldots+\beta_{n-1\,n}b_{n-1}+\lambda_n b_n.$$

Está claro que los elementos de la diagonal de B tienen que ser los valores propios $\lambda_1,\lambda_2,\ldots,\lambda_n$ de f, pues

$$p_B(x)=(-1)^n(x-\lambda_1)\ldots(x-\lambda_n)=p_f(x).$$

luego

$$p_f(f)=(-1)^n(f-\lambda_1 id_E)\circ(f-\lambda_2 id_E)\circ\ldots\circ(f-\lambda_n id_E).$$

Los endomorfismos $f-\lambda_i id_E$, $1\leqslant i\leqslant n$, conmutan entre sí, esto es,

$$(f-\lambda_i id_E)\circ(f-\lambda_j id_E)=(f-\lambda_j id_E)\circ(f-\lambda_i id_E)\ \forall i,j.$$

Probemos por inducción en n que $(f-\lambda_1 id_E)\circ(f-\lambda_2 id_E)\circ\ldots\circ(f-\lambda_n id_E)$ se anula en $a_1,a_2,\ldots,a_n$.

Para $n=1$, es evidente que

$$(f-\lambda_1 id_E)(a_1)=f(a_1)-\lambda_1 a_1=0,$$

por la definición de B y, por tanto, de $f(a_1)$.

Supongamos que $(f-\lambda_1 id_E)\circ\ldots\circ(f-\lambda_{n-1}id_E)$ se anula en $a_1,\ldots,a_{n-1}$ y probemos que $(f-\lambda_1 id_E)\circ\ldots\circ(f-\lambda_n id_E)$ se anula en $a_1,\ldots,a_n$.

Podemos escribir:

$$\begin{aligned} p_f(f) &= [(f-\lambda_1 id_E)\circ\ldots\circ(f-\lambda_{n-1}id_E)]\circ(f-\lambda_n id_E) \\ &= (f-\lambda_n id_E)\circ(f-\lambda_1 id_E)\circ\ldots\circ(f-\lambda_{n-1}id_E). \end{aligned}$$

De aquí que por la hipótesis de inducción,

$$\begin{aligned} p_f(f)(a_i) &= (f-\lambda_n id_E)\circ[(f-\lambda_1 id_E)\circ\ldots\circ(f-\lambda_{n-1}id_E)](a_i) \\ &= (f-\lambda_n id_E)(0)=0,\ 1\leqslant i\leqslant n-1 \end{aligned}$$

y

$$p_f(f)(a_n)=(f-\lambda_1 id_E)\circ\ldots\circ(f-\lambda_{n-1}id_E)\circ(f-\lambda_n id_E)(a_n).$$

donde

$$\begin{aligned} (f-\lambda_n id_E)\ (a_n) &= f(a_n)-\lambda_n a_n=\beta_{1n}a_1+\ldots+\beta_{n-1,n}a_{n-1}+\lambda_n a_n-\lambda_n a_n \\ &= \beta_{1n}a_1+\ldots+\beta_{n-1,n}a_{n-1} \end{aligned}$$

Por tanto,

$$p_f(f)(a_n)=[(f-\lambda_1 id_E)\circ\ldots\circ(f-\lambda_{n-1}id_E)](\beta_{1n}a_1+\ldots+\beta_{n-1,n}a_{n-1})=0$$

por la hipótesis de inducción y la linealidad de las $f(f-\lambda_i id_E)$: queda pues demostrado que $p_f(f)(a_i)=0$ para todo $i \Longleftrightarrow p_f(f)=0$.

¿Qué ocurrirá si f es un endomorfismo de un espacio vectorial sobre $\mathbb{R}$? Sabemos que la diferencia fundamental con el caso $K=\mathbb{C}$ está dada por el hecho de que las raíces de p_f cuando $K=\mathbb{R}$ pueden no estar en $\mathbb{R}$.

El teorema que acabamos de demostrar permite afirmar que $p_A(A)=0$ para toda matriz de $M_n(\mathbb{C})$, ya que al ser toda matriz A representativa de cierto endomorfismo f, de $p_f(f)=0$ resulta que $p_A(A)=0$. Pero $\mathbb{R}\subset\mathbb{C}$, luego si $A\in M_n(\mathbb{R})$, entonces $A\in M_n(\mathbb{C})$ y $p_A(A)=0$. De aquí que si f es un endomorfismo de un espacio vectorial sobre $\mathbb{R}$, también se cumple que $p_f(f)=0$.

Nota. La teoría de los espacios vectoriales puede generalizarse al caso en que los escalares, en lugar de ser números reales o complejos, son elementos de una estructura algebraica más general llamada cuerpo. En este caso también se cumple el teorema de Hamilton-Cayley.

Hemos dicho que el teorema de Hamilton-Cayley que acabamos de demostrar nos ayudará a saber cuál es la relación existente entre el polinomio minimal w_f y el polinomio característico p_f de un endomorfismo f de un espacio vectorial E de dimensión finita sobre K y nos facilitará, por tanto, el cálculo del polinomio minimal.

Vimos que el conjunto I de los polinomios que se vinculan a un endomorfismo fijo f es de la forma $I=\{q(x)w_f(x)/q(x)\in K[x]\}$, y del teorema de Hamilton-Cayley resulta de inmediato que el polinomio característico p_f de un endomorfismo f es un múltiplo del polinomio minimal w_f. Esto es,

$$p_f=w_f\cdot q(x) \text{ para cierto } q(x)\in K[x].$$

Obsérvese que no se excluye la posibilidad $p_f=w_f$.

¿Cómo incidirá esta relación en la búsqueda de un método de cálculo para el polinomio minimal? Veamos.

Cálculo del polinomio minimal

Supongamos que f es un endomorfismo de un espacio vectorial de dimensión n sobre K y

$$p_f(x) = (-1)^n(x-\lambda_1)^{m_1}\dots(x-\lambda_k)^{m_k}$$

su polinomio característico, donde $\lambda_1, \lambda_2, \dots, \lambda_k$ son las raíces distintas de este. Entonces w_f debe ser de la forma

$$w_f(x) = (x-\lambda_1)^{r_1}\dots(x-\lambda_k)^{r_k} \text{ donde } 0 \leqslant r_i \leqslant m_i.$$

Ejemplo

El endomorfismo f de $\mathbb{R}^4$ cuya matriz en base canónica es

$$A = \begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -2 & 4 \end{pmatrix}$$

tiene el polinomio característico

$$p_f(x) = p_A(x) = (x-3)(x-2)^3.$$

Según lo que acabamos de ver, el polinomio minimal de f debe ser uno de los siguientes:

$$\begin{array}{ll} x-3 & (x-3)(x-2) \\ x-2 & (x-3)(x-2)^2 \\ (x-2)^2 & (x-3)(x-2)^3. \\ (x-2)^3 & \end{array}$$

¿Cómo comprobar cuál de ellos es el minimal?

Podemos ir probando uno a uno, comenzando por los de menor grado, hasta hallar el de grado mínimo que se anula en A, y por tanto, en f. Es decir, debemos calcular

$$A-3I_4,\ A-2I_4,\ (A-2I_4)^2,\ (A-2I_4)^3,\ \dots$$

El estudiante puede comprobar que en este caso $w_f = (x-3)(x-2)^2$.

Como vemos, aún en este caso, en el cual el polinomio característico es sencillo, pues tiene solo dos raíces, el proceso es largo y tedioso. Este proceso se puede aligerar ya que podemos comprobar que toda raíz del polinomio característico es también raíz del polinomio minimal.

PROPOSICIÓN 5.3

Sea f un endomorfismo de un espacio vectorial de dimensión n sobre K. Entonces toda raíz del polinomio característico de f es raíz de su polinomio minimal.

Demostración

Sea λ una raíz del polinomio característico p_f de f. Entonces existe $v \neq 0$ en E tal que $f(v) = \lambda v$.

Supongamos que $w_f(x) = x^p + a_{p-1} x^{p-1} + \ldots + a_0$. Como $w_f(f) = 0$, tenemos que

$$w_f(f) = f^p + a_{p-1} f^{p-1} + \ldots + a_0 id_E = 0.$$

Evaluemos $w_f(f)$ en un vector $v \neq 0$ tal que $f(v) = \lambda v$.

$$w_f(f)(v) = f^p(v) + a_{p-1} f^{p-1}(v) + \ldots + a_0 v = 0.$$

Pero $f(v) = \lambda v \Rightarrow f^2(v) = f(f(v)) = f(\lambda v) = \lambda f(v) = \lambda^2 v$ y

$$f^k(v) = f(f^{k-1}(v)) = f(\lambda^{k-1} v) = \lambda^{k-1} f(v) = \lambda^k v.$$

De aquí que

$$w_f(f(v)) = \lambda^p v + a_{p-1} \lambda^{p-1} v + \ldots + a_0 v = v(\lambda^p + a_{p-1}\lambda^{p-1} + \ldots + a_0) = 0$$

Como $v \neq 0$, resulta:

$$\lambda^p + a_{p-1}\lambda^{p-1} + \ldots + a_0 = 0.$$

Por tanto, λ es raíz de w_f, según queríamos demostrar.

Este resultado precisa el ya obtenido sobre la forma que debe tener el polinomio minimal si se conoce la del polinomio característico, ya que si f es un endomorfismo de un espacio vectorial E de dimensión n sobre K y $p_f = (-1)^n (x-\lambda_1)^{m_1} \ldots (x-\lambda_k)^{m_k}$, entonces,

$$w_f = (x-\lambda_1)^{r_1} \ldots (x-\lambda_k)^{r_k}, \quad 1 \leqslant r_i \leqslant m_i.$$

Esto facilitará en la práctica la búsqueda del polinomio minimal. Si volvemos al ejemplo dado veremos que las posibilidades de polinomio minimal para f se reducen a

$$(x-3)(x-2), \ (x-3)(x-2)^2, \ (x-3)(x-2)^3.$$

Calculemos:

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

$$(A-3I_4)(A-2I_4)^2 = \begin{pmatrix} -1 & 1 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -2 & 1 \end{pmatrix} \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 1 \\ 0 & 0 & -2 & 2 \end{pmatrix} = 0$$

En resumen, el método de cálculo del polinomio minimal de un endomorfismo f de un espacio vectorial de dimensión finita sobre K consiste en:

1ro. Buscar la matriz representativa de f en una base conveniente.

2do. Calcular p_f y expresarlo en la forma

$p_f(x) = (-1)^n(x-\lambda_1)^{m_1}\dots(x-\lambda_k)^{m_k}$.

3ro. Tomar todas las posibilidades del tipo

$(x-\lambda_1)^{r_1}\dots(x-\lambda_k)^{r_k}$, $1 \leqslant r_i \leqslant m_i$.

4to. Calcular

$(A-\lambda_1 I_n)^{r_1}\dots(A-\lambda_k I_n)^{r_k}$

para cada posibilidad anterior, comenzando por los polinomios de menor grado hasta encontrar el de grado mínimo que se anule con A.

Este método tiene la dificultad de que no siempre es posible descomponer en factores el polinomio característico de un endomorfismo, pero al menos resuelve el problema en un gran número de casos.

Si empleamos la técnica de las λ-matrices desarrollada en el capítulo anterior, el cálculo del polinomio minimal puede realizarse una vez que se conocen los factores invariantes y los divisores elementales de la matriz $A-\lambda I_n$ asociada a la matriz A ya que, como puede demostrarse, el polinomio minimal de la matriz A es el último factor invariante de la matriz $A-\lambda I_n$.

Nota: La definición de polinomio minimal es de especial interés en el caso de los espacios vectoriales de dimensión infinita. Obsérvese que aunque para la búsqueda de un método de cálculo para el polinomio minimal relacionamos este con el polinomio característico, la definición de polinomio minimal es completamente independiente de la de polinomio característico. Cuando la dimensión es infinita no podemos definir, por razones obvias, el polinomio característico de un endomorfismo, pero sí es válida la definición de polinomio minimal. Por supuesto, la demostración de la existencia del polinomio minimal que hemos dado es válida solamente para el caso de un espacio vectorial de dimensión finita.

Cuando E no es de dimensión finita, puede existir un endomorfismo que no posea polinomio minimal.

Ejemplo

Sea $E=K[x]$ y consideremos el endomorfismo f de $K[x]$ dado por $f(x^p)=x^{p+1}$.

Si f admite polinomio minimal, esto quiere decir que existe un polinomio no nulo $p(x)=a_0+a_1x+\dots+a_nx^n$, tal que $p(f)=0$,
esto es, $a_0 id_E + a_1 f + \dots + a_n f^n = 0$, donde no todos los a_i son nulos.

$$a_0 p(x) + a_1 f(p(x)) + \dots + a_n f^n(p(x)) = 0 \quad \forall p(x) \in K[x];$$

pero si tomamos $p(x)=1$, entonces

$$f(1)=x, \dots, f^p(1)=x^p.$$

De aquí que

$$a_0 + a_1 f(1) + \dots + a_n f^n(1) = a_0 + a_1 x + \dots + a_n x^n = 0,$$

donde no todos los coeficientes son nulos, lo cual es una contradicción, pues el sistema $(1, x, \dots, x^n, \dots)$ es linealmente independiente en $K[x]$.

Precisamente la importancia del polinomio minimal en el caso de los endomorfismos de espacio de dimensión infinita radica, en que en este caso el estudio de las condiciones para diagonalizar un endomorfismo, las propiedades de los valores propios y otras propiedades se realiza mediante el polinomio minimal.

5.3 Teorema de descomposición primaria de un endomorfismo de un espacio vectorial de dimensión finita

En el epígrafe 5.1 señalamos que la vía que utilizaremos para obtener la forma de Jordán para la matriz de un endomorfismo f de un espacio E de dimensión finita sobre K es la de encontrar subespacios estables por f de forma adecuada y tales que, mediante la unión de bases de dichos subespacios, se obtenga una base de E que permita representar a f de forma simple.

En dicho epígrafe vimos cómo tomando el polinomio característico de un endomorfismo y los subespacios característicos para cada valor propio, obtenemos una representación matricial en bloques diagonales. Vamos a estudiar ahora, en general, el tipo de descomposición que mostramos en el ejemplo. El problema consistirá en lo siguiente:

Si f es un endomorfismo de un K-espacio de dimensión n sobre K tal que las raíces de su polinomio característico pertenecen a K y

$$p_f(x)=(-1)^n(x-\lambda_1)^{m_1}(x-\lambda_2)^{m_2}...(x-\lambda_k)^{m_k},$$

donde $\lambda_1, \lambda_2, ..., \lambda_k$ son los valores propios distintos de f, ¿se cumplirá que

$$E=\operatorname{Ker}(f-\lambda_1 id_E)^{m_1} \oplus \operatorname{Ker}(f-\lambda_2 id_E)^{m_2} \oplus ... \oplus \operatorname{Ker}(f-\lambda_k id_{Ea)})^{m_k}? \quad (*)$$

Ya sabemos que los subespacios $\operatorname{Ker}(f-\lambda_i id_E)^{m_i}$ son estables por f. Luego si se cumple (*), podemos escoger una base de E formada por la unión de bases de los subespacios $\operatorname{Ker}(f-\lambda_i id_E)^{m_i}, i=1,...,k,$ en la cual la matriz del endomorfismo tome la forma diagonal por bloques.

Ahora bien, si en lugar del polinomio característico p_f tomamos el polinomio minimal w_f, *por ejemplo,*

$$w_f=(x-\lambda_1)^{r_1}\ (x-\lambda_2)^{r_2}...(x-\lambda_k)^{r_k},\quad 1\leqslant r_i \leqslant m_i,$$

¿obtendremos los mismos resultados que con el polinomio característico si planteamos la posibilidad

$$E=\operatorname{Ker}(f-\lambda_1 id_E)^{r_1} \oplus ... \oplus \operatorname{Ker}(f-\lambda_k id_E)^{r_k}?$$

Está claro que también podemos escoger esta opción, ya que los polinomios característico y minimal de f tienen en común el hecho de que ambos se anulan en f, y en la forma en que los hemos presentado, cada uno de ellos aparece como un producto de polinomios primos relativos, con factores del tipo $(x-\lambda_i)^{m_i}$ en el característico y del tipo $(x-\lambda_i)^{r_i}$ en el minimal.

Ahora daremos una serie de resultados sobre polinomios de endomorfismos que, por ser más generales, incluyen los casos de los polinomios característico y minimal de un endomorfismo y permiten obtener los resultados de descomposición de E mencionados anteriormente.

Los resultados que daremos serán una muestra de la interrelación entre las diferentes partes del Álgebra, ya que los obtendremos haciendo uso básicamente de las propiedades de los polinomios estudiadas anteriormente. Llamamos la atención sobre las técnicas de demostración usadas a continuación.

TEOREMA 5.2

Sean f un endomorfismo de un espacio vectorial E de dimensión n sobre K y p un polinomio de $K[x]$ tal que:

(1) $p(f)=0$

(2) $p=p_1 \cdot p_2$, donde p_1 y p_2 son primos relativos.

Entonces $\operatorname{Ker} p_1(f)$ y $\operatorname{Ker} p_2(f)$ son subespacios de E estables por f tales que

$$E=\operatorname{Ker}[p_1(f)] \oplus \operatorname{Ker}[p_2(f)].$$

Demostración

Recordemos que todo polinomio en f conmuta con f y este es el caso de $p_1(f)$ y $p_2(f)$ y que cuando dos endomorfismos conmutan, el núcleo de uno de ellos es estable por el otro.

Tenemos que probar que $E=\operatorname{Ker}[p_1(f)] \oplus \operatorname{Ker}[p_2(f)]$, esto es, que $\operatorname{Ker}[p_1(f)]$ y $\operatorname{Ker}[p_2(f)]$ son subespacios suplementarios de E.

1) Probemos que para todo v de E se cumple:

$v=v_1+v_2$ con $v_1 \in \operatorname{Ker}[p_1(f)]$, $v_2 \in \operatorname{Ker}[p_2(f)]$.

Como $p_1(x)$ y $p_2(x)$ son primos relativos, se tiene por la identidad de Bezout que existen $q_1(x)$ y $q_2(x)$ de $K[x]$ tales que

$1=p_1(x) \cdot q_1(x)+p_2(x) \cdot q_2(x)$.

Si evaluamos ambos miembros de la identidad en f, obtenemos $id_E=p_1(f) \circ q_1(f)+p_2(f) \circ q_2(f)$, y de aquí para todo v de E:

$$v=p_1(f) \circ q_1(f)(v)+p_2(f) \circ q_2(f)(v).$$

Como $p(f)=p_1(f) \circ p_2(f)=0$, no es difícil comprobar que

$$\begin{aligned} p_1(f)\,[p_2(f) \circ q_2(f)](v) &= q_2(f)\,[(p_1 \cdot p_2)(f)\,(v)] \\ &= q_2(f)(p(f)(v)) = q_2(f)(0)=0, \end{aligned}$$

es decir,

$$p_2(f) \circ q_2(f)(v) \in \operatorname{Ker}[p_1(f)].$$

Análogamente

$$p_2(f)[p_1(f)\circ q_1(f)](v)=q_1(f)[(p_2\cdot p_1)(f)(v)]=q_1(f)(p(f)(v))=q_2(f)(0)=0.$$

y por tanto,
$[p_1(f)\circ q_1(f)](v)\in \operatorname{Ker} p_2(f)$.
De aquí,
$v_1=[p_2(f)\circ q_2(f)](v)$ y $v_2=[p_1(f)\circ q_1(f)](v)$
y así hemos obtenido la descomposición $v=v_1+v_2$ buscada.

2) Probemos ahora que $\operatorname{Ker}[p_1(f)]\cap\operatorname{Ker}[p_2(f)]=\{0\}$.
Sea $v\in\operatorname{Ker}[p_1(f)]\cap\operatorname{Ker}[p_2(f)]$. Entonces $p_1(f)(v)=p_2(f)(v)=0$. Pero vimos en 1) que

$$\begin{aligned} v &=[p_1(f)\circ q_1(f)](v)+[p_2(f)\circ q_2(f)](v)\\ &=q_1(f)\ [p_1(f)(v)]+q_2(f)\ [p_2(f)\ (v)]\\ &=q_1(f)\ (0)+q_2(f)\ (0)=0. \end{aligned}$$

De aquí que
$\operatorname{Ker}[p_1(f)]\cap\operatorname{Ker}[p_2(f)]=\{0\}$
y el teorema queda demostrado.

El estudiante no debe agobiarse por los detalles de la demostración anterior, especialmente por la notación. En cuanto a esta demostración, es conveniente recordar que un polinomio en f no es más que un endomorfismo y en este sentido hemos trabajado con dichos polinomios.

Si consideramos una base B de E formada mediante la unión de una base B_1 de $\operatorname{Ker}[p_1(f)]$ y una base B_2 de $\operatorname{Ker}[p_2(f)]$, la matriz de f en dicha base será de la forma

$$A=\begin{pmatrix} A_1 & 0\\ 0 & A_2\end{pmatrix},$$

por ser $\operatorname{Ker}[p_1(f)]$ y $\operatorname{Ker}[p_2(f)]$ estables por f.

Si consideramos las aplicaciones

$f_1:\operatorname{Ker}[p_1(f)]\longrightarrow\operatorname{Ker}[p_1(f)]$ dada por $f_1(x)=f(x)\ \forall x\in\operatorname{Ker}[p_1(f)]$,
$f_2:\operatorname{Ker}[p_2(f)]\longrightarrow\operatorname{Ker}[p_2(f)]$ dada por $f_2(x)=f(x)\ \forall x\in\operatorname{Ker} p_2(f)$,

f_1 y f_2 serán endomorfismos de $\operatorname{Ker}[p_1(f)]$ y $\operatorname{Ker}[p_2(f)]$ respectivamente, a los que llamamos *restricciones* de f a $\operatorname{Ker}[p_1(f)]$ y $\operatorname{Ker}[p_2(f)]$.

A_1 es la matriz de f_1 en la base B_1 de $\operatorname{Ker}[p_1(f)]$ y A_2 la de f_2 en la base B_2 de $\operatorname{Ker}[p_2(f)]$.

Las características de estas restricciones serán de interés para nosotros, por lo cual demostraremos las siguientes proposiciones.

PROPOSICIÓN 5.4

Sean E un espacio vectorial de dimensión finita sobre K y $E=E_1\oplus E_2$, donde E_1 y E_2 son subespacios de E estables por un endomorfismo f de E. Si denotamos por

f_1 la restricción de f a E_1, esto es, $f_1 : E_1 \longrightarrow E_1$ dada por

$$f_1(v) = f(v) \ \forall v \in E_1$$

f_2 la restricción de f a E_2, esto es, $f_2 : E_2 \longrightarrow E_2$ dada por

$$f_2(v) = f(v) \ \forall v \in E_2,$$

entonces,

(1) w_f es el mínimo común múltiplo de w_{f_1} y w_{f_2}.

(2) p_f es el producto de p_{f_1} y p_{f_2}.

Demostración

(1) Recordemos la definición de mínimo común múltiplo de dos polinomios $p_1(x)$ y $p_2(x)$.

Se dice que $m(x)$ es un mínimo común múltiplo de $p_1(x)$ y $p_2(x)$ si y solo si $m(x)$ es múltiplo de $p_1(x)$, $m(x)$ es múltiplo de $p_2(x)$, y si $n(x)$ es múltiplo de $p_1(x)$ y de $p_2(x)$ entonces es múltiplo de $m(x)$. De entre todos, tomamos el mónico y lo denotamos por mcm $(p_1(x), p_2(x))$.

Utilizaremos la definición anterior para probar que $w_f = \text{mcm}\ (w_{f_1}, w_{f_2})$.

Recordemos, además, que según la definición de polinomio minimal de un endomorfismo f, este es el polinomio de menor grado que se anula en dicho endomorfismo y los otros polinomios que se anulan en f son exactamente los múltiplos del minimal, de modo que si probamos que w_f se anula en f_1 y f_2, quedará probado que w_f es múltiplo de w_{f_1} y w_{f_2}.
En efecto,

$$w_f(f_1)(v) = w_f(f(v)) = 0 \quad \forall v \in E_1 \Longleftrightarrow w_f(f_1) = 0,$$
$$w_f(f_2)(v) = w_f(f(v)) = 0 \quad \forall v \in E_2 \Longleftrightarrow w_f(f_2) = 0.$$

Falta probar que todo múltiplo de w_{f_1} y de w_{f_2} es un múltiplo de w_f.

Sea $p(x)$ un múltiplo de w_{f_1} y de w_{f_2}. Usaremos el mismo procedimiento anterior para probar que $p(x)$ es múltiplo de w_f, es decir, debemos probar que $p(f) = 0$, $p(f) = 0 \Longleftrightarrow p(f)(v) = 0 \ \forall v \in E$

Pero $v = v_1 + v_2$ con $v_1 \in E_1$, $v_2 \in E_2$, luego

$$\begin{aligned} p(f)(v) &= p(f)(v_1) + p(fv_2) \\ &= p(f_1)(v_1) + p(f_2)(v_2) \end{aligned}$$

y como $p(x)$ es un múltiplo de w_{f_1} y de w_{f_2}, entonces

$$p(f_1) = 0 \Longleftrightarrow p(f_1)(v) = 0 \quad \forall v \in E_1$$
$$p(f_2) = 0 \Longleftrightarrow p(f_2)(v) = 0 \quad \forall v \in E_2$$

Por tanto $p(f)(v) = 0 \ \forall v \in E$ según queríamos demostrar, y como w_f es mónico, entonces $w_f = \text{mcm}\ (w_{f_1}, w_{f_2})$.

(2) En el caso del polinomio característico su cálculo depende de una representación matricial de f, y las condiciones de E_1 y E_2 permiten obtener una representación matricial simple y asociada a las restricciones f_1 y f_2. Tomemos, pues, una base de E formada por la unión de una base B_1 de E_1 y una

base B_2 de E_2 y busquemos la matriz de f en dicha base. Esta será

$$A=\begin{pmatrix} A_1 & 0 \\ 0 & A_2 \end{pmatrix},$$

donde $A_1=M(f_1, B_1)$, $A_2=M(f_2, B_2)$.
Entonces,

$$A-\lambda I_n=\begin{pmatrix} A_1-\lambda I_{n_1} & 0 \\ 0 & A_2-\lambda I_{n_2} \end{pmatrix}, \quad n_1+n_2=n$$

y det $(A-\lambda I_n)=\det(A_1-\lambda I_{n_1}) \cdot \det(A_2-\lambda I_{n_2})$, de donde resulta que

$$p_f=p_{f_1} \cdot p_{f_2}.$$

Del teorema 5.2 y la proposición 5.4 podemos llegar a la conclusión siguiente:

PROPOSICIÓN 5.5

Sean f un endomorfismo de un espacio vectorial E de dim n sobre K, w_f su polinomio minimal. Sea $w_f=p_1 \cdot p_2$, donde p_1 y p_2 son mónicos y primos relativos. Entonces si f_1 es la restricción de $[p_1(f)]$ y f_2 la restricción de f a Ker $[p_2 f]$, se tiene:

$$w_{f_1}=p_1 \text{ y } w_{f_2}=p_2.$$

Demostración

Como f_1 y f_2 son las restricciones de f a Ker $[p_1(f)]$ y Ker $[p_2(f)]$, respectivamente, resulta:

$$p_1(f_1)=0$$

ya que

$$p_1(f_1): \text{Ker } [p_1(f)] \longrightarrow \text{Ker } [p_1(f)] \text{ y}$$
$$p_1(f_1)(v)=p_1(f)(v)=0 \quad \forall v \in \text{Ker } p_1(f)$$

Análogamente se demuestra que $p_2(f_2)=0$.

Por tanto, p_1 es un múltiplo de w_{f_1} y p_2 es un múltiplo de w_{f_2} y w_{f_1} y w_{f_2} son primos, pues de lo contrario p_1 y p_2 no podrían serlo.

Según la proposición 5.4,

$$w_f=\text{mcm }(w_{f_1}, w_{f_2}),$$

y utilizando el hecho de que mcd $(w_{f_1}, w_{f_2})=1$, obtenemos

$$w_f=\text{mcm }(w_{f_1}, w_{f_2})=w_{f_1} \cdot w_{f_2}.$$

Entonces

$$p_1 \cdot p_2=w_{f_1} \cdot w_{f_2}.$$

Como $p_1=q_1 \cdot w_{f_1}$ y $p_2=q_2 \cdot w_{f_2}$ para ciertos q_1 y q_2, y w_1, w_2, p_1 y p_2 son mónicos, tenemos que $q_1 \cdot q_2=1$, y de aquí

$$p_1=w_{f_1} \text{ y } p_2=w_{f_2},$$

según queríamos demostrar.

Analicemos ahora qué ocurre con los polinomios característicos de f y los de las restricciones a Ker $[p_1(f)]$ y Ker $[p_2(f)]$.

PROPOSICIÓN 5.6

Sean f un endomorfismo de un espacio vectorial de dimensión n sobre K y p_f su polinomio característico. Si todas las raíces de p_f están en K, $p_f=p_1 \cdot p_2$ con p_1 y p_2 primos relativos, y f_1 y f_2 son las restricciones de f a Ker $[p_1(f)]$ y Ker $[p_2(f)]$ respectivamente, entonces,

$$p_{f_1}=ap_1 \text{ y } p_{f_2}=\frac{1}{a}\, p_2, \quad a\in K.$$

Demostración

Según la proposición 5.4, $p_f=p_{f_1} \cdot p_{f_2}$, de aquí que las raíces de p_{f_1} y p_{f_2} sean raíces de p_f y deban estar en K por estar las de p_f. Luego si λ es una raíz de p_{f_1}, es decir, existe $v\neq 0$, $v\in$Ker $[p_1(f)]$ tal que $f_1(v)=\lambda v$, λ debe ser raíz de p_f, o sea, $f(v)=\lambda v$, o lo que es lo mismo, $(f-\lambda id_E)\ (v)=0$.

Por otra parte, λ es raíz de $p_1(x)-p_1(\lambda)$, luego $p_1(x)-p_1(\lambda)$ es divisible por $x-\lambda$ y podemos escribir para cierto $q(x)\in K[x]$:

$$p_1(x)-p_1(\lambda)=(x-\lambda)\ q(x).$$

De la igualdad anterior resulta que

$$p_1(f)-p_1(\lambda)\, id_E=(f-\lambda id_E)\circ idq(f).$$

Si evaluamos en el elemento v de Ker $[p_1(f)]$ tal que $f_1(v)=f(v)=\lambda v$ obtenemos:

$$p_1(f)(v)-p_1\ (\lambda)\, v=q(f)(f-\lambda id_E)\ (v)=0.$$

Además de $p_1(f)(v)=0$, ya que $v\in$Ker $[p_1(f)]$, se llega a $p_1(\lambda)\, v=0$ y como $v\neq 0$, $p_1\ (\lambda)=0$.

Por tanto, toda raíz de p_{f_1} es una raíz de p_1, de donde p_{f_1} es un divisor de p_1 (y no de p_2 ya que p_1 y p_2 son primos relativos).

De manera similar se demuestra que p_{f_2} es un divisor de p_2, y como

$$p=p_1 \cdot p_2=p_{f_1} \cdot p_{f_2}$$

y p_{f_1} y p_{f_2} son primos relativos por serlo p_1 y p_2, existe $a\in K^*$ tal que

$$p_{f_1}=ap_1 \text{ y } p_{f_2}=\frac{1}{a}\, p_2.$$

El teorema y las proposiciones que acabamos de enunciar y demostrar pueden generalizarse al caso en que p, así como w_f y p_f sean un producto finito de polinomios primos entre sí.

TEOREMA 5.3

Sea f un endomorfismo de un espacio vectorial de dimensión n sobre K. Sean $p(x)$ un polinomio de $K[x]$ tal que $p(f)=0$ y $p_1, p_2, \ldots, p_m$ polinomios primos relativos tales que $p=p_1 \cdot p_2 \cdot \ldots \cdot p_m$. Entonces,

(1) Los subespacios Ker $[p_1(f)]$, Ker $[p_2(f)]$, ..., Ker $[p_m(f)]$ son estables por f y $E=\text{Ker }[p_1(f)] \oplus \text{Ker }[p_2(f)] \oplus \ldots \oplus \text{Ker }[p_m(f)]$.

(2) Si $p=w_f$ es el polinomio minimal de f y los p_1 son unitarios, entonces para las restricciones $f_1, \ldots, f_m$ a Ker $[p_1(f)], \ldots,$ Ker $[p_m(f)]$, respectivamente, se tiene:

$$w_{f_1}=p_1, \ldots, w_{f_m}=p_m.$$

(3) Si $p=p_f$ es el polinomio característico de f y p_f tiene todas sus raíces en K, para las restricciones $f_1, \ldots, f_m$ se tiene que

$$p_{f_1}=a_1 p_1, \ldots, p_{f_m}=a_m p_m,$$

donde $a_1, \ldots, a_m$ son constantes no nulas.

Demostración

La demostración se realiza por inducción en m. Haremos en detalle la de (1) y dejaremos al estudiante que realice de forma análoga la de (2) y (3).

Para $m=2$ la propiedad está demostrada por el teorema 5.2.

Supongamos que se cumple para $m-1$ polinomios, esto es, si $p=p_1, p_2, \ldots, p_{m-1}$ con las demás condiciones de la hipótesis, entonces Ker $[p_i(f)]$ es estable para $i=1, \ldots, m-1$ y

$$E=\text{Ker }[p_1(f)] \oplus \text{Ker }[p_2(f)] \oplus \ldots \oplus \text{Ker }[p_{m-1}(f)].$$

Probemos que se cumple para m.

Sea $p=p_1 \cdot p_2 \cdot \ldots \cdot p_m$ con $p_1, p_2, \ldots, p_m$ primos relativos y $p(f)=0$. Entonces podemos escribir:

$$p=(p_1 \cdot p_2 \cdot \ldots \cdot p_{m-1}) \cdot p_m = q \cdot p_m.$$

Del hecho de que los p_i sean primos relativos resulta que p_m y q son primos relativos; de aquí que

$$E=\text{Ker }[q(f)] \oplus \text{Ker }[p_m(f)] \ (*).$$

Ker $[q(f)]$ y Ker $[p_m(f)]$ son estables por f; además, Ker $q(f)$ es un subespacio de E estable por f.

Luego consideramos

f_1 : Ker$[q(f)]= \longrightarrow$Ker$[q(f)]$ restricción de f a Ker $[q(f)]$ y dado por $f_1(v)=f(v)$ $\forall v \in$ Ker $[q(f)]$. Entonces

$$q(f_1)=0 \text{ ya que } q(f_1)(v)=q(f)(v)=0 \quad \forall v \in \text{Ker }[q(f)].$$

Como $q=p_1 \cdot p_2 \cdot \ldots \cdot p_{m-1}$, podemos aplicarle la hipótesis de inducción y obtenemos:

$$\text{Ker}\,[q(f)]=\text{Ker}\,[p_1(f)]\oplus\text{Ker}\,[p_2(f)]\oplus\ldots\oplus\text{Ker}\,[p_m(f)],$$

donde cada Ker $[p_i\,(f)]$ es estable por f_1, y por tanto, por f.

Sustituyendo esta expresión para Ker $[q(f)]$ en (*), obtenemos:

$$E=\text{Ker}\,[p_1(f)]\ \oplus\text{Ker}\,[p_2(f)]\oplus\ldots\oplus\ \text{Ker}\,[p_m(f)]$$

según queríamos demostrar.

Llamamos la atención del estudiante sobre la demostración que acabamos de hacer, que le proporciona el método a seguir para demostrar las partes (2) y (3) del teorema. Obsérvese que aunque se trata de un proceso de inducción común, hay que ser cuidadoso en los pasos intermedios y comprobar que se cumplen las hipótesis planteadas.

Los teoremas y las proposiciones que acabamos de estudar nos permitirán de inmediato avanzar un paso más en la búsqueda de la matriz de Jordán mediante el teorema siguiente:

TEOREMA 5.4 (Teorema de descomposición primaria de un endomorfismo)

Sea f un endomorfismo de un espacio vectorial E de dimensión n sobre K tal que todas las raíces de p_f estén en K. Entonces·

(1) Si $p_f(x)=(-1)^n\,(x-\lambda_1)\,\ldots\,(x-\lambda_k)^{m_k}$, donde $\lambda_1,\ldots,\lambda_k$ son los valores propios distintos de f, y si
$N_1=\text{Ker}(f-\lambda_1\,id_E)^{m_1},\ldots,N_k=\text{Ker}\,(f-\lambda_k\,id_E)^{m_k}$,
entonces cada N_i es estable por f y $E=N_1\oplus N_2\oplus\ldots\oplus N_k$.
Además, la restricción f_i de f a N_i tiene a $(x-\lambda_i)^{m_i}$ como polinomio característico.

(2) Si $w_f(x)=(x-\lambda_1)^{r_1}\,(x-\lambda_2)^{r_2}\ldots\,(x-\lambda_k)^{r_k}$ es el polinomio minimal de f y $M_1=\text{Ker}\,(f-\lambda_1\,id_E)^{r_1},\ldots,M_k=\text{Ker}\,(f-\lambda_k id_E)^{r_k}$, entonces cada M_i es estable por f y
$E=M_1\oplus M_2\oplus\ldots\oplus M_k$.
Además, la restricción f_i *de* f a M_i tiene a $(x-\lambda_i id_E)^{r_i}$ como polinomio minimal.

Demostración

El teorema se demuestra considerando en el teorema 5.3, $p=p_f$ para (1) y $p=w_f$ para (2). Dichos polinomios cumplen en ambos casos las hipótesis del teorema.

Demostremos (1).

Tomemos $p=p_f$, $p_i=(x-\lambda_i)^{m_i}$. Entonces los p_i son primos relativos, y resulta del teorema 5.3 que los N_i son estables por f, y $E=N_1\oplus\ldots\oplus N_k$ y las restricciones f_i de f a N_i tienen como polinomio característico $(x-\lambda i)^{m_i}$.

El estudiante puede demostrar (2) tomando $p=w_f$ y aplicando el teorema 5.3.

Así queda demostrado el teorema de descomposición primaria.

Ahora vamos a extraer algunas conclusiones de este teorema y a explicar cómo se utiliza en la práctica.

En primer lugar, hemos encontrado dos descomposiciones distintas de E mediante los subespacios N_i y M_i ¿qué relación existe entre dichas descomposiciones?

Es fácil comprobar que

$$M_i=\operatorname{Ker}\,(f-\lambda_i\, id_E)^{n_i} \subset N_i=\operatorname{Ker}(f-\lambda_i id_E)^{m_i},$$

lo cual dejamos para que sea verificado por el estudiante.

Como además $E=N_1\oplus N_2\oplus\ldots\oplus N_k=M_1\oplus M_2\oplus\ldots\oplus M_k$, está claro que de la definición de suma directa resulta que

$$N_i=M_i \text{ para } i=1,\ldots, k.$$

¿Cómo se expresa este resultado en la práctica?

Evidentemente, el hecho de que ambas descomposiciones de E obtenidas una a partir de p_f y otra a partir de w_f sean iguales, permite escoger indistintamente una vía u otra con los mismos resultados. Cuando realicemos un ejemplo práctico señalaremos las ventajas y desventajas de cada vía.

Por supuesto que la matriz de f con respecto a una base B formada por unión de las bases B_i *de los* N_i(o M_i) es de la forma

$$M(f,B)=\begin{pmatrix} A_1 & 0 & \ldots & 0\\ 0 & A_2 & \ldots & 0\\ \cdot & & & \cdot\\ \cdot & & & \cdot\\ \cdot & & & \cdot\\ \cdot & & & \cdot\\ 0 & 0 & & A_k \end{pmatrix}$$

y cada A_i es la matriz de la restricción f_i de f a N_i (o de la restricción f_i *de* f a M_i) en las bases B_i.

Ejemplo

Sea f el endomorfismo de $\mathbb{R}^5$ cuya matriz en la base canónica de dicho espacio es

$$A=\begin{pmatrix} 1 & 1 & -1 & 2 & -1\\ 2 & 0 & 1 & -4 & -1\\ 0 & 1 & 1 & 1 & 1\\ 0 & 1 & 2 & 0 & 1\\ 0 & 0 & -3 & 3 & -1 \end{pmatrix}$$

Para este endomorfismo,

$$p_f(x)=p_A(x)=-(x+1)^2\,(x-1)^3.$$

Entonces por el teorema de descomposición primaria tenemos que

$$\mathbb{R}^5=\operatorname{Ker}\,(f+id_{\mathbb{R}^5})^2 \oplus \operatorname{Ker}\,(f-id_{\mathbb{R}^5})^3.$$

Calculemos Ker $(f+id_{\mathbb{R}^5})^2$ y Ker $(f-id_{\mathbb{R}^5})^3$:

$$(A+I_5)^2=\begin{pmatrix} 6 & 4 & 4 & -2 & -2 \\ 6 & 0 & -4 & -6 & -6 \\ 2 & 4 & 4 & 2 & 2 \\ 2 & 4 & 4 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

$$(A-I_5)^3=\begin{pmatrix} 0 & 0 & -28 & 28 & -8 \\ 0 & 0 & 28 & -28 & 8 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & -8 & 0 \\ 0 & 0 & -36 & 36 & -8 \end{pmatrix}$$

Al resolver el sistema

$$\begin{pmatrix} 6 & 4 & 4 & -2 & 2 \\ 6 & 0 & -4 & -6 & -6 \\ 2 & 4 & 4 & 2 & 2 \\ 2 & 4 & 4 & 2 & 2 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

encontramos:

$$\text{Ker } (f+id_{\mathbb{R}^5})^2=\langle\langle(1,-1,0,0,1),\ (1,-1,0,1,0)\rangle\rangle,$$

y al resolver

$$\begin{pmatrix} 0 & 0 & -28 & 28 & -8 \\ 0 & 0 & 28 & -28 & 8 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 8 & -8 & 0 \\ 0 & 0 & -36 & 36 & -8 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \\ x_5 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

resulta

$$\text{Ker}(f-I_5)^3=\langle\langle(1,0,0,0,0),\ (0,1,0,0,0),\ (0,0,1,1,0)\rangle\rangle.$$

Tomando en $\mathbb{R}^5$ la base

$(b_1=1,-1,0,0,1)$, $b_2=(1,-1,0,1,0)$, $b_3=(1,0,0,0,0)$, $b_4=(0,1,0,0,0)$, $b_5=(0,0,1,1,0))$,

la matriz de f en la base $(b_1\ b_2,\ b_3,\ b_4,\ b_5)$ es una matriz diagonal por bloques:

$$M(f,(b_i\ ))=\left(\begin{array}{cc:ccc} -1 & 3 & & 0 & \\ 0 & -1 & & & \\ \hdashline & & 1 & 1 & 1 \\ & 0 & 2 & 0 & -3 \\ & & 0 & 1 & 2 \end{array}\right)$$

donde,

$A_1=\begin{pmatrix} -1 & 3 \\ 0 & -1 \end{pmatrix}$ es la matriz de la restricción f_1 de f a $(\mathrm{Ker}(f+id_{\mathbf{R}^5})^2$

$A_2=\begin{pmatrix} 1 & 1 & 1 \\ 2 & 0 & -3 \\ 0 & 1 & 2 \end{pmatrix}$ es la matriz de la restricción f_2 de f a $\mathrm{Ker}(f-id_{\mathbf{R}^5})^3$.

Hubiéramos podido escoger la vía de utilizar el polinomio minimal de f en lugar del polinomio característico, aunque en este caso

$$w_f(x)=(x+1)^2(x-1)^3=p_f(x).$$

La vía del polinomio minimal tiene la ventaja aparente sobre la del polinomio característico, de que al ser minimal, en general, de menor grado que el característico, el número de productos de matrices a realizar para el cálculo de los subespacios $\mathrm{Ker}(f-\lambda_i\, id_E)^{r_i}$ debe ser menor. Decimos que esta ventaja es aparente porque el cálculo del propio polinomio minimal requiere un número de productos de matrices que, en general, equipara el volumen de cálculo por ambas vías.

Una aplicación inmediata del teorema de descomposición primaria es la siguiente caracterización de los endomorfismos diagonalizables mediante el polinomio minimal, la cual resulta de gran utilidad.

TEOREMA 5.5

Sea f un endomorfismo de un K-espacio E de dimensión finita sobre K. Entonces f es diagonalizable si y solo si su polinomio minimal w_f se descompone completamente como un producto de factores de primer grado en K.

Demostración

Obsérvese que del propio enunciado del teorema resulta obvio que se cumple que las raíces de p_f están en K.

Sea K diagonalizable y $\lambda_1, \lambda_2, \ldots, \lambda_k$ los valores propios distintos de f. Probemos que

$$w_f=(x-\lambda_1)(x-\lambda_2)\ldots(x-\lambda_k).$$

Como f es diagonalizable, tenemos:

$$E=E(\lambda_1)\oplus E(\lambda_2)\oplus\ldots\oplus E(\lambda_k).$$

Bastará probar, entonces, que

$$(f-\lambda_1 id_E)\circ(f-\lambda_2 id_E)\circ\ldots\circ(f-\lambda_k id_E)(v)=0 \quad \forall v\in E,$$

ya que es obvio que el polinomio anterior es el de grado mínimo que posee todas las raíces del polinomio característico de f. En particular, bastará probar que para cualquier base $(b_1, b_2, \ldots, b_n)$ de E, se cumple:

$$[(f-\lambda_1 id_E)\circ(f-\lambda_2 id_E)\circ\ldots\circ(f-\lambda_k id_E)](b_i)=0,\ i=1,\ldots,n.$$

Tomemos una base B de E formada por la unión de las bases B_i de los $E(\lambda_i)$, $i=1,\dots,k$.

Sea b_i un vector de B_i, esto es, un vector propio de f asociado al valor propio λ_i. Entonces podemos escribir:

$$(f-\lambda_1 id_E)\circ\dots\circ(f-\lambda_{i-1}id_E)\circ(f-\lambda_{i+1}id_E)\circ\dots\circ(f-\lambda_R id_E)\circ(f-\lambda_i id_E)\ (b_i)=$$
$$=[(f-\lambda_1 id_E)\circ\dots\circ(f-\lambda_k\ id_E)](f-\lambda_i\ id_E)(b_i)=0$$

ya que $(f-\lambda_i id_E)\ (b_i)=f(b_i)-\lambda_i b_i\ =0\ \forall i=1,\dots,k$. De aquí que

$$w_f(x)=(x-\lambda_1)\ \dots(x-\lambda_k).$$

Recíprocamente, supongamos que $w_f\ (x)=(x-\lambda_1)\dots(x-\lambda_k)$ y probemos que f es diagonalizable.

Por el teorema de descomposición primaria tenemos que

$$E=\mathrm{Ker}(f-\lambda_1 id_E)\oplus\mathrm{Ker}(f-\lambda_2 id_E)\oplus\dots\oplus\mathrm{Ker}(f-\lambda_k)id_E=E(\lambda_1)\oplus\dots\oplus E(\lambda_k).$$

Luego f es diagonalizable y el teorema queda demostrado.

Realmente hemos usado el teorema de descomposición primaria solamente en una de las implicaciones del teorema anterior, pero esto puso en evidencia la utilidad del enunciado de dicho teorema en términos del polinomio minimal.

En cuanto a la otra implicación, esta puede demostrarse utilizando solamente la definición de polinomio minimal; sin embargo, en la demostración se usa un recurso muy útil que el estudiante no debe olvidar: la posibilidad de tomar en el espacio una base formada por vectores propios de un endomorfismo f.

El valor de la caracterización dada por el teorema en un ejemplo práctico dependerá de si en dicho ejemplo resulta sencilla o no la búsqueda del polinomio minimal, o si por las características de dicho ejemplo es necesaria de todas formas la búsqueda de dicho polinomio.

La matriz diagonal por bloques que hallamos en el ejemplo anterior es una matriz bastante sencilla, pero no olvidemos que queremos encontrar una matriz aún más simple. Además, los elementos obtenidos en los bloques dependen de la base seleccionada en los subespacios, y ya hemos analizado que lo idóneo es encontrar una representación matricial del endomorfismo que dependa solamente de este y no de la base tomada en el espacio.

En el ejemplo dado, ¿qué faltaría para tener al menos una matriz triangular? Evidentemente si pudiéramos obtener en cada subespacio una base en la que la matriz de la restricción de f a dicho subespacio tomara la forma triángular, la unión de las bases de cada subespacio daría una base del espacio en la cual la matriz de f sería una matriz diagonal por bloques, pero cada bloque y, por tanto, la matriz serían triangulares.

Está claro que el análisis que hicimos en el ejemplo es válido en el caso general de una descomposición

$$E=\mathrm{Ker}(f-\lambda_1 id_E)^{m_1}\oplus\dots\oplus\mathrm{Ker}\ (f-\lambda_k\ id_E)^{m_k}.$$

ya que la estabilidad de los subespacios permite que los cambios de bases hechos en estos no afecten la forma diagonal por bloques de la matriz representativa de f en una base formada por la unión de las bases de los $\operatorname{Ker}(f-\lambda_i id_E)^{m_i}$.

Por otra parte, en la demostración del teorema de Hamilton-Cayley dejamos pendiente la demostración de la posibilidad de encontrar, para endomorfismos de un K-espacio E cuyo polinomio característico tenga sus raíces en K, una base en la cual la matriz de dicho endomorfismo fuese triangular. Pasemos pues a enunciar y demostrar dicho resultado.

TEOREMA 5.6

Sea f un endomorfismo de un espacio E de dimensión n sobre K tal que su polinomio característico tiene todas las raíces en K. Entonces existe una base $(a_1, a_2, \dots, a_n)$ de E en la cual la matriz T *de* f es triangular superior y los elementos de la diagonal de T son los valores propios de f.

Demostración

Haremos la demostración por inducción en $n = \dim E$.

Para $n=1$, el teorema es evidente ya que los endomorfismos de un espacio vectorial E de dimensión 1 son las homotecias y su matriz, de tamaño 1×1, es triangular en cualquier base de E.

Supongamos que la propiedad se cumple para $\dim E = n-1$, esto es, que para todo endomorfismo de un K-espacio vectorial de dimensión $n-1$ cuyo polinomio característico tenga sus raíces en K, existe una base de dicho espacio en la cual la matriz del endomorfismo sea triangular superior.

Probemos la propiedad para $\dim E = n$.

Sean f un endomorfismo de E cuyo polinomio característico tiene sus raíces en K y λ_1 una de las raíces de dicho polinomio, es decir, un valor propio de f. Entonces existe un vector a_1 de E, $a_1 \neq 0$, tal que

$$f(a_1) = \lambda_1 a_1.$$

Consideremos el subespacio de E, $Ka_1 = \{ka_1/k \in K\}$, y un suplementario de Ka_1; sea $(a'_2, \dots, a'_n)$ una base de dicho suplementario y busquemos la matriz de f en la base $(a_1, a'_2, \dots, a'_n)$ de E.

$$A = M(f, (a_1, a'_2, \dots, a'_n)) = \begin{pmatrix} \lambda_1 & a_{12} \dots a_{1n} \\ 0 & a_{22} \dots a_{2n} \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ \cdot & \cdot \quad\quad \cdot \\ 0 & a_{n2} \dots a_{nn} \end{pmatrix}$$

Una vez encontrada la matriz anterior, no es difícil suponer que fue buscada con el objetivo de poder aplicar la hipótesis de inducción, ya que el bloque

$$A'=\begin{pmatrix} a_{22} & \cdots\cdots & a_{2n} \\ \cdots & \cdots & \cdots \\ a_{n2} & \cdots\cdots & a_{nn} \end{pmatrix}$$

de tamaño $n-1$ representa un endomorfismo de un espacio de dimensión $n-1$, el cual, una vez comprobado que cumple las hipótesis del teorema y establecida su relación con el endomorfismo f, se puede representar en forma triangular y obtener una representación triangular para f. Veamos.

Primero debemos esclarecer a qué endomorfismo representa A'.

Teníamos la descomposición

$$E=Ka_1\oplus\langle a_2', \ldots, a_n'\rangle.$$

Está claro que en general $E'=\langle a_2',\ldots, a_n'\rangle$ no es estable por f, pero si consideramos $p\circ f: E\longrightarrow E'$, donde p es la proyección de E sobre E', es decir, la aplicación dada por

$$p\,(x_1a_1+x_2a_2'+\ldots+x_na_n')\ =x_2a_2'+\ldots+x_na_n',$$

entonces si $f(x)=y_1a_1+y_2a_2'\ +\ldots+y_na_n'$, se tiene $(p\circ f)\ (x)=y_2a_2'+\ldots+y_na_n'$.

Evidentemente E' es estable por $p\circ f$ y en particular

$$(p\circ f)\ (a_i')\ =a_{2i}a_2'+\ldots+a_{ni}a_n',\quad i=2,\ldots,n.$$

Si consideramos el endomorfismo f' de E' dado por

$$f'(x)=(p\circ f)\ (x)\quad \forall x\in E',$$

la matriz de f' en la base $(a_2',\ldots, a_n')$ será

$$M(f',\ (a_2',\ldots, a_n'))\ =A'=\begin{pmatrix} a_{22} & \cdots\cdots & a_{2n} \\ \cdots & \cdots & \cdots \\ a_{n2} & \cdots\cdots & a_{nn} \end{pmatrix}$$

Luego, f' es un endomorfismo de un espacio de dimensión $n-1$ al cual representa A'.

Veamos si f' cumple las hipótesis del teorema y podemos aplicarle la hipótesis de inducción. Para ello será necesario comprobar que el polinomio característico de f' tiene todas sus raíces en K.

De la expresión de la matriz A resulta que $p_f(x)=p_A(x)=\det\,(A-xI_n)$ $=(\lambda_1-x)\ \det(A'-xI_{n-1})=(\lambda_1-x)\,p_{f_1}(x)$
y como p_f tiene todas sus raíces en K, podemos afirmar que p_{f_1} tendrá sus raíces en K y aplicarle la hipótesis de inducción a f'.

Existe entonces en E' una base $(a_2,\ldots, a_n)$ tal que

$$M(f',(a_2,\ldots, a_n))=\begin{pmatrix} \beta_{22} & \cdots\cdots & & \beta_{2n} \\ 0 & \beta_{33} & \cdots\cdots & \beta_{3n} \\ \cdots & \cdots & \cdots & \cdots \\ 0 & \cdots\cdots & & \beta_{nn} \end{pmatrix}$$

Tomemos ahora la base $(a_1, a_2, \ldots, a_n)$ de E y busquemos la matriz de f en dicha base. Tenemos:

$$
\begin{aligned}
f(a_1) &= \lambda_1 a_1 \\
f(a_2) &= \beta_{12}\, a_1 + \beta_{22} a_2 + \ldots + \beta_{n2} a_n \\
&\;\;\vdots \\
f(a_n) &= \beta_{1n} a_1 + \beta_{2n} a_2 + \ldots + \beta_{nn} a_n
\end{aligned}
$$

Las expresiones de $f(a_2), \ldots, f(a_n)$ difieren de las de $f'(a_2), \ldots, f'(a_n)$ en el término correspondiente al vector a_1 de la base, ya que ahora estamos trabajando en E y con la base $(a_1, a_2, \ldots, a_n)$. Luego

$$
B = M(f, (a_1, a_2, \ldots, a_n)) = \begin{pmatrix} \lambda_1 & \beta_{12} & \ldots & \beta_{1n} \\ 0 & \beta_{22} & \ldots & \beta_{2n} \\ \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & \ldots & \beta_{nn} \end{pmatrix}
$$

que es una representación triangular de f.

Solo falta verificar que los elementos de la diagonal son exactamente los valores propios de f. Esto resulta, como ya señalamos en la demostración del teorema de Hamilton-Cayley, del hecho de que

$$
p_f(x) = p_B(x) = (\lambda_1 - x)(\beta_{22} - x) \ldots (\beta_{nn} - x)
$$

dada la forma triangular de B, y de ahí resulta de manera inmediata que las raíces de p_f y, por tanto, los valores propios de f son exactamente λ_1, $\beta_{22}, \ldots, \beta_{nn}$. De este modo, el teorema queda demostrado.

Ejemplo

Sea f el endomorfismo de $\mathbb{R}^3$ representado en la base canónica de dicho espacio por la matriz

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

Busquemos una base en la cual la matriz de f sea triangular superior. Veamos si podemos aplicar el teorema 5.6 verificando si se cumplen las condiciones dadas en él.

Tenemos que

$$
p_f(x) = p_A(x) = (2 - x)^3 \text{ y } E_2 = \langle (0,1,0) \rangle.
$$

f no es diagonizable y podemos aplicar el teorema ya que el único valor propio de f es $2 \in \mathbb{R}$.

La demostración del teorema da el método a seguir pues la inducción de la demostración se expresa en la práctica mediante un proceso iterativo. Tomemos la base formada por

$$
a_1 = (0,1,0) = e_2,\; a_2' = (1,0,0) = e_1,\; a_3' = (0,0,1) = e_3.
$$

Entonces

$$M(f, (a_1, a_2', a_3')) = \begin{pmatrix} 2 & 1 & -1 \\ 0 & 2 & 0 \\ 0 & 1 & 2 \end{pmatrix},$$

ya que

$$f(a_1) = 2a_1,$$
$$f(a_2') = f(e_1) = 2e_1 + e_2 + e_3 = 2a_2' + a_1 + a_3',$$
$$f(a_3') = f(e_3) = -e_2 + 2e_3 = -a_1 + 2a_3'.$$

El bloque $A' = \begin{pmatrix} 2 & 0 \\ 1 & 2 \end{pmatrix}$ representa el endomorfismo f' de $\langle a_2', a_3' \rangle$, cuyo polinomio característico es

$$p_f'(x) = p_A(x) = (x-2)^2$$

y al cual le podemos volver a aplicar el método; es decir, tomar un vector propio, completar una base del espacio y hallar la matriz en dicha base. Busquemos el subespacio propio en este caso

$$E_2 = \{(x, y) / x = 0\} = \langle a_3' \rangle$$

y consideremos la base (a_1, a_2, a_3), donde $a_2 = a_3'$, $a_3 = a_2'$.
Obtenemos:

$$T = M(f, \ (a_1, a_2, a_3)) = \begin{pmatrix} 2 & -1 & 1 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix},$$

ya que

$$f(a_1) = 2a_1,$$
$$f(a_2) = f(a_3') = -a_1 + 2a_3' = -a_1 + 2a_2,$$
$$f(a_3') = f(a_2') = a_1 + 2a_2' + a_3' = a_1 + 2a_3 + a_2.$$

La matriz T es triangular y no tenemos que continuar el proceso.

Obsérvese que en la diagonal de dicha matriz, como hemos demostrado, aparecen los valores propios de f.

Queremos hacer algunos comentarios acerca del ejemplo que acabamos de presentar.

Dado el tamaño de la matriz, el cual fue escogido para evitar cálculos laboriosos, el método de triangulización consistió solamente en dos pasos. En una matriz de tamaño mayor puede ocurrir que haya que hacer más de una iteración, por lo que recomendamos al estudiante que tenga en cuenta en cada paso a qué endomorfismo representa el bloque con el cual esté trabajando y el subespacio en el que está definido dicho endomorfismo.

Por otra parte, independientemente de que en el teorema se indica tomar un vector propio y completar la base del espacio para obtener el bloque al cual debemos volver a aplicar el método, en la práctica podemos tomar en lugar de un solo vector propio, el número máximo de vectores propios linealmente independientes que aparezcan en un subespacio propio, ya que esto reduce el tamaño del bloque que queda por triangulizar e incluso en matrices de tamaño pequeño permite triangulizar en un solo paso.

Ejemplo

Consideremos el endomorfismo f de $\mathbb{R}^3$ que en la base canónica de dicho espacio está representado por la matriz

$$M(f, (e_i)) = A = \begin{pmatrix} 2 & 1 & 0 \\ 0 & 1 & -1 \\ 0 & 2 & 4 \end{pmatrix}$$

En este caso,

$$p_f(x) = p_A(x) = (x-2)^2(3-x),$$
$$E_2 = \langle\langle(1,0,0)\rangle\rangle,$$
$$E_3 = \langle\langle(1,1,-2)\rangle\rangle.$$

De aquí podemos tomar:

$$a_1 = (1,1,-2) = e_1 + e_2 - 2e_3,$$
$$a_2 = (1,0,0) = e_1,$$
$$a_3 = (0,0,1) = e_3.$$

Busquemos $M(f, (a_i))$.
Tenemos que

$$f(a_1) = 3a_1,$$
$$f(a_2) = 2a_2,$$
$$f(a_3) = f(e_3) = -e_2 + 4e_3 = -a_1 + a_2 - 2a_3 + 4a_3 = -a_1 + a_2 + a_3$$

y

$$M(f, (a_i)) = \left(\begin{array}{cc:c} 3 & 0 & 1 \\ 0 & 2 & -1 \\ \hdashline 0 & 0 & 2 \end{array}\right)$$

que es triangular, pues el bloque que quedaría en este caso es de tamaño 1.

Para terminar lo referente al método de triangulización, queremos señalar que lo hemos enunciado y demostrado para una representación matricial mediante una matriz triangular superior, pero que hubiéramos podido plantear la posibilidad de encontrar una representación mediante una matriz triangular inferior. Una vez demostrado el teorema para una representación triangular superior, queda demostrada la posibilidad de una representación triangular inferior, ya que si en la base $(a_1, a_2, \ldots, a_n)$ la matriz de f es triangular superior, en la base $(a_n, a_{n-1}, \ldots, a_1)$ será triangular inferior.

Por último, es evidente que el teorema siempre es válido si $K = \mathbb{C}$.

Volvamos ahora al teorema de descomposición primaria de un endomorfismo f.

Teníamos que la descomposición del espacio E es

$$E = \mathrm{Ker}(f - \lambda_1 id_E)^{m_1} \oplus \ldots \oplus \mathrm{Ker}\ (f - \lambda_k id_E)^{m_k}.$$

Esto nos permite obtener, mediante la selección de una base B_i formada por la unión de bases de los $\mathrm{Ker}(f - \lambda_i\ id_E)^{m_i}$, una matriz diagonal por bloques que represente al endomorfismo, esto es,

$$M(f, B) = \begin{pmatrix} \boxed{A_1} & & & 0 \\ & \boxed{A_2} & & \\ & & \ddots & \\ 0 & & & \boxed{A_k} \end{pmatrix},$$

donde cada A_i representa la matriz de la restricción f_i de f a $\mathrm{Ker}\ (f - \lambda_i\ id_E)^{m_i}$ en la base B_i.

Por otra parte, acabamos de demostrar que si un endomorfismo de un K-espacio E es tal que las raíces de su polinomio característico están en K, puede encontrarse una base en la cual la matriz de dicho endomorfismo sea triangular superior. Para cada f_i las raíces de su polinomio característico están en K ya que $p_{f_i} = (x - \lambda_i)^{m_i}$. Podemos entonces encontrar para cada f_i una base B_i en la cual la matriz de f_i sea triangular superior, mientras que la estabilidad de $\mathrm{Ker}(f - \lambda_i id_E)^{m_i}$ por f garantiza que si tomamos la matriz de f con respecto a la base B' formada por la unión de las B_i, no se pierde la forma diagonal por bloques. Así,

$$M(f,\ B') = \begin{pmatrix} \boxed{T_1} & & & \\ & \boxed{T_2} & & \\ & & \ddots & \\ & & & \boxed{T_k} \end{pmatrix},$$

donde cada T_i es un bloque triangular superior.

Tomemos un bloque triangular T_i. Es fácil ver que cada bloque triangular puede descomponerse en la forma

$$T_i = \lambda_i I_{m_i} + A'_i,$$

donde A'_i es una matriz triangular superior cuyos elementos diagonales son 0.

En el ejemplo dado a continuación del teorema 5.4

$$T_1 = \begin{pmatrix} -1 & 3 \\ 0 & -1 \end{pmatrix} = \begin{pmatrix} -1 & 0 \\ 0 & -1 \end{pmatrix} + \begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}.$$

Para obtener T_2 el estudiante puede comprobar que aplicando el método de triangulización a

$$A_2=M(f_2),\ (b_3,b_4,b_5))=\begin{pmatrix}1 & 1 & 1\\ 2 & 0 & -3\\ 0 & 1 & 2\end{pmatrix}$$

como $p_{f_2}(x)=p_{A_2}(x)=(1-x)^3$
y $E(1)=\langle b_1-b_2+b_3\rangle$,
si tomamos

$$c_1=b_1-b_2+b_3,$$
$$c_2=b_2,$$
$$c_3=b_3,$$

entonces

$$f(c_1)=c_1,$$
$$f(c_2)=f(b_2)=b_1+b_3=c_1+c_2-c_3+c_3=c_1+c_2,$$
$$f(c_3)=f(b_3)=b_1-3b_2+2b_3+2b_3=c_1+c_2-c_3-3c_2+2c_3=c_1-2c_2+c_3$$

y

$$T_2=M(f_2,\ (c_1,c_2,c_3))=\begin{pmatrix}1 & 1 & 1\\ 0 & 1 & -2\\ 0 & 0 & 1\end{pmatrix}$$

Así,

$$T_2=\begin{pmatrix}1 & 1 & 1\\ 0 & 1 & -2\\ 0 & 0 & 1\end{pmatrix}=\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}+\begin{pmatrix}0 & 1 & 1\\ 0 & 0 & -2\\ 0 & 0 & 0\end{pmatrix}$$

$\lambda_i I_{m_i}$ y A_i representan exactamente a $\lambda_i\ id_{N_i}$ y $f_i-\lambda_i\ id_{N_i}$ en la base B_i' y $f_i-\lambda_i id_{N_i}$ es tal, que su único valor propio es el 0.

Por otra parte, si queremos seguir simplificando la matriz de f_i y por tanto la de f, observemos que la acción de un cambio de base sobre

$$T_i=\lambda_i I_{m_i}+A_i'$$

tiene como resultado:

$$\begin{aligned}P^{-1}T_iP&=P^{-1}\lambda_i I_{m_i}P+P^{-1}\ A_i'P\\ &=\lambda_i I_{m_i}+P^{-1}A_i'P.\end{aligned}$$

En otras palabras, es necesario y suficiente realizar los cambios de base de forma tal que se simplifique A_i'. Es por ello que vamos a estudiar endomorfismos tales que su único valor propio sea el 0. A estos endomorfismos se les llama *nilpotentes* y veremos inmediatamente por qué.

5.4 Endomorfismos nilpotentes

En el epígrafe anterior llegamos a la conclusión de la necesidad de estudiar endomorfismos cuyo único valor propio fuese el 0. Vamos a caracterizar dichos endomorfismos.

En primer lugar, está claro que el polinomio característico de un tal endomorfismo f es

$$p_f(x) = (-1)^n x^n$$

y su polinomio minimal debe ser de la forma

$$w_f(x) = x^p, \quad 1 \leqslant p \leqslant n.$$

Por el teorema de Hamilton-Cayley y la definición de polinomio minimal, podemos concluir que

$$f^n = 0, \; f^p = 0 \text{ y } f^{p-1} \neq 0.$$

Recíprocamente, supongamos que f sea un endomorfismo de E tal que exista $p \geqslant 1$, p entero, tal que

$$f^p = 0 \text{ y } f^{p-1} \neq 0.$$

Probemos que el único valor propio de f es el 0.

Supongamos que λ es un valor propio de f, entonces existe $v \neq 0$ de E tal que

$$f(v) = \lambda v$$

Luego,

$$f^2(v) = f(f(v)) = f(\lambda v) = \lambda f(v) = \lambda^2 v,$$

y se puede comprobar por inducción que

$$f^i(x) = \lambda^i v.$$

Por tanto,

$$f^p(v) = \lambda^p v = 0.$$

Como $v \neq 0$ entonces $\lambda^p = 0$, luego si λ es valor propio de f, $\lambda = 0$.

Por otra parte,

$$f^{p-1} \neq 0 \Longleftrightarrow \text{existe } v \in E \text{ tal que } f^{p-1}(v) \neq 0 \text{ y}$$

$$f(f^{p-1}(v)) = f^p(v) = 0.$$

Si llamamos $y = f^{p-1}(v)$, tendremos que

$$f(y) = 0 \cdot y = 0 \text{ con } y \neq 0,$$

o sea, 0 es un valor propio de f; más aún, es el único valor propio de f.

De todo lo anterior resulta natural que podamos dar la definición siguiente:

DEFINICIÓN 5.3

Se dice que un endomorfismo f de un espacio vectorial de dimensión n sobre K es *nilpotente* de índice p si y solo si existe un entero $p \geqslant 1$ tal que $f^p=0$ y $f^{p-1} \neq 0$

Análogamente sabemos que una matriz A de $M_n(K)$ es nilpotente de índice p si y solo si existe un entero $p \geqslant 1$ tal que

$A^p=0$ y $A^{p-1} \neq 0$.

Está claro que los endomorfismos nilpotentes estarán representados en cualquier base por matrices nilpotentes y, recíprocamente, toda matriz nilpotente representará endomorfismos nilpotentes. Esto se debe a que si la matriz A es nilpotente, toda matriz semejante a A es nilpotente, lo que puede ser verificado fácilmente por el estudiante.

Resumamos las características de los endomorfismos nilpotentes:

Sea f un endomorfismo nilpotente *de índice p.* Entonces,

(1) El único valor propio de f es el 0, y esto lo caracteriza como endomorfismo nilpotente; esto es, si un endomorfismo es tal que su único valor propio es el 0, es nilpotente.
(2) El polinomio característico de f es $p_f=(-1)^n x^n$ y su polinomio minimal $w_f=x^p$.
(3) La matriz representativa de f en cualquier base del espacio es nilpotente de índice p.

Nota: La definición de endomorfismo nilpotente, así como la caracterización de un endomorfismo nilpotente a partir de que su único valor propio es el 0, son también válidos para espacios vectoriales de dimensión infinita. Sin embargo, para dichos espacios no son válidas las consideraciones sobre el polinomio característico y la representación matricial. Dejamos al estudiante que analice lo referente al polinomio minimal de un endomorfismo nilpotente en el caso de un espacio vectorial de dimensión infinita.

Nuestro objetivo inmediato es la búsqueda de una base en la cual un endomorfismo nilpotente se pueda representar por una matriz lo más simple posible.

El hecho de que el único valor propio de un endomorfismo nilpotente de índice p sea el 0, permite afirmar que si $v \neq 0$ y $f(v) \neq 0$, entonces

$f(v) \neq \lambda v \ \ \forall \lambda \neq 0$.

En otras palabras v y $f(v)$ son linealmente independientes.

Podríamos suponer, haciendo un razonamiento inductivo, que $v, f(v)$, $f^2(v), \ldots, f^{p-1}(v)$ son linealmente independientes. Por supuesto, que no podríamos llegar hasta $f^p(v)$ pues $f^p(v)=0$. Comprobemos si no nos engaña la intuición.

PROPOSICIÓN 5.7

Sean f un endomorfismo nilpotente de índice p de un K-espacio E de dimensión n y $v \in E$ tal que $f^{p-1}(v) \neq 0$. Entonces el sistema de vectores

$$(f^{p-1}(v), f^{p-2}(v), \ldots, f(v), v)$$

es linealmente independiente.
(Obsérvese que la suposición $f^{p-1}(v) \neq 0$ obliga a $v, f(v), \ldots, f^{p-1}(v)$ a ser no nulos. Tomamos la ordenación $f^{p-1}(v), \ldots, v$ por comodidad y por necesidades futuras.

Demostración

Supongamos que

$$a_{p-1} f^{p-1}(v) + \ldots + a_1 f(v) + a_0 v = 0.$$

Aplicando f^{p-1} a ambos miembros de la igualdad, resulta:

$$a_{p-1} f^{2(p-1)}(v) + \ldots + a_1 f^p(v) + a_0 f^{p-1}(v) = 0,$$

o sea,

$$a_0 f^{p-1}(v) = 0.$$

Como $f^{p-1}(v) \neq 0$ tenemos que $a_0 = 0$.
Si aplicamos f^{p-2} de manera similar, obtenemos $a_1 = 0$ y aplicando $f^{p-(i+1)}$ tenemos que $a_i = 0$ para $i = 1, 2, \ldots, p-1$.
Con esto queda probado que $(f^{p-1}(v), f^{p-2}(v), \ldots f(v), v)$ es linealmente independiente.

La proposición anterior nos da una idea de qué tipo de vectores podemos tomar para formar una base de E. Vamos a suponer que tenemos suficientes vectores del tipo anterior para formar una base de E. Veamos si dichos vectores nos permiten hallar una forma simple para la matriz de f. Necesitaremos n vectores; luego el índice de nilpotencia de f debe ser n.
Supongamos entonces que f es un endomorfismo del K-espacio E de dimensión n y de índice de nilpotencia n.
Según la proposición 5.7, los n vectores:

$$f^{n-1}(v), f^{n-2}(v), \ldots f(v), v$$

son linealmente independientes y, por tanto, forman una base de E.
Busquemos la matriz de f en la base

$$(b_1 = f^{n-1}(v), b_2 = f^{n-2}(v), \ldots, b_{n-1} = f(v), b_n = v).$$

Tenemos:

$$f(b_1) = f^n(v) = 0$$
$$f(b_2) = f^{n-1}(v) = b_1$$
$$\vdots$$
$$f(b_i) = f^{n-(i-1)}(v) = b_{i-1}$$
$$\vdots$$
$$f(b_{n-1}) = f^2(v) = b_{n-2}$$
$$f(b_n) = f(v) = b_n.$$

De aquí,

$$M(f,(b_1,b_2,\ldots,b_n)) = \begin{pmatrix} 0 & 1 \ldots & 0 \\ 0 & 0 \ldots & 0 \\ \ldots & \ldots & \ldots \\ 0 & 0 \ldots & 1 \\ 0 & 0 \ldots & 0 \end{pmatrix}$$

Esta es precisamente la matriz que definimos como celda de Jordán para el caso de un endomorfismo nilpotente cuyos polinomios característico y minimal son $p_f(x) = (-1)^n x^n, w_f(x) = x^n$. Por tanto, la búsqueda de la base en la cual un endomorfismo nilpotente esté representado por una matriz simple deberá estar relacionada con los vectores de la forma $f^k(v)$.

Ahora bien, no siempre el índice de nilpotencia coincide con la dimensión del espacio; esto significa que no tendríamos un número suficiente de vectores del tipo $f^k(v)$ si el índice de nilpotencia p es estrictamente menor que la dimensión n del espacio, por lo cual hay que generalizar el resultado encontrado para $p=n$ al caso $p<n$.

Esta generalización es compleja y se obtiene mediante un método constructivo de la base. En esencia vuelve a emplearse el principio de descomponer el espacio en suma directa de subespacios convenientes y aprovechar el carácter de nilpotente del endomorfismo f.

Obsérvese que en el caso $p=n$, $E=\text{Ker } f^p$, y al buscar una base de E estamos buscando una base de Ker f^p. luego en el caso $p<n$, Ker f^p es un subespacio estricto de E, de ahí que tengamos que completar la base de Ker f^p para obtener una base de E.

TEOREMA 5.7

Sea f un endomorfismo nilpotente de índice $p<n$ de un K-espacio E de dimensión n. Entonces existe una base B de E en la cual la matriz de f toma la forma

$$\begin{pmatrix} 0 & a_{12} & 0 & \dots & 0 \\ 0 & 0 & a_{23} & \dots & 0 \\ 0 & 0 & 0 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & a_{n-1,n} \\ 0 & 0 & 0 & \dots & 0 \end{pmatrix}$$

donde $a_{i,i+1}$ es 0 o es 1, $i=1,\dots,n-1$.

Demostración

Consideremos los subespacios

$V_0=\{0\}$, $V_1=\mathrm{Ker}\, f$, $V_2=\mathrm{Ker}\, f^2,\dots, V_{p-1}=\mathrm{Ker}\, f^{p-1}$,
$V_p=\mathrm{Ker}\, f^p=E$.

El método constructivo de la base parte de la descomposición de E en subespacios tomados como suplementarios de cada V_i en V_{i+1}, por ello debemos probar que:

1) La sucesión de subespacios $V_0, V_1,\dots, V_{p-1}, V_p$ es una sucesión ordenada de forma estricta por la inclusión de conjuntos. En otras palabras, $V_0 \subset V_1 \subset \dots \subset V_{p-1} \subset V_p$.
Cada una de estas inclusiones es estricta.
En efecto, sea $v \in V_i$ y probemos que $v \in V_{i+1}$.
$v \in V_{i+1} \Longleftrightarrow f^{i+1}(v)=0$
$f^{i+1}(v)=f(f^i(v))=f(0)=0$,
luego
$v \in V_i \Rightarrow v \in V_{i+1}$ y $V_i \subset V_{i+1}$.
Supongamos ahora que existe i tal que $V_i=V_{i+1}$, $0 \leqslant i \leqslant p-1$, es decir, $f^i(v)=0 \Longleftrightarrow f^{i+1}(v)=0$. Tomemos $f^p(v)$,
$f^p(v)=0$, $\forall v \in E$. Por tanto, $f^p(v)=f^{[p+i+1-(i+1)]}(v)=f^{i+1}f^{p-(i+1)}(v))=0$ $\forall v \in E$.
Pero
$f^{i+1}(f^{p-(i+1)}(v))=0 \Longleftrightarrow f^i(f^{p-(i+1)})(v))=f^{p-1}(v)=0 \quad \forall v \in E$
y esto contradice que p sea el índice de nilpotencia de f.
Concluimos que las sucesiones
$V_0 \subset V_1 \subset \dots \subset V_{p-1} \subset V_p$
son estrictas.
Sea el subespacio $V_{p-1}=\mathrm{Ker}\, f^{p-1}$ y tomemos un suplementario que llamaremos F_p de V_{p-1} en $V_p=\mathrm{Ker}\, f^p=E$; esto es,
$E=V_{p-1} \oplus F_p$.
Como V_{p-2} está contenido en V_{p-1} estrictamente, tomaremos un suplementario F_{p-1} de V_{p-2} en V_{p-1}, pero no de forma arbitraria, sino de modo que dicho suplementario F_{p-1} contenga las imágenes por f de los vectores de F_p.
Para justificar que esta posibilidad existe demostraremos lo siguiente:

2) $f(V_{i+1}) \subset V_i$, $0 \leqslant i \leqslant p-1$

Sea $y \in f(v_{i+1})$, es decir, $y=f(v)$ para algún $v \in V_{i+1}$. Debemos probar que $y \in V_i$, o sea, $f^i(y)=0$.

$f^i(y)=f^i(f(v))=f^{i+1}(v)=0$ pues $v \in V_{i+1}$.

Queda así demostrado que $f(V_{i+1}) \subset V_i$.

3) Si F es un subespacio vectorial de E tal que $F \cap V_i = \{0\}$, $1 \leqslant i \leqslant p-1$, entonces $f(F) \cap V_{i-1} = \{0\}$. Además, la restricción de f a F es inyectiva.

Sea $v \in f(F) \cap V_{i-1}$. Debemos probar que $v=0$.

Como $v \in f(F) \cap V_{i-1}$, entonces $v=f(y)$ con $y \in F$, y $f^{i-1}(v)=0$. De aquí que $f^{i-1}(f(y))=f^i(y)=0$ y $y \in V_i$ de donde resulta que $y \in F \cap V_i \Rightarrow y=0$, y como $v=f(y)$, entonces $v=0$, según queríamos demostrar.

Probemos ahora que la restricción de f a F es inyectiva.

Supongamos que $f(v)=0$ para $v \in F$. Debemos probar que $v=0$. Busquemos $f^i(v)$.

$$f^i(v)=f^{i-1}(f(v))=f(0)=0, \text{ luego } v \in V_i.$$

De $v \in V_i \cap F$ resulta que $v=0$ y, por tanto, la restricción de f a F es inyectiva.

Usaremos la inyectividad en la selección de la base adecuada de E.)

Volvamos ahora al proceso constructivo.

Teníamos $E=V_{p-1} \oplus F_p$ y debemos encontrar un suplementario F_{p-1} de V_{p-2} en V_{p-1} tal que $f(F_p) \subset F_{p-1}$, lo que nos permite escribir $V_{p-1}=V_{p-2} \oplus F_{p-1}$.

¿Será posible encontrar F_{p-1} con tales condiciones? Para ello es necesario que $V_{p-2} \cap f(F_p)=\{0\}$ y $f(F_p) \subset V_{p-1}$. $V_{p-2} \cap f(F_p)=\{0\}$ resulta de que $V_{p-1} \cap F_p=\{0\}$ y por 3, $f(F_p) \cap V_{p-2}=\{0\}$.

$f(F_p) \subset V_{p-1}$, resulta de que $F_p \subset V_p$; de aquí, $f(F_p) \subset f(V_p)$ y por 2), $f(V_p) \subset V_{p-1}$, luego $f(F_p) \subset V_{p-1}$.

Así, $V_{p-1}=V_{p-2} \oplus F_{p-1}$ donde $f(F_p) \subset F_{p-1}$, y $E=F_p \oplus F_{p-1} \oplus V_{p-2}$.

Si repetimos el proceso con los V_i, obtenemos una sucesión de subespacios F_i tales que

$$V_i=V_{i+1} \oplus F_i,$$

con $f(F_{i+1}) \subset F_i$ e $1 \leqslant i \leqslant p$.

En particular, $V_1=V_0 \oplus F_1$ y como $V_0=\{0\}$, entonces $V_1=F_1$.

Finalmente, obtenemos:

$$E=F_p \oplus F_{p-1} \oplus \ldots \oplus F_2 \oplus F_1$$

con $F_1=V_1=\operatorname{Ker} f$, y $f(F_i) \subset F_{i-1}$, y como la sucesión de los V_i es creciente, entonces

$$\dim V_i < \dim V_{i+1}.$$

El hecho de que $f(F_i) \subset F_{i-1}$ y de que la restricción de f a F_i sea inyectiva permite afirmar que $\dim F_i \leqslant \dim F_{i-1}$, es decir, que la sucesión de los F_i es decreciente.

Tomemos ahora la base de E a partir de la descomposición

$$E=F_p \oplus F_{p-1} \oplus \dots \oplus F_1,$$

lo cual no haremos por una simple unión de bases de los F_i, sino mediante un algoritmo de formación dado.

Sea $(a_{p1}, a_{p2}, \dots, a_{pn_p})$ una base de F_p.

Como base de F_{p-1} tomaremos:

$$(f(a_{p1}), f(a_{p2}), \dots, f(a_{pn_p}), a_{p-1\, n_p}, \dots, a_{p-1\, n_{p-1}}).$$

Podemos hacer esto ya que $f(F_p) \subset F_{p-1}$, y al ser la restricción de f a F_p inyectiva, los vectores $f(a_{p1}), \dots f(a_{p\,n_p})$ son linealmente independientes.

En general, si tenemos una base de F_i, formamos la de F_{i+1} tomando las imágenes por f de la base de F_i y completando hasta tener una base de F_{i+1}. Obsérvese que es posible que no sea necesario completar ya que $\dim F_i \leqslant \dim F_{i-1}$ y puede cumplirse la igualdad.

Los siguientes vectores forman una base de E:

De F_p: $a_{p\,1}, a_{p\,2}, \dots, a_{p\,n_p}$.
De F_{p-1}: $f(a_{p\,1}), f(a_{p\,2}), \dots f(a_{p\,n_p}), a_{p-1\,n_p+1}, \dots, a_{p-1\,n_{p-1}}$.
De F_{p-2}: $f^2(a_{p\,1}), f^2(a_{p\,2}), \dots f^2(a_{p\,n_p}), f(a_{p-1\,n_p+1}), \dots f(a_{p-1\,n_{p-1}})$;
$a_{p-2\,n_{p-1}}+1, \dots, a_{p-2\,n_{p-2}}$.

....

De V_1: $f^{p-1}(a_{p\,1}) f^{p-1}(a_{p2}), \dots f^{p-1}(a_{p\,n_p}) f^{p-2}(a_{p-1\,n_p+1}), \dots f^{p-2}(a_{p-1\,n_{p-1}})$,
$f^{p-3}(a_{p-2\,n_{p-1}}+1), \dots f^{p-3}(a_{p-2\,n_{p-2}}), \dots, a_{1\,n_2}+1, \dots, a_{1\,n_1}$.

Para formar la base tomaremos los vectores en el orden siguiente:

$f^{p-1}(a_{p\,1}), f^{p-2}(a_{p\,1}), \dots, a_{p\,1}$,
$f^{p-1}(a_{p2}), f^{p-2}(a_{p\,2}), \dots, a_{p2}$,

....

$f^{p-1}(a_{p\,n_p}), f^{p-2}(a_{p\,n_p}), \dots, a_{p\,n_p}$,
$f^{p-2}(a_{p-1\,n_p+1}) f^{p-3}(a_{p-1\,n_p+1}), \dots, a_{p-1\,n_p+1}$,

....

$f^{p-2}(a_{p-1\,n_{p-1}}), f^{p-3}(a_{p-1\,n_{p-1}}), \dots, a_{p-1\,n_{p-1}}$,
$f^{p-3}(a_{p-2\,n_{p-1}+1}), f^{p-4}(a_{p-2\,n_{p-1}+1}), \dots, a_{p-2\,n_{p-1}+1}$,

....

$a_{1\,n_2+1}, a_{1\,n_2+2}, \dots, a_{1\,n_1}$.

Esto es, tomamos los elementos del cuadro anterior por columnas comenzando por la primera columna de la izquierda, de abajo hacia arriba, y no comenzamos a tomar elementos de una nueva columna hasta no agotar la anterior. De esta forma, si llamamos $(b_1, b_2, \dots, b_n)$ a esta base, obtenemos:

$$f(b_i)=0 \text{ o } f(b_i)=b_{i-1}.$$

Obtendremos $f(b_i)=0$ cada vez que hallemos la imagen de un b_i tomado de la última fila del cuadro, ya que en esta fila estaban situados los elementos de $V_1=F_1=\operatorname{Ker} f$, y obtendremos $f(b_i)=b_{i-1}$ en cualquier otro caso, pues según la ordenación que le dimos a la base, si por ejemplo, $b_i = f^{p-k}(a_{ij})$, entonces $b_{i-1} = f^{p-k+1}(a_{ij})$ y evidentemente

$$f(b_i) = f(f^{p-k}(a_{ij})) = f^{p-k+1}(a_{ij}) = b_{i-1}.$$

Luego la matriz de f en la base $(b_1, b_2, \ldots, b_n)$ tomará la forma deseada.

$$\begin{pmatrix} 0 & a_{12} & 0 & \ldots & 0 \\ 0 & 0 & a_{23} & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots & a_{n-1\,n} \\ 0 & 0 & 0 & \ldots & 0 \end{pmatrix}$$

El teorema queda así demostrado.

Obsérvese que si $p = n$, la cadena

$$\{0\} \subset \operatorname{Ker} f \subset \operatorname{Ker} f^2 \subset \ldots \operatorname{Ker} f^{n-1} \subset \operatorname{Ker} f^n = E$$

está formada por subespacios cuya dimensión aumenta una unidad en cada paso. Por tanto, los subespacios F_i del teorema son todos de dimensión 1 y cada uno aporta un solo vector a la base lo cual simplifica el proceso. Esta es, entre otras, la razón para considerar aparte el caso $p = n$.

El teorema quedó demostrado, pero como a su vez proporciona el método de búsqueda de la base en la práctica, queremos hacer algunos comentarios sobre los resultados obtenidos.

La matriz que resulta del teorema es una matriz diagonal por bloques y el número y el tamaño de los bloques queda determinado por el endomorfismo f. El número de bloques está dado por el número de columnas nulas y este número está dado a su vez por el número de elementos de la base de $F_1 = V_1 = \operatorname{Ker} f$, esto es, por la dimensión de $\operatorname{Ker} f$.

Por otra parte, aparece al menos un bloque de orden p correspondiente a los vectores

$$f^{p-1}(a_{p1}), \ldots, a_{p1}$$

y habrá tantos bloques de orden p como el número de elementos de la base de F_p, esto es, la dimensión de F_p suplementario de $V_{p-1} = \operatorname{Ker} f^{p-1}$.

El tamaño de los bloques restantes será menor o igual que p.

Ejemplo

Sea f el endomorfismo de $\mathbb{R}^5$ representado en la base canónica de dicho espacio por

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

f es nilpotente ya que por simple inspección de su matriz representativa vemos que su único valor propio es el 0 y que su polinomio característico es $p_f = p_A = -x^5$.

Ahora bien, hay que determinar el índice de nilpotencia o, lo que es lo mismo, su polinomio minimal:

$$A^2=\begin{pmatrix} 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \quad \text{y } A^3=0.$$

Luego f y A son nilpotentes del índice 3 y

$$w_f=w_A=x^3.$$

Siguiendo la demostración del teorema, la sucesión de V_i que debemos formar es

$$V_0=\{0\}, \ V_1= \operatorname{Ker} f, \ V_2=\operatorname{Ker} f^2,$$

ya que $V_3=\operatorname{Ker} f^3= \mathbb{R}^5$.

Calculemos $V_1=\operatorname{Ker} f$ y $V_2=\operatorname{Ker} f^2$.

$$\begin{aligned} V_1=\operatorname{Ker} f&=\{(x_1,x_2,x_3,x_4,x_5)/x_2+x_3+x_5=0, \ x_3+x_4+x_5=0\} \\ &=\{(x_1,-x_3-x_5,x_3,-x_3-x_5,x_5) \ /x_1,x_2,x_3,x_4,x_5\in \mathbb{R}\}. \\ V_2=\operatorname{Ker} f&=\{(x_1,x_2,x_3,x_4,x_5)/x_3+x_4+x_5=0\} \\ &=\{(x_1,x_2,x_3,x_4,-x_3-x_4)/x_1,x_2,x_3,x_4,x_5\in \mathbb{R}\}. \end{aligned}$$

Debemos buscar un suplementario F_3 de V_2. De la forma de V_2 resulta de inmediato que podemos tomar

$$F_3=\langle\langle(0,0,0,0,1)\rangle\rangle$$

y $\mathbb{R}^5=V_2\oplus F_3$.

Ahora debemos buscar un suplementario F_2 de V_1 en V_2 que contenga a $f(F_3)$, o sea, a $f(0,0,0,0,1)=(1,1,0,0,0)$

Como dim $V_1=3$ y dim $V_2=4$, un suplementario F_2 de V_1 en V_2 tendrá dimensión 1; luego basta tomar $F_2=\langle\langle(1,1,0,0,0)\rangle\rangle$ y tendremos $V_2=V_1\oplus F_2$; luego

$$\mathbb{R}^5=V_1\oplus F_2\oplus F_3=F_1\oplus F_2\oplus F_3.$$

Falta encontrar una base de $V_1=F_1=\operatorname{Ker} f$. Tomaremos

$$V_1=\langle\langle(1,0,0,0,0), \ (0,-1,1,-1,0), \ (0,-1,0,-1,1)\rangle\rangle,$$

donde $f(F_2)=f(1,1,0,0,0)=(1,0,0,0,0)$.

Tomemos:

$$\begin{aligned} &a_{31}=(0,0,0,0,1), \\ &f(a_{31})=(1,1,0,0,0), \\ &f^2(a_{31})=(1,0,0,0,0), \ a_{12}=(0,-1,1,-1,0), \ a_{13}=(0,-1,0,-1,1). \end{aligned}$$

Entonces,

$b_1=(1,0,0,0,0)=f^2(a_{31})$
$b_2=(1,1,0,0,0)=f(a_{31})$
$b_3=a_{31}$
$b_4=a_{12}=(0,-1,1,-1,0)$
$b_5=a_{13}=(0,-1,0,-1,1)$

y

$$M(f,(b_i))=\left(\begin{array}{ccc|c|c} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \\ \hline 0 & 0 & 0 & 0 & 0 \end{array}\right)$$

tiene un bloque de orden 3 que es el índice de nilpotencia de f; en total 3 bloques y dim $V_1=\operatorname{Ker} f=3$.

5.5 Teorema de Jordán

El resultado de los epígrafes anteriores permite demostrar de forma rápida y elegante el teorema de Jordán, ya estudiado en el capítulo anterior (teorema 4.8).

Teorema de Jordán

Sea f un endomorfismo de un K-espacio E de dimensión n tal que su polinomio característico tiene todas sus raíces en K. Entonces existe una base B de E con respecto a la cual la matriz de f es una matriz de Jordán, es decir,

$$M(f,B)=\begin{pmatrix} J_{11} & & 0 \\ & \ddots \; J_{ij} \; \ddots & \\ 0 & & J_{kh} \end{pmatrix},$$

donde

$$J_{ij}=\begin{pmatrix} \lambda_i & 1 & & 0 \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ 0 & & & \lambda_i \end{pmatrix}$$

o sea, es una celda de Jordán.

Esta representación es única salvo el orden de las celdas de Jordán.

Demostración

Supongamos que

$$p_f(x)=(-1)^n(x-\lambda_1)^{m_1}\ldots(x-\lambda_k)^{m_k} \text{ y}$$

$$w_f(x)=(x-\lambda_1)^{r_1}\ldots(x-\lambda_k)^{r_k} \text{ con } 1\leqslant r_i\leqslant m_i,$$

donde $\lambda_1, \lambda_2, \ldots, \lambda_k$ son los valores propios distintos de f.

El teorema de descomposición primaria de un endomorfismo asegura que tomando una base de E formada mediante la unión de bases B_i de los $N_i=\mathrm{Ker}(f-\lambda_i id_E)^{m_i}$, obtenemos una matriz que represente a f de la forma diagonal por bloques:

$$\begin{pmatrix} A_1 & & & 0 \\ & A_2 & & \\ & & \ddots & \\ 0 & & & A_k \end{pmatrix},$$

donde cada A_i representa la matriz del endomorfismo f_i, restricción de f el subespacio estable N_i.

En cada subespacio N_i podemos descomponer f_i como

$$f_i=\lambda_i\ id_{N_i}+(f_i-\lambda_i\ id_{N_i}),$$

donde el endomorfismo $(f_i-\lambda_i\ id_{N_i})$ de N_i es nilpotente de índice r_i porque si $x\in \mathrm{Ker}\ (f-\lambda_i\ id_E)^{r_i}$, entonces $(f-\lambda_i\ id_E)^{r_i}(x)=0$, lo que implica que $(f_i-\lambda_i\ id_{N_i})^{r_i}(x)=0$ para todo $x\in N_i$.

Dejamos al estudiante que analice por qué el índice de nilpotencia es precisamente r_i.

Podemos encontrar para cada N_i una base B_i en la cual la matriz de $f_i-\lambda_i id_{N_i}$ tome la forma de celdas elementales de Jordán, para el endomorfismo nilpotente hallado en el teorema 5.7.

Luego en cada N_i la matriz de $f_i-\lambda_i\ id_{N_i}$ vuelve a descomponerse en bloques menores; entre ellos habrá un bloque de tamaño r_i y habrá tantos bloques como la dimensión de $\mathrm{Ker}(f_i-\lambda_i id_{N_i})$, o sea, la dimensión del subespacio propio asociado a λ_i.

Como los subespacios N_i son estables por f, los cambios de base que se realicen en cada N_i no alteran la estructura diagonal por bloques dada por el teorema de descomposición primaria.

Por otra parte, como la matriz de $\lambda_i\ id_{N_i}$ es diagonal en cualquier base, el cambio de base de B_i a B'_i en cada N_i no afecta la representación matricial de $\lambda_i\ id_{N_i}$, y una vez hallada la forma de las celdas de Jordán para el endomorfismo $(f_i - \lambda_i\ id_{N_i})$, podemos regresar a f_i y tendremos en B_i' una representación para f_i de la forma:

$$\begin{pmatrix} \lambda_i & a_{12} & & \\ & \ddots & \ddots & \\ & & \ddots & a_{m_i-1\ m_i} \\ & & & \lambda_i \end{pmatrix} \qquad \text{donde } a_{ij}=1 \text{ o } 0.$$

Esto es, cada f_i está representado por una matriz compuesta por celdas de Jordán correspondientes al valor λ_i.

Si tomamos en E una base formada por la unión de las bases B_i', obtendremos la representación de Jordán de f buscada.

De la demostración del teorema, resulta evidente la unicidad de la forma de Jordán de un endomorfismo salvo el orden en que aparezcan las celdas.

En la diagonal de la matriz de Jordán de f aparecen los valores propios de esta; el número y el tamaño de las celdas de Jordán para cada valor propio están dados por el teorema 5.7 de descomposición de los endomorfismos nilpotentes y dependen en cada caso exclusivamente del endomorfismo nilpotente

$$f_i - \lambda_i\ id_{N_i}.$$

Esto completa la demostración del teorema de Jordán.

Por otra parte, si un endomorfismo es tal que puede ser representado en una cierta base del K-espacio por una matriz de Jordán, es evidente que su polinomio característico tiene por raíces los elementos de la diagonal de dicha matriz de Jordán y, por tanto, su polinomio característico tiene todas sus raíces en K. Entonces podemos afirmar:

Una condición necesaria y suficiente para que un endomorfismo de un K-espacio E de dimensión finita pueda ser representado por una matriz de Jordán, es que su polinomio característico tenga todas sus raíces en K.

Si $K=\mathbb{C}$ es obvio que siempre es posible representar cualquier endomorfismo de un $\mathbb{C}$-espacio por una matriz de Jordán.

A continuación haremos algunos señalamientos sobre la búsqueda en la práctica de la matriz de Jordán de un endomorfismo.

En general, para determinar la matriz de Jordán de un endomorfismo f basta conocer el polinomio característico p_f de f, el polinomio minimal w_f de f y la dimensión de los subespacios propios $E(\lambda)$ asociados a los diferentes valores propios de f, ya que si

$$p_f=(-1)^n\ (x-\lambda_1)^{m_1} \dots (x-\lambda_k)^{m_k}, \text{ y}$$
$$w_f=(x-\lambda_1)^{r_1} \dots (x-\lambda_k)^{r_k},$$

entonces para cada valor propio λ_i, aparece una celda de Jordán de tamaño r_i; las demás celdas son de tamaño menor o igual que r_i y el número total de celdas está dado por la dimensión de $E(\lambda_i)$. Otro elemento a tomar en cuenta es que λ_i aparece en la diagonal m_i veces.

Ejemplo

Sea f un endomorfismo de $\mathbb{R}^7$ cuyos polinomios característico y minimal son:

$$p_f = -(x-2)^4(x-3)^3$$
$$w_f = (x-2)^2(x-3)^2.$$

Para $\lambda=2$, aparecen una celda de Jordán de orden 2 y otra de orden 2 o dos de orden 1.
Para $\lambda=3$ aparece una celda de Jordán de orden 2, y en este caso la única posibilidad es otra de orden 1.
Así obtenemos para f dos posibles matrices de Jordán (salvo el orden de las celdas):

$$J_1 = \begin{pmatrix} 2 & 1 & & & & & \\ 0 & 2 & & & & & \\ & & 2 & 1 & & & \\ & & 0 & 2 & & & \\ & & & & 3 & 1 & \\ & & & & 0 & 3 & \\ & & & & & & 3 \end{pmatrix}$$

$$J_2 = \begin{pmatrix} 2 & 1 & & & & & \\ 0 & 2 & & & & & \\ & & 2 & & & & \\ & & & 2 & & & \\ & & & & 3 & 1 & \\ & & & & 0 & 3 & \\ & & & & & & 3 \end{pmatrix}$$

J_1 resultará si dim $E(2)=2$ y J_2 si dim $E(2)=3$.

Ahora bien, los elementos anteriormente dados no determinan completamente la forma de la matriz de Jordán de un endomorfismo en todos los casos, como lo muestra el ejemplo siguiente:

Ejemplo

Sea el endomorfismo f de $\mathbb{R}^9$ cuyos polinomios característico y minimal son:

$$p_f(x) = -(x-2)^7(x-3)^2$$
$$w_f(x) = (x-2)^3(x-3)^2$$

y dim $E(2)=3$.

Aún cuando conocemos la dimensión de $E(2)$, las posibilidades son dos:

$$\begin{pmatrix} 2 & 1 & 0 & & & & & & \\ 0 & 2 & 1 & & & & & & \\ 0 & 0 & 2 & & & & & & \\ & & & 2 & 1 & 0 & & & \\ & & & 0 & 2 & 1 & & & \\ & & & 0 & 0 & 2 & & & \\ & & & & & & 2 & & \\ & & & & & & & 3 & 1 \\ & & & & & & & 0 & 3 \end{pmatrix}, \quad \begin{pmatrix} 2 & 1 & 0 & & & & & & \\ 0 & 2 & 1 & & & & & & \\ 0 & 0 & 2 & & & & & & \\ & & & 2 & 1 & & & & \\ & & & 0 & 2 & & & & \\ & & & & & 2 & 1 & & \\ & & & & & 0 & 2 & & \\ & & & & & & & 3 & 1 \\ & & & & & & & 0 & 3 \end{pmatrix}$$

En estos casos hay que acudir a datos más precisos aportados por el teorema de los endomorfismos nilpotentes, ya que de dicho teorema resulta, por ejemplo, que la cantidad de bloques de tamaño p para un endomorfismo nilpotente de índice p está dada por la dimensión de F_p, suplementario de $V_{p-1}=\operatorname{Ker} f^{p-1}$. En este ejemplo habría que determinar la dimensión de $\operatorname{Ker}(f-2\, id_{\mathbf{R}^9})^2$ y de aquí la de un suplementario de este en $\operatorname{Ker}(f-2\, id_{\mathbf{R}^9})^7$.

El teorema de los endomorfismos nilpotentes aporta aún datos más precisos y determina completamente la forma de la matriz de Jordán de un endomorfismo en todos los casos.

La situación de este último ejemplo no es la más usual, ya que en la mayoría de los casos se puede determinar la forma de la matriz de Jordán de f mediante los datos: polinomio característico de f, polinomio minimal de f y dimensión de los subespacios propios de f.

En cuanto a la búsqueda de la base en la cual se obtiene la matriz de Jordán, debemos seguir el método constructivo dado para encontrar esta matriz. Veamos.

Dado un endomorfismo f del K-espacio de dimensión n, que tiene el polinomio característico

$$p_f(x)=(-1)^n(x-\lambda_1)^{m_1}\ldots(x-\lambda_x)^{m_k}$$

y el polinomio minimal

$$w_f(x)=(x-\lambda_1)^{r_1}\ldots(x-\lambda_k)^{r_k},$$

donde los valores $\lambda_1,\ldots,\lambda_k$ son todos distintos y pertenecen a K y $r_1 \leqslant m_i$, el método consiste en:

1ro. Aplicar a f el teorema de descomposición primaria y obtener $E=\operatorname{Ker}(f-\lambda_1\, id_E)^{m_1}\oplus\ldots\oplus\operatorname{Ker}(f-\lambda_k id_E)^{m_k}$.

Tomando una base de E formada por la unión de las bases B_i de $\operatorname{Ker}(f-\lambda_i\, id_E)^{m_i}$, obtener en dicha base una matriz que represente a f de la forma diagonal por bloques

$$\begin{pmatrix} A_1 & & & \\ & A_2 & & \\ & & \ddots & \\ & & & A_k \end{pmatrix}$$

donde A_i es la matriz de la restricción de f a $N_i = \mathrm{Ker}(f-\lambda_i\ id_E)^{m_i}$.

2do. Aplicar en cada N_i el proceso de triangulización a f_i, luego a A_i, obteniendo bases B_i' de cada N_i cuya unión es una base de B en la cual la matriz de f tiene la forma

$$\begin{pmatrix} T_1 & & & \\ & T_2 & & \\ & & \ddots & \\ & & & T_k \end{pmatrix}$$

y cada T_i es triangular superior.

3ro. Separar cada T_i como una matriz diagonal más una nilpotente N_i. Así resulta

$$\begin{pmatrix} \lambda_1 I_{m_1} & & & \\ & \lambda_2 I_{m_2} & & \\ & & \ddots & \\ & & & \lambda_k\ I_{m_k} \end{pmatrix} + \begin{pmatrix} L_1 & & & \\ & L_2 & & \\ & & \ddots & \\ & & & L_k \end{pmatrix}$$

Esta L_i representa al endomorfismo $(f_i - \lambda_i id_{N_i})$ de N_i que es un endomorfismo nilpotente y que permitirá buscar en cada N_i la base B_i'' en la cual el endomorfismo nilpotente $f_i - \lambda_i id_{N_i}$ toma la forma de Jordán J_i. Como $\lambda_i I_{m_i}$ sigue teniendo la misma forma en B_i'', la unión de las bases B_i'' dará la base de E en la cual el endomorfismo toma la forma de Jordán. Esta forma se encuentra sumando de nuevo $\lambda_i I_{m_i}$ a J_i.

Este proceso es largo y laborioso, aunque en algunos casos prácticos no es necesario realizar todos los pasos, como ocurre en el ejemplo presentado para ilustrar el teorema de descomposición primaria.

Ejemplo

En el caso mencionado teníamos el endomorfismo de $\mathbf{R}^5$ dado en la base canónica de dicho espacio por

$$A=\begin{pmatrix} 1 & 1 & -1 & 2 & -1 \\ 2 & 0 & 1 & -4 & -1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & -3 & 3 & -1 \end{pmatrix},$$

con $p_f=p_A=-(x+1)^2(x-1)^3$ y $w_f=w_A=(x+1)^2(x-1)^3$.

Aplicando el teorema de descomposición primaria obtenemos que en la base

$(b_1=(1,-1,0,0,1),\ b_2=(1,-1,\ 0,1,0),\ b_3=(1,0,0,0,0),$
$b_4=(0,1,0,0,0),\ b_5=(0,0,1,1,0)),$

se obtiene la matriz diagonal por bloques:

$$M(f,(b_1,b_2,b_3,b_4,b_5))=\left(\begin{array}{cc|ccc} -1 & 3 & & & \\ & & & 0 & \\ 0 & -1 & & & \\ \hline & & 1 & 1 & 1 \\ & & 2 & 0 & -3 \\ & 0 & 0 & 1 & 2 \end{array}\right)$$

con $A_1=M\ (f_1,(b_1,b_2))$, $A_2=M(f_2,(b_3,b_4,b_5))$ y f_1, f_2 las restricciones a Ker $(f+id_{\mathbf{R}^5})^2$ y Ker$(f-id_{\mathbf{R}^5})^3$, respectivamente.

A_1 es triangular y podemos mantener la base (b_1,b_2), pero es necesario triangulizar A_2 y obtenemos:

$$T_2=\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

En la base,

$c_3=b_1-b_2+b_3,$
$c_4=b_4,$
$c_5=b_5.$

Así, en la base (c_1,c_2,c_3,c_4,c_5),

$$M(f,(c_1,c_2,c_3,c_4,c_5))=\left(\begin{array}{cc|ccc} -1 & 3 & & 0 & \\ 0 & -1 & & & \\ \hline & & 1 & 1 & 1 \\ & 0 & 0 & 1 & -2 \\ & & 0 & 0 & 1 \end{array}\right)$$

Por tanto,

$$T_1=\begin{pmatrix} -1 & 3 \\ 0 & -1 \end{pmatrix},\quad T_2=\begin{pmatrix} 1 & 1 & 1 \\ 0 & 1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

Ahora podemos escribir:

$$\left(\begin{array}{cc|ccc} -1 & 3 & & 0 & \\ 0 & -1 & & & \\ \hline & & 1 & 1 & 1 \\ 0 & & 0 & 1 & -2 \\ & & 0 & 0 & 1 \end{array}\right)=\left(\begin{array}{c|c} -I_2 & 0 \\ \hline 0 & I_3 \end{array}\right)+\left(\begin{array}{cc|ccc} 0 & 3 & & 0 & \\ 0 & 0 & & & \\ \hline & & 0 & 1 & 1 \\ 0 & & 0 & 0 & 2 \\ & & 0 & 0 & 0 \end{array}\right),$$

donde

$$L_1=\begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix} \text{ y } L_2=\begin{pmatrix} 0 & 1 & 1 \\ 0 & 0 & 2 \\ 0 & 0 & 0 \end{pmatrix}$$

son nilpotentes de índice 2 y 3, respectivamente:

En este caso, en que los índices de nilpotencia de L_1 y L_2, representativos de $f_1+id_{N_1}$ y $f_2-id_{N_2}$, respectivamente, coinciden con las dimensiones de $N_1=\mathrm{Ker}(f+id_E)^2$ y $N_2=\mathrm{Ker}\,(f-id_E)^3$, se simplifica la aplicación del teorema de los endomorfismos nilpotentes ya que las bases de N_1 y N_2 están dadas por

$d_1=(f_1+id_{N_1})\,(c_2)=3c_1$ (no tomamos $(f_1+id_{N_1})(c_1)$ ya que es nulo),
$d_2=c_2$,
$d_3=(f_2-id_{N_2})^2(c_5)=2c_3$,
$d_4=(f_2-id_{N_2})(c_5)=c_3+2c_4$,
$d_5=c_5$.

Entonces:

$$\begin{aligned}
&(f_1+id_{N_1})(d_1)=(f_1+id_{N_1})^2(d_1)=0\\
&(f_1+id_{N_1})(d_2)=(f_1+id_{N_1})(c_2)=d_1\\
&(f_2-id_{N_2})^2(d_3)=(f_2-id_{N_2})^2(f_2-id_{N_2})(c_5)=(f_2-id_{N_2})^3(c_5)=0\\
&(f_2-id_{N_2})(d_4)=(f_2-id_{N_2})\ (f_2-id_{N_2})(c_5)=(f_2-id_{N_2})^2(c_5)=d_3\\
&(f_2-id_{N_2})(d_5)=(f_2-id_{N_2})(c_5)=d_4
\end{aligned}$$

De aquí que

$$M(f_1+id_{N_1},(d_1,d_2))=\begin{pmatrix} 0 & 1 \\ 0 & 1 \end{pmatrix},$$

$$M(f_2+id_{N_2}(d_3,d_4,d_5))=\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}\quad,$$

$$M(f, (d_1,d_2,d_3,d_4,d_5)) = \left(\begin{array}{cc|ccc} -1 & 1 & & 0 & \\ 0 & -1 & & & \\ \hline & & 1 & 1 & 0 \\ & 0 & 0 & 1 & 1 \\ & & 0 & 0 & 1 \end{array}\right)$$

que es la matriz de Jordán de f.

Obsérvese que se cumplen todas las afirmaciones hechas sobre la forma de la matriz de Jordán, ya que al ser en este caso $w_f=(x+1)^2(x-1)^3$, existe una celda de orden 2 asociada al valor propio -1 y una de orden 3 correspondiente al valor propio 1.

La expresión explícita de la base (d_1,d_2,d_3,d_4,d_5) podemos buscarla mediante las relaciones entre las diferentes bases halladas. Así,

$$\begin{aligned} &d_1=3c_1=3b_1 \\ &d_2=c_2=b_2 \\ &d_3=2c_3=2(b_1-b_2+b_3)=2b_1-2b_2+2b_3 \\ &d_4=c_3+2c_4=b_1-b_2+b_3+2b_4 \\ &d_5=c_5=b_5 \end{aligned}$$

$$\begin{aligned} &d_1=(3,-3,0,0,3) \\ &d_2=(1,-1,0,1,0) \\ &d_3=(1,0,0,-1,1) \\ &d_4=(1,2,0,-1,1) \\ &d_5=(0,0,1,1,0) \end{aligned}$$

De la demostración del teorema resulta, y se pone en evidencia en el ejemplo anterior, que la base en la que se obtiene la forma de Jordán de un endomorfismo no es única, ya que en los diferentes pasos dados, la selección de las bases se podía haber realizado en más de una forma.

El ejemplo anterior puede catalogarse entre los más sencillos, ya que no ha sido necesario aplicar el teorema de los endomorfismos nilpotentes en su forma más compleja, esto es, cuando el índice de nilpotencia es estrictamente menor que la dimensión del espacio. No obstante, el estudiante debe notar que el volumen de los cálculos realizados es considerable. Un ejemplo más complejo requeriría tomar una matriz de mayor tamaño con la consecuente dificultad de cálculo para los polinomios minimal y característico, por lo que preferimos explicar en un ejemplo aparte el uso del teorema de los endomorfismos nilpotentes en el caso $p<n$.

Por otra parte, el estudiante no debe olvidar que los ejemplos resueltos en los textos tienen un fin didáctico y como tal están preparados para que los cálculos resulten sencillos. En la práctica esto no ocurre así en general; las matrices del modelo matemático de un problema dado, pueden ser grandes o aún siendo pequeñas no tener polinomios característico o minimal con raíces enteras. Por supuesto que en estos casos se utilizan los métodos aproximados y la computación.

En ocasiones la búsqueda de una base en la cual se obtiene la matriz de Jordán se realiza por otro método muy útil, sobre todo en matrices de tamaño pequeño. Este método se basa en el conocimiento previo de la forma de la matriz de Jordán del endomorfismo y lo ejemplificaremos en el caso de un endomorfismo de $\mathbf{R}^3$.

Ejemplo

Sea f el endomorfismo de $\mathbf{R}^3$ representado en la base canónica de dicho espacio por la matriz

$$A=\begin{pmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 2 \end{pmatrix}$$

Busquemos la forma de Jordán de f y una base en la cual se obtenga dicha forma.

Tenemos,

$$p_f(x)=(1-x)(x-2)^2,$$
$$w_f(x)=(x-1)(x-2)^2.$$

Luego la forma de Jordán de f es

$$J=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 2 & 1 \\ 0 & 0 & 2 \end{pmatrix}.$$

¿Cómo buscar la base?

Supongamos que (b_1,b_2,b_3) es una base en la cual $M(f,(b_i))=J$, entonces,

$$f(b_1)=b_1,$$
$$f(b_2)=2b_2,$$
$$f(b_3)=b_2+b_3.$$

De ahí que para buscar b_1 y b_2 bastará tomar vectores propios asociados a 1 y 2, respectivamente.

$E(1)=\langle\langle(1,0,2)\rangle\rangle$, luego podemos tomar $b_1=(1,0,2)$.
$E(2)=\langle\langle(1,1,2)\rangle\rangle$; tomemos $b_2=(1,1,2)$.

Pero, ¿y b_3?

Tenemos $f(b_3)=b_2+2b_3$, es decir, $f(b_3)-2b_3=b_2$, o lo que es lo mismo,

$(f-id_{\mathbf{R}^3}(b_3)=b_2$.

Luego, $b_3=(x_1,x_2,x_3)$ debe ser solución del sistema de ecuaciones lineales no homogéneo:

$$\begin{pmatrix} 1 & 1 & -1 \\ 2 & 0 & -1 \\ 2 & 2 & -2 \end{pmatrix}\cdot\begin{pmatrix} x_1 \\ x_2 \\ x_3 \end{pmatrix}=\begin{pmatrix} 1 \\ 1 \\ 2 \end{pmatrix}$$

Resolvamos dicho sistema.

$$\begin{cases} x_1 + x_2 - x_3 = 1 \\ 2x_1 \quad\;\; - x_3 = 1 \\ x_1 + 2x_2 - 2x_3 = 2 \end{cases}$$

Este sistema es compatible pues $rg\ A=2$ y el rango de la matriz ampliada es 2 como puede verificar fácilmente el estudiante. La solución está dada por

$$x_3 = 1 - 2x_1,$$
$$x_2 = x_1.$$

Cualquier solución particular del sistema puede ser tomada como vector b_3, por ejemplo:

$$b_3 = (0,0,-1).$$

Así resulta que una base en la cual se obtiene la forma de Jordán para f es

$$(b_1=(1,0,2),\ b_2=(1,1,2),\ b_3=(0,0,-1))$$

lo cual puede ser verificado por el estudiante.

Queremos llamar la atención sobre el método que acabamos de ejemplificar. Como observará el estudiante, la búsqueda de uno o más vectores de la base puede depender de la solución de un sistema de ecuaciones lineales no homogéneo y es bien conocido que estos sistemas no siempre tienen solución, esto es, no siempre son compatibles. Los vectores de la base se escogen tomando vectores propios y vectores que satisfagan sistemas no homogéneos del tipo $(f-\lambda_i\ id_E)(b_{i+1})=b_i$.

En nuestro ejemplo, $E(2)$ tiene dimensión 1 y por ello cualquier vector de $E(2)$ hubiera servido, ya que todos los vectores de $E(2)$ son proporcionales entre sí. En este caso, sin embargo, si dim $E(\lambda_i)$ es mayor que 1, lo que proporciona varias opciones de vectores linealmente independientes para b_i en el sistema $(f-\lambda_i id_E)(b_{i+1})=b_i$, puede ocurrir que algún b_i escogido no haga compatible el sistema. Entonces debe tomarse otro vector en $E(\lambda_i)$, pues siempre existirá en $E(\lambda_i)$ al menos un vector que haga compatible el sistema y permita hallar la base por este método.

Aplicaciones del Teorema de Jordán

Hemos analizado anteriormente que el problema de determinar en la práctica si dos matrices cuadradas son semejantes o no, es complejo, aun para el caso de matrices cuadradas de orden 2. Por la unicidad (salvo el orden de las celdas) de la forma de Jordán de un endomorfismo cuyo polinomio característico tenga todas sus raíces en K y, por tanto, de una matriz que cumpla este requisito, cada matriz será semejante a una matriz de Jordán única. Y como dos matrices semejantes a una tercera son semejantes entre sí, resulta que una condición necesaria y suficiente para que dos matrices de $M_n(K)$ cuyos polinomios característicos tengan sus raíces en K sean semejantes, es que tengan la misma forma de Jordán.

El criterio anterior permite resolver problemas como los siguientes: ¿Será una matriz compleja cuyos valores propios sean reales, semejante a una matriz real?

Dada A cuyo polinomio característico tenga sus raíces en K, serán A y ${}^tA$ semejantes?

La respuesta a la primera pregunta es inmediata y afirmativa ya que su matriz de Jordán tendrá solamente coeficientes reales y toda matriz es semejante a su matriz de Jordán.

La respuesta a la segunda pregunta es un poco menos inmediata. Sugerimos al estudiante que piense en la forma de la traspuesta de una matriz de Jordán y en la relación que existe, dada una matriz A y su matriz de Jordán J, entre A, ${}^tA$, J y ${}^tJ$.

Preguntas semejantes a las anteriores las encontrará el estudiante en los ejercicios al final del capítulo.

Quizás una de las aplicaciones más importantes del teorema de Jordán sea su uso en la parte de la Matemática conocida como Ecuaciones diferenciales, la cual es de gran interés práctico. Debido a que no es objeto de este texto abordar de manera profunda en las ecuaciones diferenciales, daremos una idea muy general de cómo se utiliza en estas el teorema de Jordán, sin realizar demostraciones.

Se denomina *sistema de ecuaciones diferenciales homogéneo de primer orden con coeficientes constantes* a un sistema de la forma

$$\frac{dx_1}{dt} = a_{11}x_1(t) + a_{12}x_2(t) + \ldots + a_{1n}x_n(t)$$

$$\frac{dx_2}{dt} = a_{21}x_1(t) + a_{22}x_2(t) + \ldots + a_{2n}x_n(t)$$

$$\vdots$$

$$\frac{dx_n}{dt} = a_{n1}x_1(t) + a_{n2}x_2(t) + \ldots + a_{nn}x_n(t)$$

donde $a_{ij} \in \mathbb{C}$ para todo i,j y $x_i(t) \in C^\infty(\mathbb{R})$ para todo $i=1,2,\ldots,n$.

Si consideramos el vector columna

$$X = \begin{pmatrix} x_1(t) \\ \cdot \\ \cdot \\ \cdot \\ x_n(t) \end{pmatrix}$$

y la matriz $A=(a_{ij})$, podemos dar al sistema la forma matricial:

$$\frac{dX}{dt}=AX.$$

Puede demostrarse que el conjunto de soluciones de $\frac{dX}{dt}=AX$ forman un subespacio vectorial, de dimensión n, del espacio de las funciones $C^{\infty}(\mathbf{R})$. Una base de tal espacio se denomina *sistema fundamental de soluciones.*

Consideremos entonces el problema de encontrar un sistema fundamental de soluciones de $\frac{dX}{dt}=AX.$

Puede probarse que si $X(t)$ es una solución de $\frac{dX}{dt}=AX$ para todo valor de t y para algún valor t_0 de t, X es un vector propio asociado al valor propio λ de A, es decir,

$$AX(t_0)=\lambda X(t_0).$$

Entonces se cumple que

$$AX(t)=\lambda X(t) \text{ para todo } t.$$

Llamaremos a $X(t)$ *solución propia* de $\frac{dX}{dt}=AX$ asociada a λ si y solo si $AX(t)=\lambda X(t)$ para todo t. Una solución es una solución propia asociada a λ si y solo si para algún t_0, es $X(t_0)$ un vector propio de A asociado a λ.

La importancia de las soluciones propias consiste en que pueden ser expresadas en forma muy simple.

Puede demostrarse que para cada valor propio de A existe al menos una solución propia, $X(t)$ de $\frac{dX}{dt}=AX$ la que se puede escribir en la forma

$$X(t)=(x_1e^{\lambda t}, x_2e^{\lambda t}, \ldots, x_ne^{\lambda t}),$$

donde $(x_1, x_2, \ldots, x_n)$ es un vector propio de A asociado a λ. Esto permitirá resolver el problema de hallar un sistema fundamental de soluciones de $\frac{dX}{dt}=AX.$

Está claro que sí A es una matriz diagonalizable, el problema está resuelto, pues podemos encontrar una base formada por vectores propios $X_1(t_0), \ldots, X_n(t_0)$ de A. El sistema correspondiente, $(X_1(t), \ldots, X_n(t))$ será un sistema fundamental de soluciones.

Ahora bien, sabemos que en general la matriz no será diagonalizable. ¿Cómo resolver el problema en este caso general?

En el caso anterior era fundamental el hecho de que si $X(t)$ es una solución de $\frac{dX}{dt}=AX(t)$ y para cierto t_0, era $X(t_0)$ un valor propio de A asociado a λ, entonces

$AX(t)=\lambda X(t)$ para todo t.

Ahora podemos generalizar este resultado como sigue:

Si $X(t)$ es una solución de $\frac{dX}{dt}=AX(t)$ y para algún valor t_0 de t $X(t_0)$ es un vector propio asociado al valor propio λ de A de multiplicidad m, entonces

$(A-\lambda I_m)^m\ X(t_0)=0$

$(A-\lambda I_m)^m\ X(t)=0$ *para todo t.*

Este último se puede probar.

Llamaremos entonces *solución primitiva* de $\frac{dX}{dt}=AX$ asociada al valor propio λ de multiplicidad m, a cualquier solución no trivial de $\frac{dX}{dt}=AX$ para la cual exista un valor propio λ de A y un entero m tal que $(A-\lambda I_n)^m X(t)=0$ para todo t, y $X(t)$ será una solución primitiva de $\frac{dX}{dt}=AX$ si y solo si para algún valor de t, es $X(t_0)$ un vector propio asociado al valor propio λ de A de multiplicidad m.

De la misma forma que las soluciones propias, las soluciones primitivas se pueden expresar en forma simple; exactamente:

Si $X(t)$ es una solución primitiva de $\frac{dX}{dt}=AX$ asociada al valor propio λ de multiplicidad m de A, entonces $X(t)$ es de la forma

$$X(t)=(P_1(t)e^{\lambda t},\ P_2(t)\ e^{\lambda t},\dots,P_n(t)\ e^{\lambda t}),$$

donde los $P_i\ (t)$ son polinomios de grado menor o igual que $(m-1)$.

¿Por qué podemos garantizar esto? En parte por el teorema de Jordán y por la descomposición que a partir de él se obtiene de toda matriz que cumpla las condiciones del teorema, como la suma de la matriz diagonal, D, y una matriz nilpotente N. Además, para ello debemos introducir el concepto *exponencial de una matriz.*

DEFINICIÓN 5.4

Sea A una matriz de orden n. Si la serie

$$I_n+A+\frac{A^2}{2!}+\dots+\frac{A^n}{n!}+\dots=\sum_{n=0}^{\infty}\frac{A^n}{n!}$$

es convergente, la suma de la serie se denota por e^A

$$e^A = \sum_{n=0}^{\infty} \frac{A^n}{n!}.$$

Esta función exponencial que hemos definido, cumple algunas de las propiedades que ya conocemos de la exponencial. Nos interesan especialmente las siguientes:

(1) Si A y B son matrices que conmutan, $AB=BA$, entonces $e^{A+B}=e^A \cdot e^B$.

(2) La derivada de las funciones $f(t)=e^{tA}$ es la función $f'(t)=Ae^{tA}$.

A partir de esta exponencial de matrices, podemos regresar al sistema

$$\frac{dX}{dt}=AX$$

con la condición inicial $X(0)=X_0$.

Podemos plantear el problema en la forma siguiente:

Supongamos que todas las raíces del polinomio característico de A estén en K. Entonces por el teorema de Jordán existe una matriz de Jordán J tal que $J=P^{-1}AP$ con P una matriz inversible. Además, J se puede descomponer en una suma:

$$J=D+N,$$

donde D es una matriz diagonal y N es una matriz nilpotente, y por otra parte, D y N conmutan.

Consideremos entonces el sistema

$$\frac{dY}{dt}=JY$$

con la condición inicial $Y(0)=Y_0$.

La solución de este sistema es el espacio generado por las columnas de la matriz $e^{Jt}Y_0$;

pero como $J=D+N$ tenemos:

$$e^{Jt}=e^{(D+N)t}=e^{Dt}e^{Nt}.$$

Calculemos e^{Dt} y e^{Nt}.

Si

$$D = \begin{pmatrix} \lambda_1 & & & 0 \\ & \lambda_2 & & \\ & & \ddots & \\ 0 & & & \lambda_n \end{pmatrix}$$

entonces

$$e^{Dt} = \begin{pmatrix} e^{\lambda_1 t} & & & 0 \\ & e^{\lambda_2 t} & & \\ & & \ddots & \\ 0 & & & e^{\lambda_n t} \end{pmatrix}$$

Para el otro término e^{Nt} empleamos la definición:

$$e^{Nt} = \left(I_n + Nt + \frac{N^2t^2}{2!} + \ldots + \frac{N^kt^k}{k!} \right),$$

pues como N es nilpotente, a partir de un cierto orden todas sus potencias se anulan. Por tanto e^{Nt} es una matriz cuyos elementos son polinomios en t con coeficientes en K.

e^{Jt} está formada por productos de polinomios en t por exponenciales de la forma $e^{\lambda_i t}$ (λ_i valor propio de A).

Las soluciones del sistema $\dfrac{dY}{dt} = JY$ están dadas como combinaciones lineales de vectores cuyos componentes son los descritos. Realizando el cambio de base inverso, podemos escribir la solución del sistema $\dfrac{dX}{dt} = AX$.

Veamos un ejemplo

Ejemplo

Consideremos el sistema de ecuaciones diferenciales:

$$\begin{cases} \dfrac{dx}{dt}=2x \\[2ex] \dfrac{dy}{dt}=x+y \\[2ex] \dfrac{dz}{dt}=x+y+z \end{cases}$$

Podemos escribir este sistema en la forma

$$\frac{dX}{dt}=AX,$$

con las condiciones iniciales

$$X(0)=X_0=(a_1,a_2,a_3) \text{ y } A=\begin{pmatrix} 2 & 0 & 0 \\ -1 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix}$$

Llevamos A a la forma de Jordán:

$$J=\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}$$

La matriz P inversible tal que $J=P^{-1}AP$ es

$$P=\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix}$$

Así obtenemos un nuevo sistema:

$$\frac{dY}{dt}=JY,$$

con la condición inicial $Y(0)=(c_1,c_2,c_3)$. Entonces $J=D+N$, donde

$$J=\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 2 \end{pmatrix}+\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix} \text{ y } N^2=0.$$

Por tanto,

$$e^D=\begin{pmatrix} e & 0 & 0 \\ 0 & e & 0 \\ 0 & 0 & e^2 \end{pmatrix}, \quad e^N=\begin{pmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix},$$

$$e^{Dt}=\begin{pmatrix} e^t & 0 & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{2t} \end{pmatrix}, \quad e^{Nt}=\begin{pmatrix} 1 & t & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Luego, la solución del sistema es:

$$T(t)=e^{Jt}Y_0=e^{Dt}\ e^{Nt}Y_0$$

$$=\begin{pmatrix} e^t & te^t & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{2t} \end{pmatrix} Y_0$$

$$=\begin{pmatrix} e^t & te^t & 0 \\ 0 & e^t & 0 \\ 0 & 0 & e^{2t} \end{pmatrix} \cdot \begin{pmatrix} c_1 \\ c_2 \\ c_3 \end{pmatrix}$$

$$=(c_1e^t+c_2te^t;\ c_2e^t,\ c_3e^{2t}).$$

Haciendo el cambio inverso de coordenadas obtenemos:

$$X(t)=PY(t)$$

$$X(t)=\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & -1 \\ 1 & -1 & 0 \end{pmatrix} \cdot \begin{pmatrix} c_1e^t+c_2te^t \\ c_2e^t \\ c_3e^{2t} \end{pmatrix}$$

Por tanto,

$$X(t)=(c_3c^{2t},\ c_2e^t-c_3e^{2t},\ (c_1-c_2)e^t+c_2te^t).$$

Y como la relación entre (c_1,c_2,c_3) y (a_1,a_2,a_3) es conocida, podemos plantear:

$$X(t)=(a_1e^{2t},\ (a_1+a_2)\ e^t-a_1e^{2t},\ a_3e^t+(a_1+a_2)te^t).$$

Comprobamos:

$$X(0)=(a_1,a_2,a_3).$$

El cambio de base que hacemos con la matriz P es un cambio en el espacio de las soluciones del sistema de ecuaciones $\dfrac{dX}{dt}=AX$. Este método no solamente ayuda a resolver un sistema de ecuaciones diferenciales, sino que a partir de este enfoque se obtienen muchos resultados analíticos sobre las soluciones del sistema.

Ejercicios

1. Pruebe que si A y P son matrices de $M_n(K)$, con P inversible y $p(x)\in K\{x\}$, entonces
   $p(P^{-1}AP)=P^{-1}(p(A))\ P.$
2. Pruebe que dos matrices semejantes tienen el mismo polinomio minimal.

3. Busque los polinomios minimales de las matrices siguientes:

a) $\begin{pmatrix} 2 & 1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 0 & 0 & -2 & 4 \end{pmatrix}$

b) $\begin{pmatrix} 2 & 5 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 \\ 0 & 0 & 4 & 2 & 0 \\ 0 & 0 & 3 & 5 & 0 \\ 0 & 0 & 0 & 0 & 7 \end{pmatrix}$

c) $\begin{pmatrix} 3 & 1 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 \\ 0 & 0 & 3 & 1 & 0 \\ 0 & 0 & 0 & 3 & 1 \\ 0 & 0 & 0 & 0 & 3 \end{pmatrix}$

d) $\begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 0 \\ 0 & 0 & 1 \end{pmatrix}$

e) $\begin{pmatrix} 2 & 0 & 0 \\ 0 & 2 & 2 \\ 0 & 0 & 1 \end{pmatrix}$

f) $\begin{pmatrix} \lambda & a & 0 \\ 0 & \lambda & a \\ 0 & 0 & \lambda \end{pmatrix}, a \neq 0$

g) $\begin{pmatrix} \lambda & a & 0 & 0 \\ 0 & \lambda & a & 0 \\ 0 & 0 & \lambda & a \\ 0 & 0 & 0 & \lambda \end{pmatrix}\ a \neq 0$

4. Halle el polinomio minimal de

$$A = \begin{pmatrix} 2 & 8 & 0 & 0 & 0 & 0 & 0 \\ 0 & 2 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 4 & 2 & 0 & 0 & 0 \\ 0 & 0 & 1 & 3 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5 \end{pmatrix}$$

Sugerencia: Observe que

$$A = \begin{pmatrix} A_1 & 0 & 0 & 0 \\ 0 & A_2 & 0 & 0 \\ 0 & 0 & A_3 & 0 \\ 0 & 0 & 0 & A_4 \end{pmatrix}$$

con $A_1 = \begin{pmatrix} 2 & 8 \\ 0 & 2 \end{pmatrix}$, $A_2 = \begin{pmatrix} 4 & 2 \\ 1 & 3 \end{pmatrix}$, $A_3 = \begin{pmatrix} 0 & 3 \\ 0 & 0 \end{pmatrix}$, $A_4 = (5)$.

5. Sean E un espacio vectorial de dimensión n sobre K y f un endomorfismo inversible de E. Pruebe que f^{-1} es un polinomio en f cuyo grado es menor o igual que n.
Sugerencia: Considere el polinomio minimal w_f de f y recuerde que f es inversible si y solo si 0 no es raíz del polinomio minimal de f y que $w_f(f) = 0$.

6. Considere la matriz de $M_n(K)$:

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

Pruebe que $p_A(x)=w_A(x)=(x-\lambda)^n$.

7. Pruebe que un endomorfismo f de un espacio E es una homotecia de parámetro k (esto es, $f(x)=kx \ \forall x \in E$) si y solo si $w_f(x)=x-k$.

8. Pruebe que una matriz y su traspuesta tienen el mismo polinomio minimal.

9. Sea $p(x)$ un polinomio mónico de $K[X]$, de grado n.

$p(x)= x^n+a_{n-1}x^{n-1}+\dots+a_1x+a_0$.

Se denomina *matriz compañera* de $p(x)$ a:

$$A=\begin{pmatrix} 0 & 0 \dots 0 & -a_0 \\ 1 & 0 \dots 0 & -a_1 \\ 0 & 1 \dots 0 & -a_2 \\ \dots & \dots & \dots \\ 0 & 0 \dots 1 & -a_{n-1} \end{pmatrix}.$$

Pruebe que $p(x)$ es el polinomio minimal de A.

Sugerencia: Halle el polinomio característico de A desarrollando el determinante de $A-xI_n$ por la primera columna y usando inducción en n.

10. Sea el espacio de los polinomios con coeficientes reales de grado menor o igual que n y sea D el endomorfismo dado por $D(p)=p'$ (donde p' es el polinomio derivada de p). Halle el polinomio minimal de D.

11. Sean $M_n(K)$ el espacio de las matrices cuadradas de orden n con coeficiente en K y T_A el endomorfismo de $M_n(K)$ dado por $T_A(B)=AB$ donde A es una matriz fija de $M_n(K)$. Pruebe que el polinomio minimal de T_A es el de A.

12. Para las siguientes matrices encuentre una matriz T semejante a la dada y una diagonal D y una nilpotente N tales que $D+N=T$.

a) $$A=\begin{pmatrix} 0 & 3 & 3 \\ -1 & 8 & 6 \\ 2 & -14 & -10 \end{pmatrix}$$

c) $$A=\begin{pmatrix} 3 & 1 & -1 \\ 2 & 2 & -1 \\ 2 & 2 & 0 \end{pmatrix}$$

b) $$A=\begin{pmatrix} 6 & -3 & -2 \\ 4 & -1 & -2 \\ 10 & -5 & -3 \end{pmatrix}$$

13. Aplique el teorema de descomposición primaria para obtener una matriz diagonal por bloques triangulares semejante a:

a) $\begin{pmatrix} 1 & 0 & 0 & 0 \\ -1 & 2 & 1 & 1 \\ -3 & -1 & 5 & 3 \\ 2 & 1 & -3 & -1 \end{pmatrix}$ b) $\begin{pmatrix} 1 & 1 & -1 & 2 & -1 \\ 2 & 0 & 1 & -4 & -1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & -3 & 3 & -1 \end{pmatrix}$

14. Pruebe que si u y v son dos endomorfismos nilpotentes de E que conmutan, esto es, $u \circ v = v \circ u$, entonces $u+v$ y $u \circ v$ son también nilpotentes.

15. Pruebe que en el espacio de los polinomios con coeficientes reales de grado menor o igual que n, el endomorfismo que a cada polinomio asigna su derivada es nilpotente de índice $n+1$.

16. Pruebe que dos matrices nilpotentes de orden 3 son semejantes si y solo si tienen el mismo orden de nilpotencia. Verifique mediante un ejemplo que esto no es válido para matrices de orden estrictamente mayor que 3.

17. Sean E un espacio de dimensión n sobre K y f un endomorfismo de E de rango 1. Pruebe que f no puede ser a la vez nilpotente y diagonalizable.

18. Sean $E = M_n(K)$ y A una matriz fija de $M_n(K)$. Considere el endomorfismo de E dado por
$T(B) = AB - BA$.
Pruebe que si A es nilpotente, entonces T es un endomorfismo nilpotente.

19. Aplique el teorema de descomposición de los endomorfismos nilpotentes a los endomorfismos dados por las siguientes matrices, dando en cada caso la base en la cual se obtiene la forma reducida del endomorfismo nilpotente.

a) $\begin{pmatrix} 2 & -8 & 12 & -60 \\ 2 & -5 & 9 & -48 \\ 6 & -17 & 29 & -152 \\ 1 & -3 & 5 & -26 \end{pmatrix}$ b) $\begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 0 \\ 0 & 0 & 0 & 0 & 4 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix}$

20. Determine todas las formas de Jordán posibles para los endomorfismos siguientes, atendiendo a las características dadas para dichos endomorfismos.

a) $p_f = (x-2)^4(x-3)^2, w_f = (x-2)^2(x-3)^2$
b) $p_f = (x-7)^5, w_f = (x-7)^2$
c) $p_f = (x-2)^7,\ w_f = (x-2)^3$
d) $p_f = (x-3)^4(x-5)^4,\ w_f = (x-3)^2(x-5)^2$
e) $p_f = (x-2)^3(x+7)^2,\ w_f = (x-2)^2(x+7)$
f) $p_f = (x+2)^4(x-1)^2,\ w_f = (x+2)^2(x-1)$

21. Pruebe que toda matriz compleja es semejante a su traspuesta.

22. Busque las formas de Jordán para los endomorfismos representados por las matrices siguientes:

a) $\begin{pmatrix} 0 & 3 & 3 \\ -1 & 8 & 6 \\ 2 & 14 & -10 \end{pmatrix}$ d) $\begin{pmatrix} 1 & 1 & -1 & 2 & 1 \\ 2 & 0 & 1 & -4 & -1 \\ 0 & 1 & 1 & 1 & 1 \\ 0 & 1 & 2 & 0 & 1 \\ 0 & 0 & -3 & 3 & -1 \end{pmatrix}$

b) $\begin{pmatrix} 3 & 2 & -3 \\ 4 & 10 & -12 \\ 3 & 6 & -7 \end{pmatrix}$ e) $\begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 2 & 0 \\ 0 & 0 & 0 & 3 \\ 0 & 0 & 0 & 0 \end{pmatrix}$

c) $\begin{pmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & 1 & -1 \\ 1 & -1 & 1 & 1 \\ 1 & 1 & -1 & 1 \end{pmatrix}$ f) $\begin{pmatrix} 2 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2 & 0 & 0 & 0 & 0 \\ -1 & 0 & 2 & 0 & 0 & 0 \\ 0 & 1 & 0 & 2 & 0 & 0 \\ 1 & 1 & 1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 & 1 & -1 \end{pmatrix}$

En a), b) y c) halle la base en la cual el endomorfismo se representa mediante la forma de Jordán.

23. Pruebe que si dos matrices A y B de $M_n(\mathbb{C})$ son tales que tienen el mismo polinomio característico y minimal y que la multiplicidad de ningún valor propio de A y B como raíz del polinomio característico, excede a 3, entonces son semejantes.

24. Si A es una matriz de $M_n(K)$ cuyo polinomio característico es
$P_A=(x-\lambda_1)^{m_1}...(x-\lambda_k)^{m_k}$ con $\lambda_i\in K$,
¿cuál es la traza de A?

25. Sean E un espacio vectorial de dimensión finita n *sobre* K y f un endomorfismo de E tal que en una base $(a_1,a_2,...,a_n)$ de dicho espacio tiene por matriz a $A=(a_{ij})$, donde
$a_{ij}=0$ si $j>i$,
$a_{ii}=\lambda$ para todo i
y existen i y j tales que $j<i$ y $a_{ij}\neq 0$.
Pruebe que λ no puede ser raíz simple del polinomio minimal de f.
Sugerencia: Pruebe primero con $n=2$ o 3.

26. Sea f un endomorfismo de un espacio vectorial de dimensión finita sobre K. Sea $x\in E$, $x\neq 0$, y k el mayor número natural tal que el sistema $[x, f(x),...,f^{k-1}(x))$ sea linealmente independiente.
   a) Demuestre que este sistema de vectores genera un subespacio V de E estable por f y que existen $a_0,a_1,...,a_{k-1}$ en K únicos tales que $f^k(x)=a_0x+a_1f(x)+...+a_{k-1}f^{k-1}(x)$.
   b) Halle un polinomio p de $K[x]$ de grado k tal que si f_V es la restricción de f a V, entonces $p(f_V)=0$.

Bibliografía

CHAMBADAL, L. y L.L. OVAERT: *Algebre Lineaire et Algebre Tensorielle.* Dunod, Paris, 1968.

FADDIEEV, D. y I. SOMINSKI: *Problemas de Álgebra Superior.* Editorial MIR, Moscú, 1971.

GODEMENT, R.: *Álgebra.* Ciencia y Técnica, La Habana, 1970.

HALMOS, P.: *Finite Dimensional Vector Spaces.* D. Van Nostrand Co., Princeton, 1958.

HOFFMAN, K. y R. KUNZE: *Linear Algebra.* Prentice Hall, New Jersey, 1971.

KUROSH, A.G.: *Curso de Álgebra Superior.* Editorial MIR, Moscú, 1968.

LANG, S.: *Linear Algebra.* Editorial Pueblo y Educación, La Habana, 1971.

LIPSCHUTZ, S.: *Álgebra Lineal.* Edición Revolucionaria, La Habana, 1974.

MALTSEV, A.I.: *Fundamentos de Álgebra Lineal.* Editorial MIR, Moscú, 1976.

MUTAFIAN, C.: *La Structure Vectorielle.* Vuibert, Paris, 1979.

——: *Les Applications Lineaires.* Vuibert, Paris, 1979.

QUEYSANNE, M.: *Álgebra Básica.* Editorial Vicens-Vives, Barcelona, 1973.

Índice de materias