APUNTES DE ALGEBRA 2

(revisión 2 de abril, 2017)

Juan Pablo Muszkats
Isabel Pustilnik

FACULTAD
DE INGENIERIA
Universidad de Buenos Aires

Prólogo

Prólogo

Estas notas tienen por objetivo cubrir todos los temas teóricos correspondientes a la asignatura Álgebra 2, tal y como se dicta actualmente en la Facultad de Ingeniería de la Universidad de Buenos Aires. Para que resulten de utilidad a los alumnos que las usen, consideramos oportuno hacer algunas aclaraciones y advertencias:

- Estos apuntes consisten, a grandes rasgos, de una exposición de los temas teóricos junto con algunos ejemplos y ejercicios. Hemos intentado dar algunas motivaciones y justificaciones de los teoremas expuestos, abarcando las demostraciones que consideramos más formativas para los estudiantes.
- Quienes quieran profundizar tanto en las aplicaciones del álgebra lineal como en los detalles técnicos de las demostraciones más exigentes, encontrarán a lo largo del texto indicaciones bibliográficas de los textos que consideramos adecuados para ello.
- Aunque sea una obviedad decirlo, *un texto de matemática no se lee como una novela*. Un texto de matemática se estudia de maneras diversas que cada lector debe encarar, explorar y evolucionar. Una primera lectura podría ser más superficial, omitiendo las demostraciones y concentrándose en las definiciones y resultados importantes. Una segunda (o tercera) lectura podría detenerse en los detalles de las demostraciones. Creemos que el estudio de las demostraciones es formativo y permite madurar los conceptos involucrados, lo cual las hace relevantes aun para quienes no estudian la matemática como una profesión. Además, la lectura atenta de las demostraciones suele echar luz sobre los alcances de un teorema: para un ingeniero es importante, por ejemplo, distinguir si un teorema tan solo prueba la existencia de una solución o brinda además un método para calcularla efectivamente. Esta es la clase de habilidades que deseamos fomentar en nuestros lectores.
- Para facilitar la lectura, hemos incorporado recuadros para destacar las ideas principales, como las definiciones y los teoremas, así como también delimitadores para los ejemplos:

Definición 0.1

Teorema 0.1

Ejemplo 0.1

y el siguiente símbolo para las notas, comentarios u observaciones

- Toda vez que se aluda a un teorema, fórmula, etc. como por ejemplo

  ...en el teorema 1.1 y la fórmula (1.1)

  el número funciona como vínculo, y permite llegar directamente al objeto aludido.

- A lo largo del texto aparecen algunos ejercicios breves para reforzar los conceptos que se exponen:

  Ejercicio.

  Muchas de las respuestas se encuentran al final.

Al realizar este trabajo, hemos intentado aportar nuestra experiencia en la asignatura y nuestro conocimiento de las dificultades que atraviesan los estudiantes al cursarla. Como esperamos sinceramente que les resulte de utilidad, invitamos a todos nuestros lectores a compartir cualquier aporte, corrección o crítica que nos permita mejorarlo.

Por último, agradecemos la generosidad del profesor José Luis Mancilla Aguilar que puso a nuestra disposición sus valiosos apuntes, de los que tomamos numerosas ideas y ejemplos.

Juan Pablo Muszkats e Isabel Pustilnik
Buenos Aires, diciembre de 2016.

Índice general

Esta página fue dejada intencionalmente en blanco

1 Espacios vectoriales

1.1 Introducción

El concepto de vector, que usualmente comienza por estudiarse en $\mathbb{R}^2$ y $\mathbb{R}^3$, se generaliza mediante la siguiente definición.

Definición 1.1

Un ***espacio vectorial*** es un conjunto no vacío V en el que están definidas una ley de composición interna denominada *suma*, que a cada par de elementos $u, v \in V$ le asigna un elemento $w \in V$ denotado $w = u+v$, y una ley de composición externa denominada *producto por escalares*, que a cada número $\alpha \in \mathbb{K}$ (siendo $\mathbb{K} = \mathbb{R}$ ó $\mathbb{C}$) y a cada elemento $u \in V$ les asigna un elemento $v \in V$ denotado $v = \alpha u$, de manera tal que se verifican las siguientes condiciones para todos $u, v, w \in V$, $\alpha, \beta \in \mathbb{K}$:

1. *conmutatividad*: $u + v = v + u$
2. *asociatividad*: $u + (v + w) = (u + v) + w$
3. existencia de un *elemento neutro* $0_V \in V$ tal que $u + 0_V = u$
4. para cada $u \in V$ existe un *opuesto*, denotado $-u$, tal que $u + (-u) = 0_V$
5. *distributividad* respecto de los escalares: $(\alpha + \beta)u = \alpha u + \beta u$
6. *distributividad* respecto de los vectores: $\alpha(u + v) = \alpha u + \alpha v$
7. *asociatividad* respecto de los escalares: $\alpha(\beta u) = (\alpha\beta)u$
8. 1u=u

A los elementos del espacio vectorial V se los denomina ***vectores***, mientras que a los números del conjunto $\mathbb{K}$ se los denomina ***escalares***.

En el caso de que sea $\mathbb{K} = \mathbb{R}$, se dice que V es un *espacio vectorial real* o un $\mathbb{R}$-*espacio vectorial*, mientras que si $\mathbb{K} = \mathbb{C}$ se dice que V es un *espacio vectorial complejo* o un $\mathbb{C}$-*espacio vectorial*. Diremos simplemente que V es un *espacio vectorial* o un $\mathbb{K}$-*espacio vectorial* cuando no sea preciso distinguir el conjunto de escalares. La notación habitual para los vectores será

con las letras latinas minúsculas $u, v, w, ...$, mientras que para los escalares se usarán las letras griegas $\alpha, \beta, ...$ En ocasiones se notará mediante 0 al vector nulo 0_V: el contexto debe permitir al lector saber a qué se está aludiendo en cada caso.

Veamos ahora algunos ejemplos de espacios vectoriales.

Ejemplo 1.1 Vectores en $\mathbb{R}^n$

Al conjunto de n-uplas $(x_1, ..., x_n)$ con $x_i \in \mathbb{R}$ para todo $i = 1, ..., n$ lo denotaremos $\mathbb{R}^n$. Definiendo la suma de dos n-uplas arbitrarias mediante

$$(x_1, ..., x_n) + (y_1, ..., y_n) = (x_1 + y_1, ..., x_n + y_n)$$

y el producto por escalares como

$$\alpha (x_1, ..., x_n) = (\alpha x_1, ..., \alpha x_n)$$

se obtiene un espacio vectorial. Queda como ejercicio para el lector identificar el neutro, los opuestos y comprobar que se cumplen todas las condiciones de un espacio vectorial.

Ejemplo 1.2 Vectores en $\mathbb{C}^n$

Si en lugar de n-uplas de números reales consideramos n-uplas de números complejos $(z_1, ..., z_n)$ con $z_i \in \mathbb{C}$, definiendo las operaciones de manera análoga al espacio anterior obtenemos en este caso un $\mathbb{C}$-espacio vectorial, al que denotamos $\mathbb{C}^n$.

Muchas de las propiedades conocidas para los vectores de $\mathbb{R}^2$ y $\mathbb{R}^3$ se conservan para el caso general. A partir de la definición de espacio vectorial se deducen las siguientes propiedades elementales.

Teorema 1.1 Propiedades elementales

1. El elemento neutro $0_V \in V$ es único. También es único el elemento opuesto que corresponde a cada $u \in V$
2. $\forall u \in V : 0u = 0_V$
3. $\forall \alpha \in \mathbb{K} : \alpha 0_V = 0_V$
4. Si $u \neq 0_V \wedge \alpha u = \beta u \Rightarrow \alpha = \beta$
5. Si $\alpha \neq 0 \wedge \alpha u = \alpha v \Rightarrow u = v$
6. $\forall u \in V : (-1)u = -u$

Demostración:
Para probar la unicidad del neutro, se considera que existen dos neutros 0_1 y 0_2 en V. Entonces, por la misma definición de neutro resulta que $0_1 + 0_2 = 0_1$, a la vez que $0_2 + 0_1 = 0_2$. Además, la propiedad conmutativa garantiza que $0_1 + 0_2 = 0_2 + 0_1$ de donde se deduce que $0_1 = 0_2$.
Para probar 2, consideramos que

$$u = 1u = (1 + 0)u = 1u + 0u = u + 0u$$

(el lector deberá explicar qué propiedad justifica cada paso). Entonces $0u$ es el neutro del espacio vectorial, con lo cual $0u = 0_V$.

Dejamos la demostración del resto de las propiedades como ejercicio para el lector. □

Veamos algunos ejemplos más de espacios vectoriales que serán frecuentes a lo largo del curso.

Ejemplo 1.3 Matrices en $\mathbb{K}^{m\times n}$

$\mathbb{K}^{m\times n}$ es el conjunto de matrices de m filas y n columnas cuyas componentes son elementos de $\mathbb{K}$, con las definiciones de suma y producto por escalar usuales:

$$\begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} + \begin{bmatrix} b_{11} & \cdots & b_{1n} \\ \vdots & \ddots & \vdots \\ b_{m1} & \cdots & b_{mn} \end{bmatrix} = \begin{bmatrix} a_{11}+b_{11} & \cdots & a_{1n}+b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1}+b_{m1} & \cdots & a_{mn}+b_{mn} \end{bmatrix}$$

$$\alpha \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{bmatrix} = \begin{bmatrix} \alpha a_{11} & \cdots & \alpha a_{1n} \\ \vdots & \ddots & \vdots \\ \alpha a_{m1} & \cdots & \alpha a_{mn} \end{bmatrix}$$

El elemento neutro es la matriz nula, que es aquella cuyos elementos son todos nulos. El opuesto de una matriz A con elementos de la forma a_{ij} es la matriz $-A$ con elementos de la forma $-a_{ij}$.

Ejemplo 1.4 Espacio de funciones

Un espacio vectorial real que aparece frecuentemente en las aplicaciones es el de las funciones $f : I \to \mathbb{R}$ definidas en un intervalo I de la recta real, que denotaremos $\mathbb{R}^I$, con la suma y producto por escalar como usualmente se definen para funciones:

- *suma*: dadas $f, g \in \mathbb{R}^I$, se define una nueva función $f + g : I \to \mathbb{R}$ mediante

$$(f+g)(x) = f(x) + g(x) \text{ para todo } x \in I$$

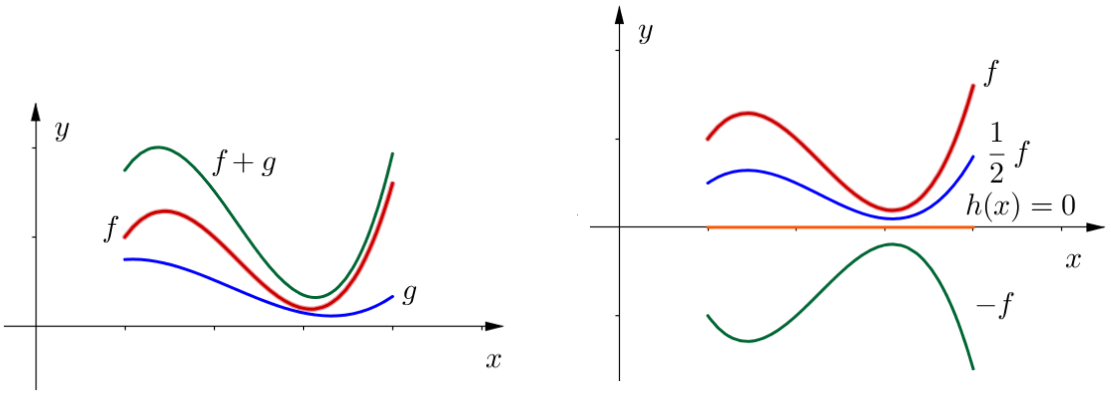
- *producto por escalar*: dados $f \in \mathbb{R}^I$ y $\alpha \in \mathbb{R}$, se define una nueva función $\alpha f : I \to \mathbb{R}$ mediante

$$(\alpha f)(x) = \alpha f(x) \text{ para todo } x \in I$$

El elemento neutro para la suma es la función idénticamente nula, es decir, la función h tal que para todo $x \in I$ verifica $h(x) = 0$ (es crucial distinguir a la función nula de las funciones que eventualmente se anulan para algún $x \in I$). Por otra parte, el opuesto de una función $f \in \mathbb{R}^I$ es la función $-f$ definida mediante $(-f)(x) = -(f(x))$ para todo $x \in I$. Por ejemplo, si consideramos el intervalo abierto $I = (0, 1)$, las funciones $f(x) = x + \frac{1}{x}$ y $g(x) = \frac{1}{x-1}$ son efectivamente elementos de $\mathbb{R}^I$ y quedan definidas

$$(f+g)(x) = x + \frac{1}{x} + \frac{1}{x-1} = \frac{x^3 - x^2 + 2x - 1}{x^2 - x}$$

$$(3f)(x) = 3x + \frac{3}{x}$$



(a) suma de funciones

(b) producto por escalar, función nula y opuesta

Figura 1.1: las funciones como vectores

Ejemplo 1.5 Espacio de funciones continuas

Consideremos un intervalo I de la recta real y denominemos $\mathscr{C}(I)$ al conjunto formado por todas las funciones $f : I \to \mathbb{R}$ continuas. $\mathscr{C}(I)$ es un espacio vectorial real con las operaciones de suma y producto por escalar definidas en el ejemplo anterior. Nótese que ambas operaciones son cerradas en $\mathscr{C}(I)$, en el sentido de que sus resultados son elementos de $\mathscr{C}(I)$, porque tanto la suma como el producto de funciones continuas son funciones continuas.

La función nula, que es continua y por ende pertenece a $\mathscr{C}(I)$, es el elemento neutro para la suma. Por otra parte, el opuesto de $f \in \mathscr{C}(I)$ es la función continua $-f$.

Ejemplo 1.6 Espacio de polinomios

Otro espacio vectorial real que se usará con frecuencia es el de los polinomios a coeficientes reales

$$\mathscr{P} = \{p : p(t) = a_n t^n + a_{n-1} t^{n-1} + \ldots + a_0, n \in \mathbb{N}_0, a_i \in \mathbb{R}, i = 0, \ldots, n\}$$

con las operaciones de suma y producto por un escalar habituales.

Queda como ejercicio para el lector comprobar que los polinomios, junto con las operaciones indicadas, satisfacen la definición de espacio vectorial, e identificar los elementos neutro y opuesto.

Los espacios vectoriales reales o complejos que mencionamos hasta aquí son algunos de los que aparecen con más frecuencia en las aplicaciones del álgebra lineal a la ingeniería. El ejercicio que sigue tiene por objeto repasar la definición de espacio vectorial y mostrar que existen muchas más posibilidades que las que convencionalmente se usan.

Ejercicio.

1.1 Considere $V = \mathbb{R}^2_+$, donde $\mathbb{R}^2_+$ es el conjunto de pares ordenados compuestos por números reales positivos. Por ejemplo $(1;1,5) \in \mathbb{R}^2_+$, mientras que $(1,0) \notin \mathbb{R}^2_+$ y $(-1,5) \notin \mathbb{R}^2_+$. En V definimos las siguientes operaciones de suma y producto para $(x_1, x_2), (y_1, y_2) \in V$ y $\alpha \in \mathbb{R}$:

- $(x_1, x_2) + (y_1, y_2) = (x_1 y_1, x_2 y_2)$
- $\alpha(x_1, x_2) = \left(x_1^\alpha, x_2^\alpha\right)$

Demuestre que el conjunto V con las operaciones definidas es un espacio vectorial.

N Luego de haber visto algunos ejemplos de espacios vectoriales, es importante notar lo siguiente: todo espacio vectorial complejo es a su vez un espacio vectorial real si se considera la misma operación suma y el producto por escalares se restringe a números reales (¿por qué?). Por ende, será importante *dejar siempre en claro cuál es el conjunto de escalares que se considera.*

Ejemplo 1.7

$\mathbb{C}^2$ es un espacio vectorial complejo y, por lo dicho recién, también es un espacio vectorial real. Sin embargo, si consideramos a $\mathbb{C}^2$ como espacio vectorial real, el producto $(2+i)(1,-i)$ carece de sentido, pues $\alpha = (2+i)$ no es un número real. Tampoco es válida la descomposición

$$(1+2i, i) = (1+2i)(1,0) + i(0,1)$$

(que sí vale si consideramos a $\mathbb{C}^2$ como espacio vectorial complejo). Una descomposición posible en este caso es

$$(1+2i, i) = 1(1,0) + 2(i,0) + 1(0,i)$$

Veremos más adelante que, a pesar de tratarse del mismo conjunto de vectores, sus propiedades como espacio vectorial real serán diferentes de sus propiedades como espacio vectorial complejo.

Convención

- De aquí en más, por comodidad, vamos a convenir en identificar $\mathbb{R}^n$ y $\mathbb{C}^n$ con $\mathbb{R}^{n\times 1}$ y $\mathbb{C}^{n\times 1}$ respectivamente. Es decir, cuando sea conveniente escribiremos
$$u = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$
en lugar de $u = (x_1, ..., x_n)$.
- A menos que se especifique lo contrario, siempre consideraremos tanto a $\mathbb{C}^n$ como a $\mathbb{C}^{m\times n}$ como $\mathbb{C}$-espacios vectoriales.

1.2 Subespacios

Algunos de los espacios vectoriales que hemos estudiado están relacionados entre sí. Por ejemplo, $\mathscr{C}(I)$ es un subconjunto de $\mathbb{R}^I$. A su vez, $\mathscr{P}$ puede pensarse como subconjunto de $\mathscr{C}(\mathbb{R})$, ya que cada polinomio define una función continua en $\mathbb{R}$.

Por cierto, las operaciones definidas en $\mathbb{R}^I$ y en $\mathscr{C}(\mathbb{R})$ valen para elementos de los subconjuntos $\mathscr{C}(I)$ y $\mathscr{P}$ respectivamente. Esta situación - la de tener un subconjunto de un espacio vectorial V que, con las operaciones definidas en V, también es un espacio vectorial - es muy común y tiene consecuencias importantes. Esto motiva la siguiente definición.

Definición 1.2

Un subconjunto S de un espacio vectorial V es un ***subespacio*** de V si S es un espacio vectorial con las operaciones definidas en V.

Ejemplo 1.8

De acuerdo con las observaciones hechas al comienzo de esta sección y con la definición de subespacio, es inmediato que $\mathscr{C}(I)$ es un subespacio de $\mathbb{R}^I$ y que $\mathscr{P}$ es un subespacio de $\mathscr{C}(\mathbb{R})$.

En principio, para comprobar que cierto subconjunto S de cierto espacio vectorial V es un subespacio, debería comprobarse formalmente que se cumplen en S todas las propiedades exigidas en la definición de espacio vectorial 1.1. Sin embargo, las propiedades $1-8$ se cumplen para todos los elementos de V: luego se cumplen en particular para los elementos del subconjunto S. Sólo resta probar que S no es vacío y es cerrado para las operaciones con sus elementos. Resumimos esta observación en el teorema siguiente.

Teorema 1.2 (del Subespacio)

Sea S un subconjunto de un espacio vectorial V. Entonces S es subespacio de V si y sólo si se verifican las siguientes condiciones:

1. S no es vacío
2. $u, v \in S \Rightarrow u + v \in S$
3. $u \in S, \alpha \in \mathbb{K} \Rightarrow \alpha u \in S$

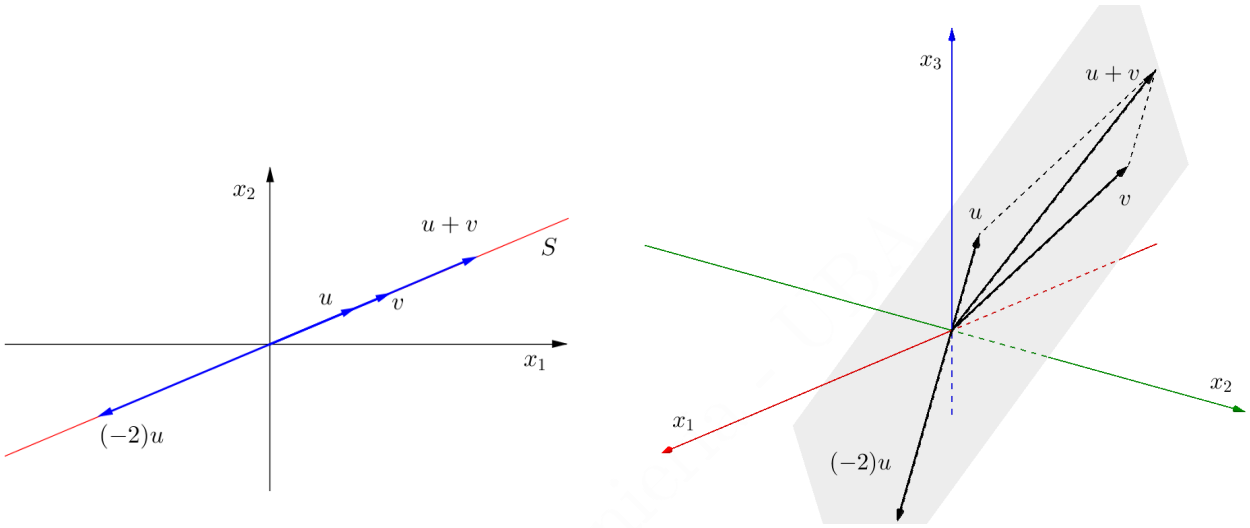
Ejemplo 1.9 Subespacios triviales

Claramente $S = \{0_V\}$ (el conjunto que solamente contiene al vector nulo del espacio) y $S = V$ son subespacios de V. A estos subespacios se los denomina *subespacios triviales* .

Ejemplo 1.10

En $\mathbb{R}^2$ toda recta que pasa por el origen es un subespacio. Esta afirmación puede justificarse mediante el teorema del subespacio apelando a las interpretaciones geométricas de la adición de vectores en el plano (regla del paralelogramo) y del producto de vectores por escalares. En efecto, S no es vacío porque el origen de coordenadas pertenece a S; si sumamos dos vectores contenidos en la recta S el vector resultante pertenece a la misma recta y, por último, si multiplicamos un vector cualquiera de S por un escalar, el vector que obtenemos tiene la misma dirección que el primero y por lo tanto también pertenece a S.

Mediante argumentos similares también se puede justificar que en $\mathbb{R}^3$ toda recta o plano que contenga al origen es un subespacio.



(a) recta en $\mathbb{R}^2$ **(b)** plano en $\mathbb{R}^3$

Figura 1.2: Subespacios en $\mathbb{R}^2$ y $\mathbb{R}^3$

Ejemplo 1.11 Subespacio asociado a un sistema homogéneo

Sea $A \in \mathbb{K}^{m\times n}$. Entonces

$$S = \{x \in \mathbb{K}^n : Ax = 0\} \tag{1.1}$$

es un subespacio de $\mathbb{K}^n$. Para comprobarlo usaremos nuevamente el teorema del subespacio.

1. S no es vacío porque $0 \in S$, dado que $A0 = 0$.
2. Sean $u, v \in S$. Entonces, por la misma definición de S se cumple que $Au = 0$ y $Av = 0$. Por lo tanto $A(u+v) = Au + Av = 0 + 0 = 0$, de donde se deduce que $u + v \in S$.
3. Sea $u \in S$: luego $Au = 0$. Dado un escalar arbitrario $\alpha \in \mathbb{K}$, tenemos que $A(\alpha u) = \alpha(Au) = \alpha 0 = 0$. Vale decir que $\alpha u \in S$.

A partir de este resultado se puede probar formalmente lo que se enunció en el ejemplo anterior. En efecto, si S es una recta en $\mathbb{R}^2$ que contiene al origen, sus puntos verifican una ecuación de la forma $ax_1 + bx_2 = 0$. Es decir que $S = \{x \in \mathbb{R}^2 : ax_1 + bx_2 = 0\}$. Definiendo la matriz $A = [a\ b]$, la recta se adapta a la forma general caracterizada por (1.1). De manera similar se demuestra que en $\mathbb{R}^3$ las rectas y planos que pasan por el origen son subespacios (los detalles quedan como ejercicio).

Ejemplo 1.12

Dado un número $n \in \mathbb{N}_0$, denominamos $\mathscr{P}_n$ al subconjunto de $\mathscr{P}$ compuesto por los polinomios de grado menor o igual a n junto con el polinomio nulo [a]:

$$\mathscr{P}_n = \{p : p(t) = a_n t^n + a_{n-1} t^{n-1} + \ldots + a_0, a_i \in \mathbb{R}, i = 0, \ldots, n\}$$

La demostración de que $\mathscr{P}_n$ es efectivamente un subespacio de $\mathscr{P}$ se deja como ejercicio.

[a] Un detalle técnico es que al polinomio nulo no se le asigna grado, de ahí que se lo mencione aparte.

Ejemplo 1.13

Dado un intervalo I de la recta real, consideremos el conjunto $\mathscr{C}^1(I)$ formado por todas las funciones $f : I \to \mathbb{R}$ derivables con continuidad. Evidentemente $\mathscr{C}^1(I)$ es un subconjunto de $\mathscr{C}(I)$, pues toda función derivable es continua. Afirmamos que $\mathscr{C}^1(I)$ es un subespacio de $\mathscr{C}(I)$. En efecto, la función nula es derivable y su derivada, que es la misma función nula, es continua. Esto prueba que $\mathscr{C}^1(I)$ no es vacío. Por otra parte, dadas $f, g \in \mathscr{C}^1(I)$, resulta que $f + g \in \mathscr{C}^1(I)$ porque $(f + g)' = f' + g'$ es continua por ser continuas f' y g'. Similarmente, si $f \in \mathscr{C}^1(I)$ y α es un número real cualquiera, $(\alpha f)' = \alpha f'$ es continua y, por lo tanto, $\alpha f \in \mathscr{C}^1(I)$.

De manera semejante se puede probar que $\mathscr{C}^k(I)$, el conjunto compuesto por las funciones definidas en I que son k veces derivables con continuidad, es subespacio de $\mathscr{C}(I)$.

1.3 Combinaciones lineales

Las operaciones de un espacio vectorial consisten en multiplicar vectores por escalares y en sumar vectores. Queda entonces bien definida cualquier expresión de la forma

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_r v_r$$

es decir, una colección finita de las operaciones permitidas en el espacio vectorial.

Definición 1.3

Dados $v_1, v_2, \ldots v_r$ vectores de un espacio vectorial V, una ***combinación lineal*** de ellos es cualquier expresión de la forma

$$\alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_r v_r$$

donde $\alpha_1, \alpha_2, \ldots, \alpha_r$ son escalares arbitrarios.

Por otra parte, se dice que un vector $v \in V$ es una ***combinación lineal de*** (o que ***depende linealmente de***) $v_1, v_2, \ldots v_r$ si existen escalares $\alpha_1, \alpha_2, \ldots, \alpha_r$ tales que $v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_r v_r$

Ejemplo 1.14

El vector nulo es combinación lineal de cualquier conjunto de vectores $v_1, v_2, \ldots v_r \in V$, pues

$$0_V = 0v_1 + 0v_2 + \ldots + 0v_r$$

Ejemplo 1.15

En $\mathbb{R}^3$, $w = (1,1,1)$ es combinación lineal de $v_1 = (2,1,3)$ y $v_2 = (1,0,2)$, pues

$$w = 1v_1 + (-1)v_2$$

Asimismo, $(2,2,2)$ es combinación lineal de $(1,1,1)$, ya que $(2,2,2) = 2(1,1,1)$

Ejercicio.

1.2 Considere en $\mathbb{C}^2$ los vectores $v_1 = (1,0)$, $v_2 = (0,1)$ y $w = (1+i,2)$. Demuestre que si se considera a $\mathbb{C}^2$ como $\mathbb{C}$ espacio vectorial, w es combinación lineal de v_1 y v_2, mientras que si se considera a $\mathbb{C}^2$ como $\mathbb{R}$ espacio vectorial, w **no** es combinación lineal de v_1 y v_2.

Ejercicio.

1.3 Demuestre que en $\mathscr{C}(\mathbb{R})$ el vector $f(x) = \operatorname{sen}\left(x + \frac{\pi}{4}\right)$ es combinación lineal de los vectores $f_1(x) = \cos x$ y $f_2(x) = \operatorname{sen} x$.

Definición 1.4

Dados los vectores $v_1, v_2, \ldots, v_r \in V$, se denomina $\operatorname{gen}\{v_1, v_2, \ldots, v_r\}$ al conjunto de todas sus posibles combinaciones lineales. Es decir

$$\operatorname{gen}\{v_1, v_2, \ldots, v_r\} = \{v \in V : v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_r v_r \wedge \alpha_1, \alpha_2, \ldots \alpha_r \in \mathbb{K}\}$$

N Es conveniente destacar la diferencia fundamental entre las notaciones $\{v_1, v_2, \ldots, v_r\}$ y $\operatorname{gen}\{v_1, v_2, \ldots, v_r\}$, que suelen confundirse. La primera alude a un conjunto *finito* de r vectores, mientras que la segunda alude al conjunto de las *infinitas* combinaciones lineales de dichos vectores.

Queda como ejercicio para el lector demostrar que $\operatorname{gen}\{v_1, v_2, \ldots, v_r\}$ es subespacio.

Definición 1.5

Al subespacio $S = \operatorname{gen}\{v_1, v_2, \ldots, v_r\}$ se lo denomina ***subespacio generado*** por los vectores $v_1, v_2, \ldots, v_r$, y a estos últimos ***generadores*** de S.

Se dice que $\{v_1, v_2, \ldots, v_r\}$ es un ***conjunto generador*** de S y que S es un subespacio ***finitamente generado***.

Ejemplo 1.16

En $\mathbb{R}^2$, $\text{gen}\{(1,1)\}$ es el conjunto de todos los múltiplos del vector $(1,1)$, el cual coincide con la recta de pendiente 1 que pasa por el origen.

Ejercicio. Demuestre que en $\mathscr{P}$, $\text{gen}\{1, t, t^2\} = \mathscr{P}_2$.

Ejemplo 1.17

Sea $S = \{x \in \mathbb{R}^3 : 2x_1 + x_2 + x_3 = 0\}$. S es subespacio de $\mathbb{R}^3$ (ver ejemplo 1.11). Además,

$$x \in S \Leftrightarrow 2x_1 + x_2 + x_3 = 0 \Leftrightarrow x_3 = -2x_1 - x_2$$

Entonces los $x \in S$ son de la forma $x = (x_1, x_2, -2x_1 - x_2) = x_1(1,0,-2) + x_2(0,1,-1)$ con $x_1, x_2 \in \mathbb{R}$.
En otras palabras, $S = \text{gen}\{v_1, v_2\}$ donde $v_1 = (1,0,-2)$ y $v_2 = (0,1,-1)$.

N En el último ejemplo, lo correcto es referirse a $\{(1,0,-2),(0,1,-1)\}$ como **un** conjunto generador de S, y no como **el** conjunto generador. Esto es así porque existen otros conjuntos generadores de S, tales como $\{(2,0,-4),(0,2,-2)\}$ y $\{(1,0,-2),(0,1,-1),(1,1,-3)\}$ (verifíquelo).

Veamos algunos ejemplos más.

Ejemplo 1.18

$\mathbb{R}^2 = \text{gen}\{e_1, e_2\}$ con $e_1 = (1,0)$ y $e_2 = (0,1)$. En efecto, si $x = (x_1, x_2)$ con $x_1, x_2 \in \mathbb{R}$, resulta

$$x = (x_1, 0) + (0, x_2) = x_1(1,0) + x_2(0,1) = x_1 e_1 + x_2 e_2$$

Con mayor generalidad, se prueba que $\mathbb{R}^n = \text{gen}\{e_1, e_2, \ldots, e_n\}$ siendo $e_1 = (1,0,\ldots,0), e_2 = (0,1,\ldots,0), \ldots, e_n = (0,0,\ldots,1)$.

Ejemplo 1.19

$\mathbb{C}^n = \text{gen}\{e_1, e_2, \ldots, e_n\}$ con $e_1, e_2, \ldots, e_n$ definidos como en el ejemplo anterior.

Ejemplo 1.20

$\mathscr{P}_n = \text{gen}\{1, t, \ldots, t^n\}$.

Ejemplo 1.21

$\mathbb{K}^{2\times 2} = \text{gen}\{E_{11}, E_{12}, E_{21}, E_{22}\}$ siendo

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

En efecto, para cualquier $A = [a_{ij}]$ tenemos la combinación lineal $A = a_{11}E_{11} + a_{12}E_{12} + a_{21}E_{21} + a_{22}E_{22}$.

En el caso general, $\mathbb{K}^{m\times n} = \text{gen}\{E_{11}, E_{12}, ..., E_{1n}, ..., E_{m1}, ..., E_{mn}\}$ donde E_{ij} es la matriz que tiene todos sus elementos nulos, salvo el correspondiente a la posición ij que es 1.

Ejemplo 1.22

$\mathscr{P}$ no es un espacio vectorial finitamente generado.

Para demostrarlo consideramos cualquier conjunto finito de polinomios no nulos $\{p_1, p_2, ..., p_N\}$ y llamamos g_i al grado del polinomio p_i. Sea entonces G el máximo entre los grados de los p_i, es decir

$$G = \text{máx}\{g_1, g_2, ..., g_N\}$$

Entonces cualquier combinación lineal de los polinomios $p_1, p_2, ..., p_N$ será un polinomio nulo o un polinomio de grado $\leq G$, con lo cual

$$\text{gen}\{p_1, p_2, ..., p_N\} \subsetneq \mathscr{P}$$

1.4 Independencia lineal y bases

Como ya señalamos en la nota del ejemplo 1.17, es posible que un subespacio finitamente generado tenga diferentes conjuntos generadores, incluso con diferentes cantidades de vectores. Por ejemplo cada uno de los conjuntos $\{(1,0),(0,1)\}$, $\{(1,1),(1,-1)\}$ y $\{(1,1),(1,-1),(1,0)\}$ genera $\mathbb{R}^2$. En el caso de los dos primeros conjuntos generadores, para cada vector de $\mathbb{R}^2$ hay una única combinación lineal adecuada que lo genera. En cambio, existen distintas combinaciones lineales de los elementos del tercero que generan un mismo vector, como se aprecia en la figura 1.3. (Dejamos la demostración formal de esta afirmación como ejercicio para el lector).

Usualmente será preferible trabajar con conjuntos de generadores tales que la combinación lineal que hay que efectuar para obtener cada vector del espacio es única.

Consideremos ahora el caso de un subespacio S generado por un solo vector v. Si $v \neq 0$, entonces combinaciones lineales diferentes darán origen a diferentes vectores. En efecto, usando la propiedad 4 del teorema 1.1, si $\alpha v = \beta v \Rightarrow \alpha = \beta$. De manera equivalente: $\alpha \neq \beta \Rightarrow \alpha v \neq \beta v$. Obviamente, si $v = 0$, distintas combinaciones lineales de v producen el mismo resultado; por ejemplo $1v = 2v = 0$.

La definición siguiente nos permitirá caracterizar a los conjuntos de generadores para los cuales cada vector del subespacio generado se alcanza con una única combinación lineal.

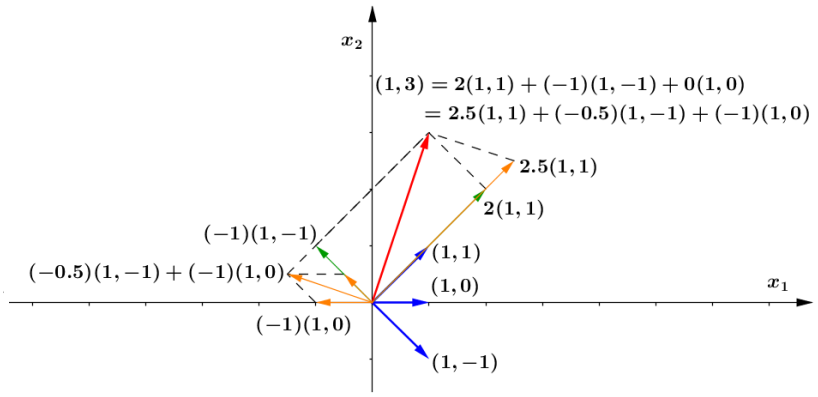


Figura 1.3: Dos combinaciones lineales que generan al mismo vector

Definición 1.6

Un conjunto de vectores $\{v_1, v_2, ..., v_r\}$ de un espacio vectorial V se dice ***linealmente independiente*** si la única solución de la ecuación

$$\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_r v_r = 0_V \tag{1.2}$$

es la solución trivial $\alpha_1 = \alpha_2 = ... = \alpha_r = 0$.

A su vez, $\{v_1, v_2, ..., v_r\}$ se dice ***linealmente dependiente*** si existe una solución no trivial de la ecuación (1.2); es decir si existen escalares $\alpha_1, \alpha_2, ..., \alpha_r$ no todos nulos para los cuales se verifica (1.2).

Ejemplo 1.23

Apliquemos la definición de independencia lineal a los conjuntos de vectores $\{(1,1),(1,-1)\}$ y $\{(1,1),(1,-1),(1,0)\}$ mencionados al comienzo de esta sección.

Al plantear la ecuación 1.2 para el primer conjunto tenemos

$$\alpha_1(1,1) + \alpha_2(1,-1) = (0,0)$$
$$(\alpha_1 + \alpha_2, \alpha_1 - \alpha_2) = (0,0)$$

lo que nos lleva al sistema de ecuaciones

$$\alpha_1 + \alpha_2 = 0$$
$$\alpha_1 - \alpha_2 = 0$$

cuya única solución es $\alpha_1 = \alpha_2 = 0$, por lo que afirmamos que el conjunto de vectores $\{(1,1),(1,-1)\}$ es linealmente independiente.

En cuanto al conjunto $\{(1,1),(1,-1),(1,0)\}$, el planteo de la ecuación 1.2 es

$$\alpha_1(1,1) + \alpha_2(1,-1) + \alpha_3(1,0) = (0,0)$$
$$(\alpha_1 + \alpha_2 + \alpha_3, \alpha_1 - \alpha_2) = (0,0)$$

lo que nos lleva al sistema de ecuaciones

$$\begin{aligned}\alpha_1+\alpha_2+\alpha_3&=0\\ \alpha_1-\alpha_2&=0\end{aligned}$$

cuyo conjunto solución es gen$\{(1,1,-2)\}$. Afirmamos entones que el conjunto $\{(1,1),(1,-1),(1,0)\}$ es linealmente dependiente.

Ejemplo 1.24

$\{1, t-1, t^2+2t+1\}$ es un conjunto linealmente independiente en $\mathscr{P}$. Para comprobarlo planteamos la ecuación (1.2) con $v_1=1, v_2=t-1, v_3=t^2+2t+1$:

$$\begin{aligned}\alpha_1 1+\alpha_2(t-1)+\alpha_3(t^2+2t+1)&=0\\ (\alpha_1-\alpha_2+\alpha_3)+(\alpha_2+2\alpha_3)\,t+\alpha_3 t^2&=0\end{aligned}$$

Como un polinomio es nulo si y sólo si todos sus coeficientes los son, necesariamente

$$\begin{aligned}\alpha_1-\alpha_2+\alpha_3&=0\\ \alpha_2+2\alpha_3&=0\\ \alpha_3&=0\end{aligned}$$

con lo cual, resolviendo el sistema obtenemos $\alpha_1=\alpha_2=\alpha_3=0$.

Ejercicio.

1.4 Demuestre que un conjunto $\{v_1, v_2, ..., v_r\}$ con $r \geq 2$ es linealmente dependiente si y sólo si hay un vector que es combinación lineal de los demás.

Ejemplo 1.25

$\{\operatorname{sen} x, \cos x, \operatorname{sen}\left(x+\frac{\pi}{4}\right)\}$ es linealmente dependiente en $\mathscr{C}(\mathbb{R})$ ya que, como mencionamos en el ejercicio 1.3, la función $\operatorname{sen}\left(x+\frac{\pi}{4}\right)$ es combinación lineal de las otras dos.

Ejemplo 1.26

$\{\operatorname{sen} x, \cos x\}$ es linealmente independiente en $\mathscr{C}(\mathbb{R})$. Para probarlo, supongamos que α_1 y α_2 son escalares tales que $\alpha_1 \operatorname{sen} x+\alpha_2 \cos x$ es la función nula, o sea que

$$\alpha_1 \operatorname{sen} x+\alpha_2 \cos x=0 \quad \forall x \in \mathbb{R} \tag{1.3}$$

Como suponemos que la igualdad vale para todo $x \in \mathbb{R}$, en particular vale para $x=0$, con lo cual

$$\alpha_1 \operatorname{sen} 0+\alpha_2 \cos 0=0 \Leftrightarrow \alpha_2=0$$

Considerando ahora $x = \frac{\pi}{2}$ y teniendo en cuenta que $\alpha_2 = 0$, resulta

$$\alpha_1 \operatorname{sen} \frac{\pi}{2} = 0 \Leftrightarrow \alpha_1 = 0$$

Con esto hemos demostrado que necesariamente $\alpha_1 = \alpha_2 = 0$, y con ello que el conjunto es linealmente independiente.

Una resolución alternativa consiste en derivar a la función nula $\alpha_1 \operatorname{sen} x + \alpha_2 \cos x$ y obtener así dos ecuaciones que deben cumplirse para todo $x \in \mathbb{R}$:

$$\alpha_1 \operatorname{sen} x + \alpha_2 \cos x = 0$$
$$\alpha_1 \cos x - \alpha_2 \operatorname{sen} x = 0$$

Se trata de un sistema homogéneo con incógnitas α_1, α_2. Podemos escribirlo matricialmente:

$$\begin{bmatrix} \operatorname{sen} x & \cos x \\ \cos x & -\operatorname{sen} x \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

Si para algún valor de x la matriz de coeficientes tiene determinante distinto de cero, el sistema es compatible determinado y la única solución posible es la trivial $\alpha_1 = \alpha_2 = 0$. En este ejemplo el determinante vale -1 para cualquier x, de donde se comprueba, una vez más, la independencia de las funciones.

La última parte del ejemplo anterior sugiere un método para probar la independencia lineal de un conjunto de funciones $\{f_1, f_2, .., f_n\}$ suficientemente derivables:

- Plantear que $\alpha_1 f_1(x) + \alpha_2 f_2(x) + ... + \alpha_n f_n(x)$ es la función nula:

$$\alpha_1 f_1(x) + \alpha_2 f_2(x) + ... + \alpha_n f_n(x) = 0 \ \ \forall x \in \mathbb{R}$$

- Como hay n incógnitas se deben obtener n ecuaciones, para lo cual se deriva $n-1$ veces a la función original:

$$\begin{aligned} \alpha_1 f_1(x) + \alpha_2 f_2(x) + ... + \alpha_n f_n(x) &= 0 \\ \alpha_1 f_1'(x) + \alpha_2 f_2'(x) + ... + \alpha_n f_n'(x) &= 0 \\ &\vdots \\ \alpha_1 f_1^{(n-1)}(x) + \alpha_2 f_2^{(n-1)}(x) + ... + \alpha_n f_n^{(n-1)}(x) &= 0 \end{aligned}$$

- La matriz de coeficientes del sistema depende de x: si para algún valor de x su determinante es no nulo, entonces el sistema es compatible determinado y es $\alpha_1 = \alpha_2 = ... = \alpha_n = 0$, de donde se deduce la independencia lineal.

Resumimos estos argumentos en una definición y un teorema.

Definición 1.7

Dado un conjunto de funciones $\{f_1, f_2, ..., f_n\}$ en $\mathscr{C}^{(n-1)}(I)$ su ***wronskiano*** es la función de $I \to \mathbb{R}$ dada por

$$W\left(f_1, f_2, ..., f_n\right)(x) = \det \begin{bmatrix} f_1(x) & f_2(x) & \cdots & f_n(x) \\ f_1'(x) & f_2'(x) & \cdots & f_n'(x) \\ \vdots & \vdots & \ddots & \vdots \\ f_1^{(n-1)}(x) & f_2^{(n-1)}(x) & \cdots & f_n^{(n-1)}(x) \end{bmatrix}$$

Teorema 1.3

Dado un conjunto de funciones $\{f_1, f_2, ..., f_n\}$ en $\mathscr{C}^{(n-1)}(I)$, si existe un $x_0 \in I$ tal que $W\left(f_1, f_2, ..., f_n\right)(x_0) \neq 0$, entonces $\{f_1, f_2, ..., f_n\}$ es linealmente independiente.

Ejercicio.

1.5 Demuestre que las funciones $f_1(x) = x$, $f_2(x) = e^x$ y $f_3(x) = e^{-x}$ son linealmente independientes en $\mathscr{C}^{(2)}(I)$ (donde I puede ser cualquier intervalo abierto).

Ejercicio.

1.6 Demuestre que si un conjunto de funciones $\{f_1, f_2, ..., f_n\}$ en $\mathscr{C}^{(n-1)}(I)$ es linealmente dependiente, entonces para todo $x \in I$ se verifica que $W\left(f_1, f_2, ..., f_n\right)(x) = 0$.

La definición 1.6 se puede reformular diciendo que un conjunto de vectores es linealmente independiente si hay una sola combinación lineal que da el vector nulo. Veremos en el teorema siguiente que con esto basta para caracterizar a los conjuntos de vectores con una única combinación lineal para cada elemento del espacio que generan.

Teorema 1.4

Sea $\{v_1, v_2, ..., v_r\}$ un conjunto generador de un subespacio S del espacio vectorial V. Entonces son equivalentes:

1. Para cada $v \in S$ existen escalares únicos $\alpha_1, \alpha_2, ..., \alpha_r$ tales que $v = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_r v_r$.
2. $\{v_1, v_2, ..., v_r\}$ es linealmente independiente.

Demostración:
Demostraremos en primer lugar que 1 ⇒ 2. Es decir, debemos probar que el conjunto $\{v_1, v_2, ..., v_r\}$ es linealmente independiente asumiendo que verifica 1. Pero esto es inmediato, porque al plantear la ecuación (1.2) se deduce que los escalares deben ser todos nulos.

A continuación, demostraremos la implicación 2⇒1. Suponemos entonces que para cierto $v \in V$ existen escalares $\alpha_1, \alpha_2, ..., \alpha_r$ y $\beta_1, \beta_2, ..., \beta_r$ tales que

$$v = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_r v_r = \beta_1 v_1 + \beta_2 v_2 + ... + \beta_r v_r$$

Entonces

$$(\alpha_1 - \beta_1)\, v_1 + (\alpha_2 - \beta_2)\, v_2 + \ldots + (\alpha_r - \beta_r)\, v_r = 0$$

y dada la independencia lineal de los vectores, debe ser para todo $i = 1, \ldots, r$:

$$(\alpha_i - \beta_i) = 0 \Rightarrow \alpha_i = \beta_i$$

□

El teorema anterior garantiza que un conjunto generador linealmente independiente tendrá una única combinación lineal posible para alcanzar cada vector del espacio generado. Esta clase de conjuntos será de gran interés a lo largo del curso, lo cual motiva la siguiente definición.

Definición 1.8

Un conjunto de vectores $B = \{v_1, v_2, \ldots, v_r\}$ del espacio vectorial V es una ***base*** de un subespacio S si y sólo si es un conjunto linealmente independiente que genera S.

N El espacio nulo carece de base, pues el conjunto $\{0_V\}$ es linealmente dependiente.

Planteamos ahora la siguiente cuestión: dado un conjunto generador de un subespacio $S \neq \{0_V\}$ ¿cómo podemos obtener una base de S? La respuesta la brinda el siguiente lema.

Lema 1.1

Sea $\{v_1, v_2, \ldots v_r, v_{r+1}\}$ un conjunto generador del subespacio S. Supongamos que v_{r+1} depende linealmente del resto, entonces $\{v_1, v_2, \ldots v_r\}$ también es un conjunto generador de S.

Demostración:
Dado un vector $v \in S$, existen escalares $\alpha_1, \alpha_2, \ldots \alpha_r, \alpha_{r+1}$ tales que

$$v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_r v_r + \alpha_{r+1} v_{r+1}$$

Por otra parte, hemos supuesto que v_{r+1} es una combinación lineal de los demás vectores. Es decir que existen escalares $\beta_1, \beta_2, \ldots, \beta_r$ tales que

$$v_{r+1} = \beta_1 v_1 + \beta_2 v_2 + \ldots + \beta_r v_r$$

Reemplazando en la ecuación anterior tenemos

$$\begin{aligned} v &= \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_r v_r + \alpha_{r+1} (\beta_1 v_1 + \beta_2 v_2 + \ldots + \beta_r v_r) \\ &= \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_r v_r + (\alpha_{r+1}\beta_1 v_1 + \alpha_{r+1}\beta_2 v_2 + \ldots \alpha_{r+1}\beta_r v_r) \\ &= (\alpha_1 + \alpha_{r+1}\beta_1)\, v_1 + (\alpha_2 + \alpha_{r+1}\beta_2)\, v_2 + \ldots + (\alpha_r + \alpha_{r+1}\beta_r)\, v_r \end{aligned}$$

y por lo tanto existe una combinación lineal de los vectores $v_1, v_2, \ldots, v_r$ que genera a $v \in S$. □

El lema anterior nos indica *un método para obtener una base a partir de un conjunto generador*. Supongamos que $\{v_1, v_2, \ldots, v_r\}$ es un conjunto generador de un subespacio $S \neq \{0_V\}$.

Si $r = 1$, es decir si el conjunto generador posee un único elemento v_1, teniendo en cuenta que $S \neq \{0_V\}$ debe ser $v_1 \neq 0_V$. Por lo tanto v_1 es base de S.
Supongamos ahora que $r > 1$. Si $\{v_1, v_2, ..., v_r\}$ es linealmente independiente, entonces es base de S. Si en cambio es linealmente dependiente, alguno de sus elementos depende linealmente del resto; eliminamos ese elemento del conjunto y obtenemos un conjunto que sigue generando S pero con $r-1$ elementos en lugar de r. Si el conjunto obtenido es linealmente independiente, es base de S; si es linealmente dependiente, algún elemento es combinación lineal del resto. Entonces eliminando ese elemento del conjunto obtenemos un conjunto de generadores de S con $r-2$ elementos. Prosiguiendo de esta manera, después de un número finito de pasos (no más de $r-1$) obtendremos un conjunto de generadores de S linealmente independiente: una base de S. En el más extremo de los casos llegaremos a un conjunto generador de S compuesto por un solo vector, que forzosamente deberá ser no nulo, pues $S \neq \{0_V\}$, y que por ende será base de S.

El argumento anterior demuestra el siguiente resultado.

Teorema 1.5

Sea $G = \{v_1, v_2, ..., v_r\}$ un conjunto generador de un subespacio $S \neq \{0_V\}$ de cierto espacio vectorial V. Entonces existe un subconjunto de G que es base de S.

Una consecuencia inmediata de este teorema es la siguiente.

Teorema 1.6

Todo subespacio finitamente generado y distinto del nulo posee una base.

En lo que sigue vamos a enunciar una serie de propiedades útiles e interesantes que se deducen de la existencia de bases [1].

Lema 1.2

Si $\{v_1, v_2, ..., v_r\}$ es una base de un subespacio S de un espacio vectorial V, entonces todo conjunto de vectores de S con más de r elementos es linealmente dependiente.

Una consecuencia del lema anterior es el resultado siguiente.

Teorema 1.7

Sea S un subespacio no nulo finitamente generado. Entonces todas las bases de S tienen el mismo número de elementos.

Este teorema garantiza que exista un número de elementos común a todas las bases y justifica la definición que sigue.

Definición 1.9

La ***dimensión*** de un subespacio no nulo finitamente generado es el número de elementos que posee una base cualquiera de S. Se define también que el subespacio nulo tiene dimensión cero. A la dimensión de S se la denota $\dim(S)$.

[1] Hemos optado por omitir las demostraciones, aunque el lector interesado puede encontrar los detalles en la bibliografía. Por ejemplo, en [4] y [5].

Ejemplo 1.27 Base canónica de $\mathbb{K}^n$

Tanto la dimensión de $\mathbb{R}^n$ como la de $\mathbb{C}^n$ ($\mathbb{C}$ espacio vectorial) es n. En efecto, el conjunto $E_n = \{e_1, e_2, ..., e_n\}$ ya estudiado en los ejemplos 1.18 y 1.19 es una base tanto de $\mathbb{R}^n$ como de $\mathbb{C}^n$. Ya vimos que E_n genera cada uno de esos espacios; por otra parte la independencia lineal es muy sencilla de probar. A la base E_n se la suele denominar ***base canónica***.

Ejemplo 1.28

La dimensión de $\mathbb{C}^n$ como $\mathbb{R}$ espacio vectorial no es n sino $2n$.

Veámoslo para el caso $n = 2$ (el caso general es análogo). En primer lugar, notemos que si bien $\{e_1, e_2\}$ es un conjunto linealmente independiente en $\mathbb{C}^2$, ya no genera todo el espacio. Por ejemplo, el vector $(2 + i, 1)$ no puede escribirse como combinación lineal de e_1 y e_2 si los escalares de la combinación son números reales.

Veamos ahora que la dimensión es 4. Para ello exhibiremos cuatro vectores linelamente independientes que generen a $\mathbb{C}^2$. Consideramos

$$(1,0) \quad (i,0) \quad (0,1) \quad (0,i) \tag{1.4}$$

En primer lugar comprobamos que efectivamente generan a $\mathbb{C}^2$: sea $(z_1, z_2) \in \mathbb{C}^2$; entonces $z_1 = x_1 + i y_1$ y $z_2 = x_2 + i y_2$ con $x_1, x_2, y_1, y_2 \in \mathbb{R}$. Luego

$$\begin{aligned}(z_1, z_2) &= (x_1 + i y_1, x_2 + i y_2) \\ &= x_1(1,0) + y_1(i,0) + x_2(0,1) + y_2(0,i)\end{aligned}$$

Luego cualquier vector de $\mathbb{C}^2$ es una combinación lineal (con escalares reales) de los cuatro vectores de (1.4).
Para probar que son linealmente independientes, planteamos la ecuación

$$\begin{aligned}\alpha_1(1,0) + \alpha_2(i,0) + \alpha_3(0,1) + \alpha_4(0,i) &= (0,0) \\ \Leftrightarrow (\alpha_1 + i\alpha_2, \alpha_3 + i\alpha_4) &= (0,0)\end{aligned}$$

Con lo cual es necesariamente

$$\alpha_1 + i\alpha_2 = 0 \wedge \alpha_3 + i\alpha_4 = 0$$

Como se trata de escalares reales, la única solución posible es $\alpha_1 = \alpha_2 = \alpha_3 = \alpha_4 = 0$. Por lo tanto, los cuatro vectores de (1.4) son linealmente independientes.

Ejemplo 1.29 Base canónica de $\mathbb{K}^{m\times n}$

Las dimensiones de $\mathbb{R}^{m\times n}$ y de $\mathbb{C}^{m\times n}$ son ambas iguales a $m \cdot n$. En efecto, como ya mostramos en el ejemplo 1.21, ambos espacios están generados por el conjunto $E_{m\times n} = \{E_{11}, E_{12}, ..., E_{1n}, ..., E_{m1}, ..., E_{mn}\}$ definido oportunamente y cuya independencia lineal se prueba fácilmente. Este conjunto es la ***base canónica*** tanto de $\mathbb{R}^{m\times n}$ como de $\mathbb{C}^{m\times n}$.

Ejemplo 1.30 Base canónica de $\mathscr{P}_n$

$\{1, t, ..., t^n\}$ es base de $\mathscr{P}_n$ (verifíquelo) y por lo tanto $\dim(\mathscr{P}_n) = n + 1$. Dicha base es la ***base canónica*** del espacio $\mathscr{P}_n$.

Ejemplo 1.31

Consideremos el conjunto $V = \{f : [0,1] \to \mathbb{R} : f(x) = (a_0 + a_1 x + a_2 x^2) e^{-x}\}$. Este conjunto de funciones resulta un espacio vectorial si se consideran las operaciones suma y producto usuales para funciones.
Más aun, $V = \text{gen}\{e^{-x}, xe^{-x}, x^2 e^{-x}\}$. Afirmamos que $\dim(V) = 3$.

Para constatarlo, basta con ver que el conjunto generador $\{e^{-x}, xe^{-x}, x^2 e^{-x}\}$ es linealmente independiente. Usaremos el mismo argumento que en el ejemplo 1.26: supongamos entonces que existen escalares $\alpha_1, \alpha_2, \alpha_3$ tales que

$$\alpha_1 e^{-x} + \alpha_2 x e^{-x} + \alpha_3 x^2 e^{-x} = 0 \ \forall x \in [0,1] \tag{1.5}$$

Evaluando en $x = 0$, $x = 0{,}5$ y $x = 1$, obtenemos el sistema de ecuaciones

$$\begin{aligned} \alpha_1 &= 0 \\ \alpha_1 e^{-0,5} + \alpha_2 0{,}5 e^{-0,5} + \alpha_3 0{,}5^2 e^{-0,5} &= 0 \\ \alpha_1 e^{-1} + \alpha_2 e^{-1} + \alpha_3 e^{-1} &= 0 \end{aligned}$$

que tiene por única solución $\alpha_1 = \alpha_2 = \alpha_3 = 0$.

Otra forma de demostrar la independencia lineal del conjunto generador es la siguiente: como $e^{-x} \neq 0$ para todo $x \in \mathbb{R}$, la ecuación (1.5) es equivalente a

$$\alpha_1 + \alpha_2 x + \alpha_3 x^2 = 0 \ \ \forall x \in [0,1] \tag{1.6}$$

Para que se cumpla (1.6) es necesario que sean $\alpha_1 = \alpha_2 = \alpha_3 = 0$ ya que $\alpha_1 + \alpha_2 x + \alpha_3 x^2$ debe ser el polinomio nulo.

Supongamos que la dimensión de un subespacio S es $\dim(S) = n$. Para demostrar que un conjunto de vectores $\{v_1, v_2, ..., v_r\} \subset S$ es base de S, según la definición de base debemos probar por una parte que es linealmente independiente y por otra que genera a S. Sin embargo, debido al siguiente resultado, la tarea es más sencilla: basta con probar que el conjunto verifica solamente una de las dos condiciones.

Teorema 1.8

Sea S un subespacio de un espacio vectorial V. Supongamos que $\dim(S) = n$. Luego

1. Si $\{v_1, v_2, ..., v_n\} \subset S$ es linealmente independiente, entonces $\{v_1, v_2, ..., v_n\}$ es base de S.
2. Si $\{v_1, v_2, ..., v_n\} \subset S$ genera a S, entonces $\{v_1, v_2, ..., v_n\}$ es base de S.

Ejemplo 1.32

$\{(1,1),(2,1)\}$ es base de $\mathbb{R}^2$ pues es linealmente independiente y $\dim(\mathbb{R}^2) = 2$.

Ejercicio.

1.7 Demuestre que $\{1 - t, 1 + t, t^2\}$ es una base de $\mathscr{P}_2$.

Para finalizar esta sección, no podemos dejar de mencionar un resultado motivado por la siguiente pregunta: dado un conjunto linealmente independiente $\{v_1, v_2, ..., v_r\}$ de vectores de un espacio vectorial V finitamente generado ¿forman parte de

alguna base de V? Dicho de otra manera: ¿se puede extender un conjunto linealmente independiente hasta completar una base?

Teorema 1.9

Sea V un espacio vectorial de dimensión n y sea $\{v_1, v_2, ..., v_r\} \subset V$ con $r < n$ un conjunto linealmente independiente. Entonces existen $v_{r+1}, ..., v_n \in V$ tales que $B = \{v_1, v_2, ..., v_r, v_{r+1}, ..., v_n\}$ es una base de V.

1.5 Subespacios fundamentales de una matriz

En la sección 1.1 identificamos los vectores de $\mathbb{K}^r$ con los de $\mathbb{K}^{r\times 1}$. A partir de esta sección y en lo que resta del texto, usaremos libremente esta identificación. Por lo general, será muy ventajoso y facilitará la notación interpretar a los vectores de $\mathbb{K}^r$ como "vectores columna" de $\mathbb{K}^{r\times 1}$.

Dada una matriz $A \in \mathbb{K}^{m\times n}$, cada una de sus columnas puede ser interpretada como un vector de $\mathbb{K}^m$. Entonces, si denominamos A_i a la i-ésima columna de A, tenemos que $A_i \in \mathbb{K}^m$ y que

$$A = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix}$$

Por ejemplo, si

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \tag{1.7}$$

sus columnas son

$$A_1 = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix}, \; A_2 = \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

Similarmente, cada fila de A puede ser interpretada como un vector columna de $\mathbb{K}^n$ *transpuesto* y, si denominamos F_i al vector columna que se obtiene transponiendo la i-ésima fila de A, podemos escribir

$$A = \begin{bmatrix} F_1^t \\ \vdots \\ F_m^t \end{bmatrix}$$

Por ejemplo, las filas transpuestas de la matriz A de (1.7) son

$$F_1 = \begin{bmatrix} 1 \\ 2 \end{bmatrix}, \; F_2 = \begin{bmatrix} 1 \\ 1 \end{bmatrix}, \; F_3 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

Con esta notación, el producto de la matriz $A \in \mathbb{K}^{m\times n}$ por un vector $x \in \mathbb{K}^n$ puede escribirse en la forma

$$Ax = \begin{bmatrix} F_1^t \\ \vdots \\ F_m^t \end{bmatrix} x = \begin{bmatrix} F_1^t x \\ \vdots \\ F_m^t x \end{bmatrix}$$

que no expresa otra cosa que la regla usual de multiplicación de una matriz por un vector: la i-ésima componente del producto es el producto de la i-ésima fila de A por el vector x.

En cada matriz podemos definir subespacios a partir de lo que generan sus filas y columnas.

Definición 1.10

Dada una matriz $A \in \mathbb{K}^{m\times n}$

- El ***espacio columna*** es el subespacio de $\mathbb{K}^m$ generado por las columnas de A y se denota Col A:

$$\text{Col } A = \text{gen}\{A_1, ..., A_n\} \subseteq \mathbb{K}^m \tag{1.8}$$

- El ***espacio fila*** es el subespacio de $\mathbb{K}^n$ generado por las filas transpuestas de A y se denota Fil A:

$$\text{Fil } A = \text{gen}\{F_1, ..., F_m\} \subseteq \mathbb{K}^n \tag{1.9}$$

Dado que transponiendo las filas de A obtenemos las columnas de A^t, tenemos que el espacio fila de A no es otro que el espacio columna de A^t, es decir

Proposición 1.1

Para toda matriz $A \in \mathbb{K}^{m\times n}$ se verifica que

$$\text{Fil } A = \text{Col}\left(A^t\right)$$

Vamos a dar ahora una interpretación del producto entre una matriz y un vector que usaremos a menudo. Comenzamos con un ejemplo: sean A la matriz definida en (1.7) y $x = [1\ 2]^t$, aplicamos la definición usual de producto de matrices y luego reescribimos el resultado

$$Ax = \begin{bmatrix} 1 & 2 \\ 1 & 1 \\ 1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} 1\cdot 1 + 2\cdot(-2) \\ 1\cdot 1 + 1\cdot(-2) \\ 1\cdot 1 + 0\cdot(-2) \end{bmatrix} = 1 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} + (-2) \begin{bmatrix} 2 \\ 1 \\ 0 \end{bmatrix}$$

Esto es un ejemplo de que *el producto de una matriz por un vector columna consiste en combinar linealmente las columnas de la matriz empleando como escalares las componentes del vector*. No es difícil demostrar el caso general.

Proposición 1.2

Dada una matriz $A \in \mathbb{K}^{m\times n}$ y dado un vector $x = [x_1\ x_2\ ...x_n]^t \in \mathbb{K}^n$, el producto se puede expresar como la siguiente combinación lineal de las columnas de la matriz

$$Ax = \begin{bmatrix} A_1 & A_2 & \cdots & A_n \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} = x_1 A_1 + x_2 A_2 + ... + x_n A_n$$

Cada vez que multiplicamos a una matriz $A \in \mathbb{K}^{m\times n}$ por un vector $x \in \mathbb{K}^n$ obtenemos un elemento de Col A y, recíprocamente, todo elemento de Col A puede escribirse en la forma Ax eligiendo un $x \in \mathbb{K}^n$ con los escalares adecuados. Luego vale la siguiente afirmación.

Proposición 1.3

Para toda matriz $A \in \mathbb{K}^{m\times n}$ se verifica que

$$\text{Col}\, A = \{Ax : x \in \mathbb{K}^n\}$$

En lo que sigue, vamos a vincular los subespacios Col A y Fil A con el rango de A, al que denotaremos rango A. Recordando que el rango de una matriz es tanto el número máximo de columnas linealmente independientes como el número máximo de filas linealmente independientes, se llega sin dificultad a la siguiente afirmación.

Proposición 1.4

Para toda matriz $A \in \mathbb{K}^{m\times n}$ se verifica que

$$\text{rango}\, A = \dim(\text{Col}\, A) = \dim(\text{Fil}\, A)$$

A continuación vincularemos a cada matriz con el sistema de ecuaciones lineales que se le asocia naturalmente. Es decir, dada $A \in \mathbb{K}^{m\times n}$, consideraremos la ecuación matricial

$$Ax = b \tag{1.10}$$

donde $b \in \mathbb{K}^m$ y $x \in \mathbb{K}^n$ es la incógnita. En el caso de que b sea el vector nulo, la ecuación (1.10) resulta $Ax = 0$ y se tiene un sistema homogéneo . El conjunto solución de dicho sistema es un subespacio de $\mathbb{K}^n$ (¡ejercicio!) y le asignamos un nombre propio:

Definición 1.11

Dada una matriz $A \in \mathbb{K}^{m\times n}$, el ***espacio nulo*** es el conjunto solución del sistema homogéneo asociado a la matriz y se denota Nul A. Es decir

$$\text{Nul}\, A = \{x \in \mathbb{K}^n : Ax = 0\}$$

Teniendo en cuenta el trabajo hecho en esta sección, podemos dar una nueva interpretación de lo que significa que un sistema de ecuaciones sea o no compatible. Volviendo a mirar la ecuación (1.10), notamos ahora que el sistema es compatible si y sólo si existe una combinación lineal de las columnas de A que resulte ser el vector b. Vale decir que el sistema será compatible si $b \in \text{Col}\, A$. Más aún, si $\text{Col}\, A = \mathbb{K}^m$ entonces el sistema será compatible para cualquier b.

Con esto, queda demostrado el siguiente teorema.

Teorema 1.10

Dada $A \in \mathbb{K}^{m\times n}$ son equivalentes

- El sistema $Ax = b$ tiene solución para todo $b \in \mathbb{K}^m$.
- $\text{Col}\, A = \mathbb{K}^m$
- $\text{rango}\, A = m$

En cuanto a la unicidad de las soluciones, cada sistema $Ax = b$ (con b arbitrario) tiene a lo sumo una solución si y sólo si la ecuación homogénea tiene solución única (trivial), esto es, $\text{Nul}\, A = \{0\}$. [2] Ahora bien, que $\text{Nul}\, A = \{0\}$ significa que la ecuación

[2] Suponiendo que $\text{Nul}\, A = \{0\}$, basta con señalar que si existen dos soluciones x e y, es decir si $Ax = Ay = b$, entonces $A(x - y) = 0$ y se deduce que $x = y$. Recíprocamente, si cada sistema $Ax = b$ tiene a lo sumo una solución, vale en particular para $b = 0$ y en tal caso la única solución es $x = 0$.

$Ax = 0$ tiene solamente solución trivial y esto equivale a que las columnas de A sean linealmente independientes (¿por qué?). En consecuencia, la solución de (1.10), en caso de existir, es única si y sólo si las columnas de A son linealmente independientes. Esto último equivale a que sea $\dim(\text{Col}\, A) = \text{rango}\, A = n$.

Reunimos las observaciones anteriores en el teorema que sigue.

Teorema 1.11

Dada $A \in \mathbb{K}^{m\times n}$ son equivalentes

- El sistema $Ax = b$ tiene a lo sumo una solución (una o ninguna) para cada $b \in \mathbb{K}^m$.
- $\text{Nul}\, A = \{0\}$
- $\dim(\text{Col}\, A) = n$
- $\text{rango}\, A = n$

Definición 1.12

A los cuatro subespacios que hemos asociado a cada matriz A: $\text{Col}\, A, \text{Fil}\, A, \text{Nul}\, A, \text{Nul}(A^t)$ se los denomina ***subespacios fundamentales*** de A.

1.6 Coordenadas

En el teorema 1.4 quedó establecido que, en los conjuntos linealmente independientes, hay una única combinación lineal adecuada para cada vector del espacio que generan. Esa combinación lineal queda caracterizada por los escalares que se usan: en esta sección estableceremos una correspondencia entre dichos escalares y el vector que producen. Para ello, comenzamos por establecer un orden en cada base del espacio.

Definición 1.13

Una ***base ordenada*** de un espacio vectorial V de dimensión finita es una base en la cual se ha establecido un orden. En general emplearemos la notación

$$B = \{v_1; v_2; \ldots; v_n\}$$

para enfatizar que la base está ordenada. Es decir, separaremos los vectores con punto y coma.

Consideremos una base ordenada $B = \{v_1; v_2; \ldots; v_n\}$[3] de cierto espacio V. Como ya mencionamos, para cada vector $v \in V$ existe y es único el conjunto de escalares $\alpha_1, \alpha_2, \ldots, \alpha_n$ (en ese orden) tales que $v = \alpha_1 v_1 + \alpha_2 v_2 + \ldots + \alpha_n v_n$. Esto nos permite establecer la correspondencia entre vectores y escalares que habíamos anticipado.

[3] Hacemos esta distinción con punto y coma porque en la notación de conjuntos $B = \{v_1, v_2, \ldots, v_n\}$ alude a un conjunto que contiene a los elementos $v_1, v_2, \ldots, v_n$ sin tener, en principio, ningún orden. Es decir, daría lo mismo escribir $B = \{v_1, v_2, \ldots, v_n\} = \{v_n, v_2, \ldots, v_1\}$.

Definición 1.14

Dados un espacio vectorial V y una base ordenada $B = \{v_1; v_2; ...; v_n\}$, para cada vector $v \in V$ definimos el ***vector de coordenadas*** de v respecto de la base B, al que denotamos $[v]_B$, mediante

$$[v]_B = \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ \alpha_n \end{bmatrix} \in \mathbb{K}^n$$

donde $\alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{K}$ son los únicos escalares tales que

$$v = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n$$

Por otra parte, si $\beta = \begin{bmatrix} \beta_1 & \beta_2 & ... & \beta_n \end{bmatrix}^t \in \mathbb{K}^n$, entonces existe un vector $w \in V$ tal que $[w]_B = \beta$. Efectivamente, tal vector es $w = \beta_1 v_1 + \beta_2 v_2 + ... \beta_n v_n$. De acuerdo con lo que acabamos de decir, al fijar una base ordenada B de V *podemos establecer una correspondencia biyectiva entre los vectores de* V *y los vectores de* $\mathbb{K}^n$. Tal correspondencia consiste en asignar a cada vector de V su vector de coordenadas en la base B.

Ejemplo 1.33

Consideramos la base ordenada $B = \{t-1; t+1; t^2+1\}$ de $\mathscr{P}_2$ (verifíquelo).

En primer lugar, buscamos las coordenadas del polinomio $p = t^2 - t + 1$ en la base B. Para ello es necesario encontrar los escalares que verifican

$$p = \alpha_1(t-1) + \alpha_2(t+1) + \alpha_3\left(t^2+1\right)$$

Operando en el segundo miembro de la igualdad llegamos a la ecuación

$$t^2 - t + 1 = \alpha_3 t^2 + (\alpha_1 + \alpha_2)\, t + (-\alpha_1 + \alpha_2 + \alpha_3)$$

y por lo tanto debe ser

$$\begin{aligned} \alpha_3 &= 1 \\ \alpha_1 + \alpha_2 &= -1 \\ -\alpha_1 + \alpha_2 + \alpha_3 &= 1 \end{aligned}$$

Resolviendo el sistema de ecuaciones obtenemos $\alpha_1 = -0{,}5$, $\alpha_2 = -0{,}5$ y $\alpha_3 = 1$, con lo cual $[p]_B = [-0{,}5 \ -0{,}5 \ 1]^t$.

Para ilustrar el camino inverso, buscamos un polinomio $q \in \mathscr{P}_2$ tal que $[q]_B = [0{,}5 \ -0{,}5 \ 2]^t$. Para ello calculamos

$$q = 0{,}5(t-1) - 0{,}5(t+1) + 2\left(t^2+1\right) = 2t^2 + 1$$

A continuación compararemos las coordenadas de la suma de vectores con la suma de coordenadas. La suma de las coordenadas de q con las de p es

$$[q]_B + [p]_B = [0 \ -1 \ 3]^t$$

mientras que las coordenadas de la suma $q + p$ surgen de

$$\begin{aligned} q + p &= \left(0{,}5(t-1) - 0{,}5(t+1) + 2\left(t^2+1\right)\right) + \left(-0{,}5(t-1) - 0{,}5(t+1) + 1\left(t^2+1\right)\right) \\ &= 0(t-1) - 1(t+1) + 3\left(t^2+1\right) \\ \Rightarrow [q+p]_B &= [0 \ -1 \ 3]^t \end{aligned}$$

con lo cual

$$[q]_B + [p]_B = [q+p]_B \tag{1.11}$$

Por último, compararemos la multiplicación de las coordenadas por un escalar con las coordenadas del vector multiplicado por un escalar. Consideramos el vector p y el escalar 2:

$$2[p]_B = [-1\ -1\ 2]^t$$

A su vez

$$\begin{aligned} 2p &= 2\left(-0{,}5(t-1) - 0{,}5(t+1) + 1\left(t^2+1\right)\right) \\ &= -1(t-1) - 1(t+1) + 2\left(t^2+1\right) \\ \Rightarrow [2p]_B &= [-1\ -1\ 2]^t \end{aligned}$$

Entonces tenemos que

$$2[p]_B = [2p]_B \tag{1.12}$$

Las fórmulas (1.11) y (1.12) del ejemplo sugieren la validez del teorema que sigue.

Teorema 1.12

Dados un espacio vectorial V y una base ordenada $B = \{v_1; v_2; ...; v_n\}$, para dos vectores cualesquiera $u, v \in V$ y cualquier escalar $k \in \mathbb{K}$ se verifica que

- $[u+v]_B = [u]_B + [v]_B$
- $[ku]_B = k[u]_B$

Demostración:
Sean $[u]_B = [\alpha_1\ \alpha_2\ ...\ \alpha_n]$ y $[v]_B = [\beta_1\ \beta_2\ ...\ \beta_n]$. Vale decir que

$$\begin{aligned} u &= \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n \\ v &= \beta_1 v_1 + \beta_2 v_2 + ... + \beta_n v_n \\ \Rightarrow u + v &= (\alpha_1 + \beta_1) v_1 + (\alpha_2 + \beta_2) v_2 + ... + (\alpha_n + \beta_n) v_n \end{aligned}$$

de donde se deduce que

$$\begin{aligned} [u+v]_B &= [\alpha_1 + \beta_1\ \alpha_2 + \beta_2\ ...\ \alpha_n + \beta_n]^t \\ &= [\alpha_1\ \alpha_2\ ...\ \alpha_n +]^t + [\beta_1\ \beta_2\ ...\ \beta_n]^t \\ &= [u]_B + [v]_B \end{aligned}$$

Por otra parte,

$$\begin{aligned} ku &= k(\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n) \\ &= (k\alpha_1) v_1 + (k\alpha_2) v_2 + ... + (k\alpha_n) v_n \end{aligned}$$

con lo cual

$$\begin{aligned} [ku]_B &= [k\alpha_1\ k\alpha_2\ ...\ k\alpha_n]^t \\ &= k[\alpha_1\ \alpha_2\ ...\ \alpha_n]^t \\ &= k[u]_B \end{aligned}$$

□

Como hemos visto recién, al fijar una base ordenada B de V, podemos reducir todas las operaciones entre vectores de V a operaciones entre sus coordenadas como vectores de $\mathbb{K}^n$. Más aún, como veremos a continuación, también es posible determinar la dependencia o independencia lineal de un conjunto de vectores de V a partir de la dependencia o independencia lineal de sus coordenadas.

Teorema 1.13

Dados un espacio vectorial V, una base ordenada $B = \{v_1; v_2; ...; v_n\}$ y un conjunto $\{u_1, u_2, ..., u_r\}$ de vectores de V, entonces son equivalentes

- $\{u_1, u_2, ..., u_r\}$ es linealmente independiente en V.
- $\{[u_1]_B, [u_2]_B, ..., [u_r]_B\}$ es linealmente independiente en $\mathbb{K}^n$.

Demostración:
Aplicando el teorema 1.12 y el hecho de que el vector nulo tiene coordenadas nulas, tenemos

$$\alpha_1 [u_1]_B + \alpha_2 [u_2]_B + ... + \alpha_r [u_r]_B = 0 \Leftrightarrow [\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_r u_r]_B = 0 \Leftrightarrow \alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_r u_r = 0$$

Vale decir que las soluciones de las ecuación en coordenadas son las mismas que las soluciones de la ecuación con los vectores originales. Por lo tanto, la ecuación $\alpha_1 [u_1]_B + \alpha_2 [u_2]_B + ... + \alpha_r [u_r]_B = 0$ tiene como única solución $\alpha_1 = \alpha_2 = ... = \alpha_r = 0$ si y sólo si la única solución de la ecuación $\alpha_1 u_1 + \alpha_2 u_2 + ... + \alpha_r u_r = 0$ es $\alpha_1 = \alpha_2 = ... = \alpha_r = 0$. En otras palabras, $\{[u_1]_B, [u_2]_B, ..., [u_r]_B\}$ es linealmente independiente en $\mathbb{K}^n$ si y sólo si $\{u_1, u_2, ..., u_r\}$ es linealmente independiente en V. □

Ejemplo 1.34

Determinar si $B = \{1 + t; 1 + t - 3t^2; -1 + 2t - 2t^2\}$ es base de $\mathscr{P}_2$.

En primer lugar, sabemos que la dimensión de $\mathscr{P}_2$ es 3. Luego (aplicando el teorema 1.8) bastará con estudiar su independencia lineal. Ahora bien, aplicando el teorema 1.13, podemos estudiar las coordenadas de los vectores con respecto a la base canónica $E = \{1, t, t^2\}$:

$$[1 + t]_E = [1\ 1\ 0]^t \quad \left[1 + t - 3t^2\right]_E = [1\ 1\ -3]^t \quad \left[-1 + 2t - 2t^2\right]_E = [-1\ 2\ -2]^t$$

Para estudiar la independencia lineal de estos vectores de $\mathbb{R}^3$ disponemos de varios métodos. Por ejemplo, podemos formar una matriz de tamaño 3×3 disponiendo cada uno de esos vectores como fila y determinar su rango: si éste es igual al número de filas de la matriz, los vectores son linealmente independientes. En nuestro caso, la matriz es

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

Para calcular su rango aplicamos eliminación gaussiana (triangulación):

$$\begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & -3 \\ -1 & 2 & -2 \end{bmatrix} \begin{matrix} F_2 - F_1 \\ \rightarrow \\ F_3 + F_1 \end{matrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & -3 \\ 0 & 3 & -2 \end{bmatrix} \begin{matrix} F_2 \leftrightarrow F_3 \\ \rightarrow \end{matrix} \begin{bmatrix} 1 & 1 & 0 \\ 0 & 3 & -2 \\ 0 & 0 & -3 \end{bmatrix}$$

Como el rango de la última matriz es 3, el rango de la matriz inicial es también 3 ya que, como es sabido, en cada paso de triangulación cambia la matriz pero no el espacio fila. Por lo tanto los vectores coordenados son linealmente

independientes y B es una base de $\mathscr{P}_2$.

Una alternativa es la siguiente: como A es una matriz cuadrada y solamente nos interesa saber si su rango es 3, calculamos su determinante; si resulta distinto de cero entonces rango $A = 3$, en caso contrario es rango $A < 3$. Dejamos como ejercicio completar esta alternativa.

1.7 Matriz de Cambio de Base

Como hemos visto, dada una base ordenada B de un espacio vectorial V de dimensión n, podemos identificar cada vector de V con su vector de coordenadas en la base B. Esto nos permite operar con las coordenadas en $\mathbb{K}^n$, lo cual suele ser más práctico (como se ilustró en el ejemplo 1.34). En algunas aplicaciones es posible que comencemos describiendo un problema usando las coordenadas de los vectores en una base B, pero que su solución se vea facilitada si empleamos las coordenadas de los vectores en otra base ordenada C. Entonces aparece naturalmente el siguiente problema:

Dadas las coordenadas de cierto vector v en la base ordenada B, obtener las coordenadas de v en la base ordenada C

cuya solución se expresa en el teorema que sigue.

Teorema 1.14 Matriz de Cambio de Base

Sean B y C bases ordenadas del espacio vectorial V, donde $B = \{v_1; v_2; ...; v_n\}$. Entonces existe una única matriz $M \in \mathbb{K}^{n\times n}$ tal que

$$[v]_C = M\,[v]_B \quad \forall\, v \in V$$

Además, M es la matriz cuyas columnas son las coordenadas de los vectores de la base B con respecto a la base C. Es decir

$$M = \begin{bmatrix} [v_1]_C & [v_2]_C & \cdots & [v_n]_C \end{bmatrix} \tag{1.13}$$

A la matriz M se la denomina ***matriz de cambio de coordenadas*** de la base B a la base C y se la denota C_{BC}.

Demostración:
Consideremos un vector $v \in V$ arbitrario. Entonces existen y son únicas sus coordenadas con respecto a la base B: $[v]_B = [\alpha_1\ \alpha_2\ ...\ \alpha_n]^t$. Vale decir que

$$v = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n$$

Tomando coordenadas y aplicando el teorema 1.12 obtenemos

$$[v]_C = \alpha_1\,[v_1]_C + \alpha_2\,[v_2]_C + ... + \alpha_n\,[v_n]_C$$

En esta última expresión se aprecia que para realizar el cambio de coordenadas alcanza con saber las coordenadas con respecto a la base C de los vectores de la base B. Tomando en consideración la matriz definida en 1.13 y pensando al producto entre matriz y vector como una combinación lineal de las columnas de la matriz, la última expresión queda

$$[v]_C = \alpha_1 M_1 + \alpha_2 M_2 + ... + \alpha_n M_n = M\,[v]_B$$

como queríamos probar. Falta ver la unicidad de la matriz de cambio de base. Para eso y como es usual, supongamos que existe cierta matriz $N \in \mathbb{K}^{n\times n}$ que realiza el mismo cambio de base que M, es decir $[v]_C = N\,[v]_B\ \ \forall\, v \in V$. Probaremos que

todas las columnas N_i (con $i = 1...n$) de la matriz N coinciden con las respectivas columnas de M. Para ello, tengamos en cuenta que las coordenadas de los vectores de la base B con respecto a esa misma base son los vectores canónicos de $\mathbb{K}^n$, es decir $[v_i]_B = e_i$. Con eso resulta

$$M_i = [v_i]_C = N[v_i]_B = \begin{bmatrix} N_1 & N_2 & \cdots & N_n \end{bmatrix} e_i = N_i$$

como queríamos demostrar. □

A continuación presentamos dos propiedades de la matriz de cambio de base.

Teorema 1.15

Sean B, C y D bases ordenadas del espacio vectorial V de dimensión finita. Entonces

1. $C_{BD} = C_{CD}C_{BC}$
2. C_{BC} es inversible y $(C_{BC})^{-1} = C_{CB}$

Demostración:
Empecemos por el punto 1. Es suficiente comprobar que para todo $v \in V$ se verifica que $[v]_D = (C_{CD}C_{BC})[v]_B$, ya que la matriz de cambio de base es única. Ahora bien,

$$(C_{CD}C_{BC})[v]_B = C_{CD}(C_{BC}[v]_B) = C_{CD}([v]_C) = [v]_D$$

En cuanto al punto 2, notemos en primer lugar que $C_{BB} = I$ donde I representa la matriz identidad. En efecto, como para todo $v \in V$ es $[v]_B = I[v]_B$, la unicidad de la matriz de cambio de base garantiza que $C_{BB} = I$. Entonces, usando el punto 1 deducimos que $C_{BC}C_{CB} = C_{CC} = I$ y $C_{CB}C_{BC} = C_{BB} = I$ y con ello que C_{BC} es inversible y $(C_{BC})^{-1} = C_{CB}$. □

Ejemplo 1.35

Consideremos la base $B = \{1-t, 1+t, 1+t+t^2\}$ de $\mathscr{P}_2$ y supongamos que deseamos hallar las coordenadas de $p = 2t^2 - 2t + 4$ en la base B.

Comencemos por señalar que las coordenadas de p con respecto a $E = \{1, t, t^2\}$, la base canónica de $\mathscr{P}_2$, son $[p]_E = [4\ -2\ 2]^t$. De acuerdo con el teorema 1.14, $[p]_B = C_{EB}[p]_E$. Por ende, debemos calcular C_{EB}. Para ello señalamos dos alternativas:

1. Obtener las coordenadas de los vectores de la base E con respecto a la base B. Esto implicaría resolver los tres sistemas de ecuaciones que surgen de plantear

$$\begin{aligned} 1 &= \alpha_1(1-t) + \alpha_2(1+t) + \alpha_3\left(1+t+t^2\right) \\ t &= \beta_1(1-t) + \beta_2(1+t) + \beta_3\left(1+t+t^2\right) \\ t^2 &= \gamma_1(1-t) + \gamma_2(1+t) + \gamma_3\left(1+t+t^2\right) \end{aligned}$$

2. Construir la matriz C_{BE} y, aplicando el teorema 1.15 obtener C_{EB} como su inversa.

Desarrollamos la segunda alternativa: la matriz C_{BE} se construye con facilidad puesto que

$$[1-t]_E = [1\ -1\ 0]^t \quad [1+t]_E = [1\ 1\ 0]^t \quad \left[1+t+t^2\right]_E = [1\ 1\ 1]^t$$

con lo cual

$$C_{BE} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 1 & 1 \\ 0 & 0 & 1 \end{bmatrix}$$

y su inversa[a] es la matriz buscada:

$$C_{EB} = (C_{BE})^{-1} = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

de modo que

$$[p]_B = C_{EB}\,[p]_E = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} & 0 \\ \frac{1}{2} & \frac{1}{2} & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 4 \\ -2 \\ 2 \end{bmatrix} = \begin{bmatrix} 3 \\ -1 \\ 2 \end{bmatrix}$$

[a] Quienes calculen la inversa con el conocido método que consiste en yuxtaponer la matriz identidad y triangular, notarán la relación con la primera alternativa que consistía en resolver tres sistemas de ecuaciones.

1.8 Operaciones entre subespacios

Para estudiar las posibles operaciones entre subespacios comenzamos con un ejemplo.

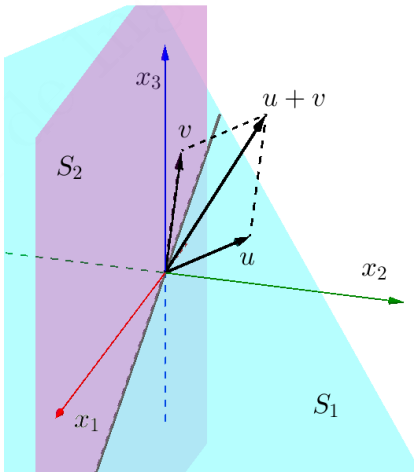


Figura 1.4: Dos subespacios de $\mathbb{R}^3$

Ejemplo 1.36

Consideremos los planos que pasan por el origen

$$S_1 = \{x \in \mathbb{R}^3 : x_1 + 2x_2 + 2x_3 = 0\} \ \wedge \ S_2 = \{x \in \mathbb{R}^3 : x_2 = 0\}$$

Si consideramos su unión, se trata del conjunto de vectores que pertenecen a un plano o al otro:

$$S_1 \cup S_2 = \{x \in \mathbb{R}^3 : x \in S_1 \vee x \in S_2\}$$

Elijamos los vectores $u = (-2,1,0) \in S_1$ y $v = (-1,0,2) \in S_2$. Al calcular la suma resulta que $u + v = (-3,1,2) \notin S_1 \cup S_2$, como se aprecia en la figura 1.4. En consecuencia, $S_1 \cup S_2$ no es subespacio pues la suma no es cerrada.

Si en cambio consideramos la intersección de estos subespacios, es decir el conjunto de vectores que pertenecen a ambos:

$$S_1 \cap S_2 = \left\{x \in \mathbb{R}^3 : x \in S_1 \wedge x \in S_2\right\}$$

queda planteado el sistema de ecuaciones

$$x_1 + 2x_2 + 2x_3 = 0 \tag{1.14}$$

$$x_2 = 0 \tag{1.15}$$

cuya solución consta de todos los vectores de la forma $(-2x_3, 0, x_3) = x_3\,(-2,0,1)$ con $x_3 \in \mathbb{R}$. Vale decir que

$$S_1 \cap S_2 = \text{gen}\,\{(-2,0,1)\}$$

y se trata efectivamente de un subespacio: en este caso es una recta que pasa por el origen.

En general, ni la diferencia entre dos subespacios ni el complemento de un subespacio pueden ser subespacios, ya que en ambos casos el vector nulo no pertenece al conjunto resultante. Ya hemos visto que la unión de subespacios no tiene por qué ser un subespacio. El teorema siguiente confirma lo que sugiere el ejemplo: que la intersección entre subespacios es a su vez un subespacio.

Teorema 1.16

Dados $S_1, S_2, ..., S_r$ subespacios de un espacio vectorial V, su intersección

$$S_1 \cap S_2 \cap ... \cap S_r = \{v \in V : v \in S_1 \wedge v \in S_2 \wedge ... \wedge v \in S_r\}$$

también es subespacio de V.

Demostración:
Probaremos las tres condiciones que exige el teorema del subespacio 1.2.

1. Es claro que $0 \in S_1 \cap S_2 \cap ... \cap S_r$, luego la intersección no es vacía.
2. Sean $u, v \in S_1 \cap S_2 \cap ... \cap S_r$. Luego para todo $i = 1, ..., r$ resulta, usando el hecho de que todos los S_i son subespacios,

$$u, v \in S_i \Rightarrow u + v \in S_i \Rightarrow u + v \in S_1 \cap S_2 \cap ... \cap S_r$$

3. Sean $u \in S_1 \cap S_2 \cap ... \cap S_r$ y $\alpha \in \mathbb{K}$. Luego para todo $i = 1, ..., r$ resulta

$$u \in S_i \Rightarrow ku \in S_i \Rightarrow ku \in S_1 \cap S_2 \cap ... \cap S_r$$

□

Hasta ahora, hemos considerado operaciones de conjuntos: unión, intersección, complemento y diferencia. A continuación definimos una nueva operación basada en las operaciones del espacio vectorial.

Definición 1.15

Dados $S_1, S_2, ..., S_r$ subespacios de un espacio vectorial V, la ***suma*** de estos subespacios se denota como $S_1 + S_2 + ... + S_r$ y se define como el conjunto de todas las sumas posibles de vectores en $S_1, S_2, ..., S_r$. Es decir,

$$S_1 + S_2 + ... + S_r = \{v \in V : v = v_1 + v_2 + ... + v_r \text{ donde } v_1 \in S_1, v_2 \in S_2, ..., v_r \in S_r\}$$

Ejemplo 1.37

Si retomamos el ejemplo 1.36, la suma de los subespacios S_1 y S_2 es, como el gráfico permite anticipar, todo el espacio $\mathbb{R}^3$. Para demostrarlo formalmente, notemos que

$$S_1 = \text{gen}\{(0,-1,1),(2,-1,0)\} \quad S_2 = \text{gen}\{(1,0,0),(0,0,1)\}$$

Luego los elementos de S_1 son de la forma $\alpha(0,-1,1) + \beta(2,-1,0)$ con α, β arbitrarios, mientras que los de S_2 son de la forma $\gamma(1,0,0) + \delta(0,0,1)$ con γ, δ arbitrarios. Por ende, los elementos de $S_1 + S_2$ serán de la forma

$$v = \alpha(0,-1,1) + \beta(2,-1,0) + \gamma(1,0,0) + \delta(0,0,1)$$

Es decir que $S_1 + S_2$ contendrá todas las combinaciones lineales de los vectores $(0,-1,1), (2,-1,0), (1,0,0), (0,0,1)$, que constituyen un conjunto generador de $\mathbb{R}^3$.

En el ejemplo anterior la suma de subespacios resultó ser un subespacio. El teorema siguiente generaliza este hecho.

Teorema 1.17

Dados $S_1, S_2, ..., S_r$ subespacios de un espacio vectorial V, la suma $S_1 + S_2 + ... + S_r$ es un subespacio de V.

Demostración:
Una vez más, probaremos las tres condiciones que exige el teorema del subespacio 1.2.

1. Dado que $0 \in S_1 \wedge 0 \in S_2 \wedge ... \wedge 0 \in S_r$, resulta $0 = 0 + 0 + ... + 0 \Rightarrow 0 \in S_1 + S_2 + ... + S_r$
2. Sean $u, v \in S_1 + S_2 + ... + S_r$. Entonces existen $u_1 \in S_1, u_2 \in S_2, ..., u_r \in S_r$ tales que $u = u_1 + u_2 + ... + u_r$. Análogamente $v = v_1 + v_2 + ... + v_r$ para ciertos $v_1 \in S_1, v_2 \in S_2, ..., v_r \in S_r$. Luego, teniendo en cuenta que los $S_1, S_2, ..., S_r$ son subespacios
$$\begin{aligned} u + v &= (u_1 + u_2 + ... + u_r) + (v_1 + v_2 + ... + v_r) \\ &= \underbrace{(u_1 + v_1)}_{\in S_1} + \underbrace{(u_2 + v_2)}_{\in S_2} + ... + \underbrace{(u_r + v_r)}_{\in S_r} \\ &\Rightarrow u + v \in S_1 + S_2 \end{aligned}$$
3. Sean $u \in S_1 + S_2 + ... + S_r$ y $\alpha \in \mathbb{K}$. Entonces existen $u_1 \in S_1, u_2 \in S_2, ..., u_r \in S_r$ tales que $u = u_1 + u_2 + ... + u_r$. Por lo tanto
$$\begin{aligned} \alpha u &= \alpha(u_1 + u_2 + ... + u_r) \\ &= \underbrace{\alpha u_1}_{\in S_1} + \underbrace{\alpha u_2}_{\in S_2} + ... + \underbrace{\alpha u_r}_{\in S_r} \\ &\Rightarrow \alpha u \in S_1 + S_2 + ... + S_r \end{aligned}$$

□

Ejercicio.

1.8 Compruebe que en $\mathbb{R}^2$ la suma de dos rectas rectas distintas S_1 y S_2 que pasan por el origen es $\mathbb{R}^2$. Interprete gráficamente. ¿Qué sucede si las rectas están contenidas en $\mathbb{R}^3$?

El método empleado en el último ejemplo para hallar un conjunto generador del subespacio suma se puede generalizar sin mayores dificultades (dejamos los detalles como ejercicio):

Proposición 1.5

Sean $\{v_{11}, v_{12}, ..., v_{1k_1}\}, \{v_{21}, v_{22}, ..., v_{2k_2}\}, ..., \{v_{r1}, v_{r2}, ..., v_{rk_r}\}$ conjuntos generadores de los subespacios $S_1, S_2, ..., S_r$ de un espacio vectorial V respectivamente. Entonces

$$S_1 + S_2 + ... + S_r = \text{gen}\{v_{11}, v_{12}, ..., v_{1k_1}, v_{21}, v_{22}, ..., v_{2k_2}, ..., v_{r1}, v_{r2}, ..., v_{rk_r}\}$$

es decir, la unión de conjuntos generadores de los subespacios $S_1, S_2, ..., S_r$, es un conjunto de generadores de $S_1 + S_2 + ... + S_r$.

Ejemplo 1.38

Obtener una base de $S_1 + S_2$, siendo

$$S_1 = \{x \in \mathbb{R}^4 : x_1 + 2x_2 - x_3 + x_4 = 0 \wedge x_1 - x_4 = 0\}$$

y

$$S_2 = \text{gen}\{(1,1,0,0), (2,1,2,1)\}$$

Para resolver el problema necesitamos un conjunto generador de S_1. Resolviendo las ecuaciones lineales homogéneas que definen el subespacio S_1, obtenemos la siguiente base de S_1: $\{(1,-1,0,1), (0,1,2,0)\}$. Como $\{(1,1,0,0), (2,1,2,1)\}$ genera S_2, aplicando la proposición 1.5 tenemos que

$$\{(1,-1,0,1), (0,1,2,0), (1,1,0,0), (2,1,2,1)\}$$

es un conjunto generador de $S_1 + S_2$. Sin embargo, este conjunto no es base, ya que es linealmente dependiente. Aplicando, por ejemplo, eliminación gaussiana, se verifica que el cuarto vector es combinación lineal de los tres primeros, y que estos son linealmente independientes, con lo cual

$$\{(1,-1,0,1), (0,1,2,0), (1,1,0,0)\}$$

es la base buscada.

N Como se desprende del ejemplo anterior, la unión de bases de los subespacios es un conjunto generador de la suma, pero no es necesariamente base de esa suma.

El teorema que enunciamos a continuación permite obtener nuevas conclusiones acerca de la suma de dos subespacios. [4]

[4] La demostración puede ser un buen repaso de los conceptos estudiados en este capítulo. Recomendamos consultarla en la bibliografía, por ejemplo en [4].

Teorema 1.18

Sean S_1 y S_2 subespacios de dimensión finita de un espacio vectorial V. Entonces $S_1 + S_2$ es de dimensión finita y

$$\dim(S_1 + S_2) = \dim(S_1) + \dim(S_2) - \dim(S_1 \cap S_2) \tag{1.16}$$

Ejemplo 1.39

Si retomamos el ejemplo 1.38, podemos aplicar el teorema anterior para deducir la dimensión de la intersección. En efecto, la fórmula (1.16) resulta en este caso

$$3 = 2 + 2 - \dim(S_1 \cap S_2)$$

y concluimos que $\dim(S_1 \cap S_2) = 1$.

Como ya hemos definido en 1.15, cada elemento v de $S_1 + S_2 + \ldots + S_r$ puede expresarse en la forma

$$v = v_1 + v_2 + \ldots + v_r \quad \text{donde } v_1 \in S_1, v_2 \in S_2, \ldots, v_r \in S_r \tag{1.17}$$

Sin embargo, esta descomposición de v como suma de vectores en los subespacios $S_1, S_2, \ldots, S_r$ no tiene por qué ser única.

Ejercicio. Considerando el ejemplo 1.36, encuentre dos descomposiciones distintas del vector $v = (1,1,1)$ como suma de elementos de S_1 y S_2.

En lo que sigue, prestaremos atención al caso en que la descomposición es única.

Definición 1.16

Dados $S_1, S_2, \ldots, S_r$ subespacios de un espacio vectorial V, se dice que la suma es ***directa*** si cada $v \in S_1 + S_2 + \ldots + S_r$ admite una única descomposición en la forma (1.17). Es decir, si los $v_1 \in S_1, v_2 \in S_2, \ldots, v_r \in S_r$ son únicos. Cuando la suma es directa emplearemos la notación $S_1 \oplus S_2 \oplus \ldots \oplus S_r$.

Ejemplo 1.40

La suma de un subespacio arbitrario S de un espacio vectorial V con el subespacio nulo $\{0\}$ es directa, es decir

$$S \oplus \{0\} = S$$

pues la única forma de expresar $v \in S$, como suma de un elemento de S con uno de $\{0\}$ es $v = v + 0$.

Ejemplo 1.41

La suma de dos rectas distintas de $\mathbb{R}^3$ que pasan por el origen es directa.

En efecto, la suma es el plano que contiene a ambas rectas (ver ejercicio 1.8) y cada vector de ese plano puede descomponerse en forma única como suma de dos vectores, uno en la primera recta y otro en la segunda.

Ejercicio.

1.9 Demuestre que en $\mathbb{R}^3$ es directa la suma de un plano S_1 que pasa por el origen con una recta S_2 que también pasa por el origen pero no está contenida en el plano.

Ejemplo 1.42

En $\mathbb{R}^3$ la suma de dos planos distintos S_1 y S_2 que contienen al origen nunca es directa.

En efecto, sea $L = S_1 \cap S_2$. Entonces L es una recta que pasa por el origen (¿por qué no puede ser tan solo un punto?). Sea w cualquier vector no nulo de L. Entonces $w \in S_1$ y $w \in S_2$, con lo cual también $-w \in S_2$. Pero entonces $0 \in S_1 + S_2$ admite las dos siguientes descomposiciones:

$$0 = 0 + 0 \text{ y } 0 = w + (-w)$$

que demuestran que la suma de S_1 con S_2 no es directa.

Señalemos que en los ejemplos anteriores en los cuales la suma de los subespacios es directa, su intersección es el subespacio nulo. En cambio en el ejemplo en que la suma no es directa, la intersección de los subespacios no es el subespacio nulo. El teorema siguiente generaliza esta observación.

Teorema 1.19

Sean S_1 y S_2 subespacios de un espacio vectorial V. Entonces la suma de S_1 y S_2 es directa si y sólo si $S_1 \cap S_2 = \{0\}$.

Demostración:
Para demostrar la doble implicación, asumimos en primer lugar que la suma de S_1 y S_2 es directa. Evidentemente, $0 \in S_1 \cap S_2$: supongamos que existe otro vector $v \in S_1 \cap S_2$. En tal caso, como v pertenece a ambos subespacios se podrían hacer dos descomposiciones distintas

$$v = \underbrace{0}_{\in S_1} + \underbrace{v}_{\in S_2} = \underbrace{v}_{\in S_1} + \underbrace{0}_{\in S_2}$$

en contra de que la suma es directa.

Asumamos ahora que $S_1 \cap S_2 = \{0\}$. Razonando nuevamente por el absurdo, supongamos que existe cierto vector v que admite dos descomposiciones distintas:

$$v = \underbrace{v_1}_{\in S_1} + \underbrace{v_2}_{\in S_2} = \underbrace{w_1}_{\in S_1} + \underbrace{w_2}_{\in S_2}$$

Pero entonces

$$v_1 - w_1 = w_2 - v_2$$

y dado que $v_1 - w_1 \in S_1$, $w_2 - v_2 \in S_2$, como la intersección es el vector nulo, deducimos que

$$v_1 - w_1 = w_2 - v_2 = 0 \Rightarrow v_1 = w_1 \wedge v_2 = w_2$$

luego la descomposición es única y la suma es directa. □

Ejemplo 1.43

En el espacio de matrices cuadradas $\mathbb{R}^{n\times n}$ consideramos el subespacio S_1 de matrices simétricas (aquellas tales que $A = A^t$) y el subespacio S_2 de matrices antisimétricas (aquellas tales que $A = -A^t$). Dejamos como ejercicio comprobar que se trata efectivamente de subespacios. Probaremos que la suma de estos subespacios es directa e igual a todo el espacio, es decir $S_1 \oplus S_2 = \mathbb{R}^{n\times n}$.

Notemos que cualquier matriz $A \in \mathbb{R}^{n\times n}$ admite la descomposición

$$A = \frac{A + A^t}{2} + \frac{A - A^t}{2}$$

El primer término de la descomposición es una matriz simétrica, mientras que el segundo es una matriz antisimétrica:

$$\left(\frac{A + A^t}{2}\right)^t = \frac{A^t + \left(A^t\right)^t}{2} = \frac{A^t + A}{2} = \left(\frac{A + A^t}{2}\right)$$

$$\left(\frac{A - A^t}{2}\right)^t = \frac{A^t - \left(A^t\right)^t}{2} = \frac{A^t - A}{2} = -\left(\frac{A - A^t}{2}\right)$$

Esto prueba que la suma de S_1 y S_2 es todo el espacio de matrices. Falta ver que la suma es directa. Basándonos en el teorema 1.19, probaremos que $S_1 \cap S_2 = \{0\}$. Sea $A \in S_1 \cap S_2$: entonces A es simétrica y antisimétrica. Luego

$$A = A^t \wedge A = -A^t \Rightarrow A = -A \Rightarrow A = 0$$

En lo que sigue enunciaremos una condición que permite determinar cuándo los subespacios $S_1, S_2, ..., S_r$ están en suma directa. Para ello, señalemos que los teoremas 1.18 y 1.19 nos permiten afirmar que la suma dos subespacios S_1 y S_2 es directa si y sólo si $\dim(S_1 + S_2) = \dim(S_1) + \dim(S_2)$. El teorema siguiente generaliza esta observación.

Teorema 1.20

Sea V un espacio vectorial y sean $S_1, S_2, ..., S_r$ subespacios de dimensión finita. Entonces la suma $S_1 + S_2 + ... + S_r$ es directa si y sólo si se verifica que

$$\dim(S_1 + S_2 + ... + S_r) = \dim(S_1) + \dim(S_2) + ... + \dim(S_r) \tag{1.18}$$

Ejemplo 1.44

La suma de los subespacios de $\mathbb{R}^5$

$$S_1 = \text{gen}\{(1,1,0,0,-1),(1,0,1,0,0)\} \quad S_2 = \text{gen}\{(2,1,0,1,-1) \quad S_3 = \text{gen}\{(1,1,-1,1,0)\}\}$$

es directa, ya que el conjunto

$$\{(1,1,0,0,-1),(1,0,1,0,0),(2,1,0,1,-1),(1,1,-1,1,0)\}$$

es linealmente independiente y por lo tanto base de $S_1 + S_2 + S_3$, con lo cual se cumple la relación establecida en (1.18):

$$\underbrace{\dim(S_1 + S_2 + S_r)}_{4} = \underbrace{\dim(S_1)}_{2} + \underbrace{\dim(S_2)}_{1} + \underbrace{\dim(S_r)}_{1}\}$$

Ejemplo 1.45

En $\mathbb{R}^3$ consideremos un plano S_1 y dos rectas S_2 y S_3, distintas y no contenidas en el plano, tales que $S_1 \cap S_2 \cap S_3 = \{0\}$. ¿Qué es $S_1 + S_2 + S_3$? ¿Es directa la suma?

En primer lugar, $S_1 + S_2 + S_3 = \mathbb{R}^3$ (¿por qué?). La suma no es directa porque no se verifica la ecuación (1.18): en efecto, $\dim(S_1 + S_2 + S_3) = 3$ mientras que $\dim(S_1) + \dim(S_2) + \dim(S_3) = 4$.

Ejercicio.

1.10 Demuestre que la suma de los subespacios de $\mathscr{P}_3$

$$S_1 = \{p \in \mathscr{P}_3 : p(1) = 0 \wedge p(-1) = 0\} \quad S_2 = \text{gen}\{t^3 + 8\} \quad S_3 = \text{gen}\{2t^3 + t^2 + 15\}$$

no es directa.

2 Transformaciones lineales

2.1 Introducción

En la sección 1.5 del capítulo 1, dedicada a los subespacios fundamentales de una matriz, propusimos una interpretación de los sistemas de ecuaciones lineales de la forma $Ax = b$, según la cual el sistema es compatible si y sólo si b es combinación lineal de las columnas de A.

Una nueva interpretación posible del sistema de ecuaciones $Ax = b$ consiste en pensar que la expresión Ax define una función de $\mathbb{K}^n \to \mathbb{K}^m$, que a cada $x \in \mathbb{K}^n$ le asigna el vector $Ax \in \mathbb{K}^m$. En este contexto, el sistema de ecuaciones $Ax = b$ será compatible si el vector b pertenece a la imagen de la función. En este capítulo presentaremos una clase de funciones que permite, entre otras cosas, representar sistemas de ecuaciones. Para ello, destaquemos dos características que tiene la función definida con la fórmula Ax:

- $\forall x, y \in \mathbb{K}^n : A(x + y) = Ax + Ay$
- $\forall x \in \mathbb{K}^n, \forall \alpha \in \mathbb{K} : A(\alpha x) = \alpha(Ax)$

Definición 2.1

Dados dos espacios vectoriales V y W con el mismo conjunto de escalares $\mathbb{K}$, la función $f : V \to W$ es una ***transformación lineal*** si verifica

1. $\forall u, v \in V : f(u + v) = f(u) + f(v)$
2. $\forall u \in V, \forall \alpha \in \mathbb{K} : f(\alpha u) = \alpha f(u)$

Ejemplo 2.1

Cualquier función $f : \mathbb{K}^n \to \mathbb{K}^m$ definida mediante $f(x) = Ax$ donde $A \in \mathbb{K}^{m \times n}$ es una transformación lineal, por los argumentos expuestos al comienzo.

Si consideramos entonces la función definida mediante

$$f:\mathbb{R}^3\to\mathbb{R}^2, f\left[(x_1,x_2,x_3)\right]=(x_1-x_3,x_2+5x_3)$$

para probar que es una transformación lineal alcanza con mostrar que su expresión es de la forma Ax, en este caso con $A\in\mathbb{R}^{2\times 3}$. Usando una vez más la identificación entre los espacios $\mathbb{K}^r$ y $\mathbb{K}^{r\times 1}$, podemos escribir

$$\begin{aligned} f:\mathbb{R}^3\to\mathbb{R}^2, f\left([x_1\ x_2\ x_3]^t\right) &= [x_1-x_3\ \ x_2+5x_3]^t \\ &= \begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & 5 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} \end{aligned}$$

Convención

En el ejemplo anterior hemos usado la notación $f\left[(x_1,x_2,x_3)\right]$ para destacar que f asigna a cada vector $(x_1,x_2,x_3)\in\mathbb{R}^3$ un vector de $\mathbb{R}^2$. Sin embargo, muchas veces simplificaremos la notación escribiendo

$$f(x_1,x_2,x_3)=\ldots$$

Ejercicio.

2.1 Determine si la siguiente función es una transformación lineal

$$f:\mathscr{P}_2\to\mathbb{R}, f(p)=p(0)$$

Ejemplo 2.2

La función

$$f:\mathbb{R}^3\to\mathbb{R}^3, f(v)=v\times w$$

siendo w un vector fijo, es una transformación lineal. Veamos que cumple las dos condiciones de la definición, usando las propiedades del producto vectorial:

1. Sean $u,v\in\mathbb{R}^3$. Entonces
$$f(u+v)=(u+v)\times w=u\times w+v\times w=f(u)+f(v)$$
2. Sean $u\in\mathbb{R}^3,\alpha\in\mathbb{R}$. Entonces
$$f(\alpha u)=(\alpha u)\times w=\alpha(u\times w)=\alpha f(u)$$

Señalemos también que las propiedades del producto vectorial permiten afirmar que

- $f(0)=0$
- $f(-v)=-v$ para todo $v\in\mathbb{R}^3$.

Estas últimas observaciones se generalizan sin dificultad, como probaremos en el siguiente teorema.

Teorema 2.1

Dados los espacios vectoriales V, W y la transformación lineal $f : V \to W$, se cumple que:

1. $f(0_V) = 0_W$
2. $\forall u \in V : f(-u) = -f(u)$

Demostración:
Para probar la propiedad 1, usamos el hecho de que el cero por cualquier vector da el vector nulo (ver 2 del teorema 1.1) y la definición de transformación lineal:

$$f(0_V) = f(0v) = 0f(v) = 0_W$$

Análogamente

$$f(-v) = f[(-1)v] = (-1)f(v) = -f(v)$$

□

Las propiedades anteriores, además de tener interés por sí mismas, sirven también para detectar funciones que no son transformaciones lineales. En efecto, como toda transformación *necesariamente* las verifica, decimos que son *condición necesaria* para ser transformación lineal.

Ejemplo 2.3

Estudiemos las funciones

$$f : \mathbb{R}^2 \to \mathbb{R}^2, f(x_1, x_2) = (2x_1 - 1, x_2)$$
$$g : \mathbb{R}^2 \to \mathbb{R}^2, g(x_1, x_2) = \left(x_1^2, x_2\right)$$

En el caso de f, notemos que $f[(0,0)] = (-1,0) \neq (0,0)$. Por ende f no cumple una condición necesaria y entonces no puede ser transformación lineal.
En el caso de g, es $g(0) = 0$ y se cumple la condición necesaria. Sin embargo, que una función transforme al vector nulo en el vector nulo no es una *condición suficiente* para ser transformación lineal. En este caso, notemos que

$$g[3(1,1)] = g(3,3) = (9,3) \neq 3g(1,1) = (3,3)$$

Proposición 2.1

Dados los vectores $v_1, v_2, ..., v_r \in V$, los escalares $\alpha_1, \alpha_2, ..., \alpha_r \in \mathbb{K}$ y la transformación lineal $f : V \to W$, se cumple que

$$f(\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_r v_r) = \alpha_1 f(v_1) + \alpha_2 f(v_2) + ... + \alpha_r f(v_r)$$

La demostración surge al aplicar sucesivamente la definición de transformación lineal y la dejamos como ejercicio.

N A menudo diremos que las transformaciones lineales *preservan* las combinaciones lineales, en el sentido que señala la propiedad anterior. Queremos destacar que, de todas las funciones posibles entre espacios vectoriales, le prestamos atención a las transformaciones lineales por propiedades como esta.

2.2 Núcleo e Imagen

Para comenzar esta sección recordamos una notación que, si bien usaremos en el contexto de las transformaciones lineales, tiene validez para cualquier función $f : A \to B$, donde A y B son respectivamente el dominio y codominio de la función.

Definición 2.2

Sea $f : A \to B$ una función.

1. Dado un subconjunto del dominio $M \subseteq A$, al conjunto de los transformados de los elementos de M por la función f lo denominamos **imagen** de M por f. Con la notación $f(M)$ resulta
$$f(M) = \{f(m) : m \in M\}$$
2. Dado un subconjunto del codominio $N \subseteq B$, al conjunto de los elementos del dominio que la función f transforma en elementos de N lo denominamos **preimagen** de N por f. Con la notación $f^{-1}(N)$ resulta
$$f^{-1}(N) = \{x \in A : f(x) \in N\}$$
Cuando el conjunto N tiene un solo elemento, o sea $N = \{y\}$, es usual suprimir las llaves:
$$f^{-1}(y) = \{x \in A : f(x) = y\}$$

N La notación $f^{-1}(N)$ suele inducir a la confusión con la función inversa. Por eso es imprescindible tener en cuenta que la preimagen de un conjunto queda bien definida para *cualquier* función $f : A \to B$, tenga o no inversa.
Por ejemplo, la función $f : \mathbb{R} \to \mathbb{R}, f(x) = 1$ no es biyectiva y por ende no tiene inversa. Sin embargo podemos calcular las preimágenes de subconjuntos del codominio tales como $f^{-1}(1) = \mathbb{R}$ y $f^{-1}([-1,0]) = \emptyset$.

En las funciones de variable real $f : \mathbb{R} \to \mathbb{R}$ es usual estudiar sus *raíces* o *ceros* , es decir los $x \in \mathbb{R}$ tales que $f(x) = 0$. De igual manera definimos al *núcleo* de una transformación lineal.

Definición 2.3

Dada una transformación lineal $f : V \to W$, se define su ***núcleo*** como el conjunto de vectores del dominio cuya imagen es el vector nulo. Adoptando la notación $\text{Nu}\, f$ resulta
$$\text{Nu}\, f = \{v \in V : f(v) = 0_W\} = f^{-1}(0_W)$$

Otro concepto que incorporamos, al igual que en las funciones de $\mathbb{R} \to \mathbb{R}$, es el de *conjunto imagen.*

Definición 2.4

Dada una transformación lineal $f : V \to W$, se define su ***imagen*** como el conjunto de vectores del codominio que son imagen por f de algún vector del dominio. Adoptando la notación $\operatorname{Im} f$ resulta

$$\operatorname{Im} f = \{f(v) : v \in V\} = f(V)$$

Ejemplo 2.4

Dada la transformación lineal

$$f : \mathbb{R}^3 \to \mathbb{R}^3, f(x_1, x_2, x_3) = (x_1 - 2x_2, 0, 2x_1 - 4x_2)$$

buscaremos su núcleo e imagen. Para encontrar el núcleo, planteamos la definición

$$\operatorname{Nu} f = \{(x_1, x_2, x_3) \in \mathbb{R}^3 : f(x_1, x_2, x_3) = (0, 0, 0)\}$$

de donde llegamos a

$$(x_1 - 2x_2, 0, 2x_1 - 4x_2) = (0, 0, 0) \Leftrightarrow \begin{cases} x_1 - 2x_2 = 0 \\ 2x_1 - 4x_2 = 0 \end{cases} \Leftrightarrow x_1 = 2x_2$$

Como no aparecen restricciones sobre x_3, la forma genérica de un elemento del núcleo es

$$(2x_2, x_2, x_3) = x_2(2, 1, 0) + x_3(0, 0, 1)$$

donde x_1 y x_3 son arbitrarios, con lo cual

$$\operatorname{Nu} f = \operatorname{gen}\{(2, 1, 0), (0, 0, 1)\}$$

y una posible base del núcleo es

$$B_{\operatorname{Nu} f} = \{(2, 1, 0), (0, 0, 1)\}$$

En cuanto a la imagen de la transformación lineal, podemos reescribir la expresión de f como

$$f(x_1, x_2, x_3) = x_1(1, 0, 2) + x_2(-2, 0, -4)$$

y como x_1 y x_2 pueden asumir cualquier valor real, concluimos que la imagen está generada por todas las combinaciones lineales de los vectores $(1, 0, 2)$ y $(-2, 0, -4)$:

$$\operatorname{Im} f = \operatorname{gen}\{(1, 0, 2), (-2, 0, -4)\}$$

y dado que son vectores linealmente dependientes

$$\operatorname{Im} f = \operatorname{gen}\{(1, 0, 2)\}$$

con lo cual una base para el conjunto imagen es

$$B_{\operatorname{Im} f} = \{(1, 0, 2)\}$$

En el ejemplo anterior, los conjuntos núcleo e imagen resultaron ser subespacios del dominio y codominio respectivamente. Veremos a continuación que esto es el caso particular de una propiedad más general.

Teorema 2.2

Dada una transformación lineal $f : V \to W$:

1. Si $S \subseteq V$ es un subespacio de V, entonces $f(S)$ es un subespacio de W.
2. Si $T \subseteq W$ es un subespacio de W, entonces $f^{-1}(T)$ es un subespacio de V.

Demostración:
En ambos casos, debemos probar que se verifican las condiciones de subespacio establecidas en el teorema 1.2.

Para demostrar 1, señalemos en primer lugar que $0_V \in S$ por ser subespacio, luego $f(0_V) = 0_W \in f(S)$.
Consideremos ahora dos vectores $w_1, w_2 \in f(S)$: esto significa que existen $v_1, v_2 \in V$ tales que $f(v_1) = w_1 \wedge f(v_2) = w_2$. Entonces $w_1 + w_2 = f(v_1) + f(v_2) = f(v_1 + v_2)$ y dado que $v_1 + v_2 \in S$, deducimos que $w_1 + w_2 \in f(S)$.
Finalmente, consideremos $w \in f(S)$ y $\alpha \in \mathbb{K}$: debe existir $v \in S$ tal que $f(v) = w$. Entonces $\alpha w = \alpha f(v) = f(\alpha v)$ y dado que $\alpha v \in S$, deducimos que $\alpha w \in f(S)$.

La demostración de la propiedad 2 es análoga y la dejamos como ejercicio. □

Ejercicio.

2.2 Demuestre que para cualquier transformación lineal $f : V \to W$, el núcleo y la imagen son subespacios de V y de W respectivamente.

Si revisamos las dimensiones del núcleo y la imagen en el ejemplo 2.4, notamos que $\dim(\text{Nu}\, f) + \dim(\text{Im}\, f) = \dim(\mathbb{R}^3)$. Generalizamos esta relación en el teorema que sigue[1].

Teorema 2.3 (de la dimensión)

Dada una transformación lineal $f : V \to W$ definida en un espacio V de dimensión finita, se cumple la relación:

$$\dim(V) = \dim(\text{Nu}\, f) + \dim(\text{Im}\, f) \tag{2.1}$$

Ejemplo 2.5

Dada la transformación lineal

$$f : \mathbb{R}^3 \to \mathbb{R}^{2\times 2}, f(x_1, x_2, x_3) = \begin{bmatrix} x_1 + x_2 & x_2 + x_3 \\ 2x_1 & 3x_2 \end{bmatrix}$$

determinaremos su núcleo e imagen.

Ahora bien, reescribiendo la fórmula de f tenemos que

$$f(x_1, x_2, x_3) = x_1 \begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix} + x_2 \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix} + x_3 \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$$

[1] La demostración es un poco trabajosa y hemos decidido omitirla. Sin embargo, puede ser un ejercicio muy instructivo leerla en algún texto de la bibliografía como [4]

con lo cual deducimos que

$$\operatorname{Im} f = \operatorname{gen}\left\{\begin{bmatrix} 1 & 0 \\ 2 & 0 \end{bmatrix}, \begin{bmatrix} 1 & 1 \\ 0 & 3 \end{bmatrix}, \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right\}$$

y siendo estas tres matrices linealmente independientes (compruébelo) podemos afirmar que $\dim\left(\operatorname{Im} f\right) = 3$. Entonces del teorema de la dimensión 2.3 deducimos que $\dim(\operatorname{Nu} f) = 0$ y que por lo tanto

$$\operatorname{Nu} f = \{(0,0,0)\}$$

En la sección 1.5 del capítulo 1 relacionamos cada matriz $A \in \mathbb{K}^{m\times n}$ y sus espacios fundamentales con el sistema de ecuaciones que se le asocia. Ahora estamos en condiciones de interpretar un sistema de ecuaciones en términos de transformaciones lineales. Como ya anticipamos al comienzo de este capítulo, el sistema de ecuaciones

$$Ax = b \tag{2.2}$$

puede representarse por medio de la transformación lineal

$$f : \mathbb{K}^n \to \mathbb{K}^m,\ f(x) = Ax \tag{2.3}$$

con lo cual el sistema (2.2) será compatible si y sólo si $b \in \operatorname{Im} f$.

El núcleo de la transformación f es, por definición, el espacio nulo de la matriz A:

$$\operatorname{Nu} f = \{x \in \mathbb{K}^n : f(x) = 0\} = \{x \in \mathbb{K}^n : Ax = 0\} = \operatorname{Nul} A$$

y el conjunto imagen de la transformación coincide con el espacio columna de la matriz A:

$$\operatorname{Im} f = \{f(x) : x \in \mathbb{K}^n\} = \{Ax : x \in \mathbb{K}^n\} = \operatorname{Col} A$$

Con estas equivalencias, el teorema de la dimensión nos permite obtener nuevas conclusiones acerca de los sistemas lineales. En estas condiciones, la fórmula (2.1) queda

$$n = \dim(\operatorname{Nul} A) + \dim(\operatorname{Col} A)$$

Teniendo en cuenta que $\dim(\operatorname{Col} A)$ es el rango de la matriz y que n es la cantidad de incógnitas del sistema, queda demostrada la proposición siguiente.

Proposición 2.2

Dado un sistema de ecuaciones de la forma $Ax = b$, se verifica la relación

$$\text{cantidad de incógnitas} = \dim(\operatorname{Nul} A) + \operatorname{rango} A \tag{2.4}$$

N El rango de una matriz es la cantidad de columnas o filas linealmente independientes; como las operaciones elementales entre filas no lo alteran, rango A es la cantidad de filas no nulas que quedan después de escalonarla. A su vez, $\dim(\operatorname{Nul} A)$ es la cantidad de *variables libres* que tendrá la solución del sistema homogéneo.

2.3 Teorema Fundamental de las Transformaciones Lineales

Ya mencionamos al comienzo de este capítulo, más precisamente en la proposición 2.1, que las transformaciones lineales preservan las combinaciones lineales. Teniendo esto en cuenta, se llega sin dificultad a la siguiente proposición.

Proposición 2.3

Sea $f : V \to W$ una transformación lineal. Entonces los transformados de una base del dominio generan la imagen. Es decir, dada una base $B = \{v_1, v_2, ..., v_n\}$ de V:

$$\operatorname{Im} f = \operatorname{gen}\{f(v_1), f(v_2), ..., f(v_n)\}$$

Demostración:
En primer lugar, $w \in \operatorname{Im} f \Leftrightarrow \exists v \in V / f(v) = w$. Ahora bien, cualquier $v \in V$ es combinación lineal de los elementos de la base B, luego

$$\begin{aligned} w \in \operatorname{Im} f &\Leftrightarrow \exists \alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{K} / f(\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n) = w \\ &\Leftrightarrow \exists \alpha_1, \alpha_2, ..., \alpha_n \in \mathbb{K} / \alpha_1 f(v_1) + \alpha_2 f(v_2) + ... + \alpha_n f(v_n) = w \\ &\Leftrightarrow w \in \operatorname{gen}\{f(v_1), f(v_2), ..., f(v_n)\} \end{aligned}$$

□

Ejemplo 2.6

Consideremos la transformación lineal

$$f : \mathbb{R}^{2\times 2} \to \mathbb{R}^3, f\left(\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}\right) = (x_{11} + x_{12} + x_{22}, -x_{22}, x_{22})$$

y busquemos una base de $\operatorname{Im} f$.

De acuerdo con la proposición 2.3, alcanza con transformar una base de $\mathbb{R}^{2\times 2}$ para obtener un conjunto generador. Transformamos entonces la base canónica de $\mathbb{R}^{2\times 2}$:

$$f\left(\begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}\right) = (1,0,0) \quad f\left(\begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}\right) = (1,0,0) \quad f\left(\begin{bmatrix} 0 & 0 \\ 1 & 0 \end{bmatrix}\right) = (0,0,0) \quad f\left(\begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}\right) = (1,-1,1)$$

De esta forma, la proposición 2.3 nos permite afirmar que

$$\operatorname{Im} f = \operatorname{gen}\{(1,0,0), (1,0,0), (0,0,0), (1,-1,1)\}$$

Para obtener una base de $\operatorname{Im} f$ suprimimos los generadores linealmente dependientes:

$$B_{\operatorname{Im} f} = \{(1,0,0), (1,-1,1)\}$$

N En el caso de una transformación lineal $f:\mathbb{K}^n \to \mathbb{K}^m$ definida mediante $f(x) = Ax$ con $A \in \mathbb{K}^{m\times n}$, los transformados de los vectores de la base canónica son precisamente las columnas de A, en concordancia con que $\operatorname{Im} f = \operatorname{Col} A$.

Consideremos ahora el siguiente problema: buscamos una transformación lineal $f:\mathbb{R}^2 \to \mathbb{R}^2$ que transforme al cuadrado rojo en el cuadrado azul, como se indica en la figura 2.1 .

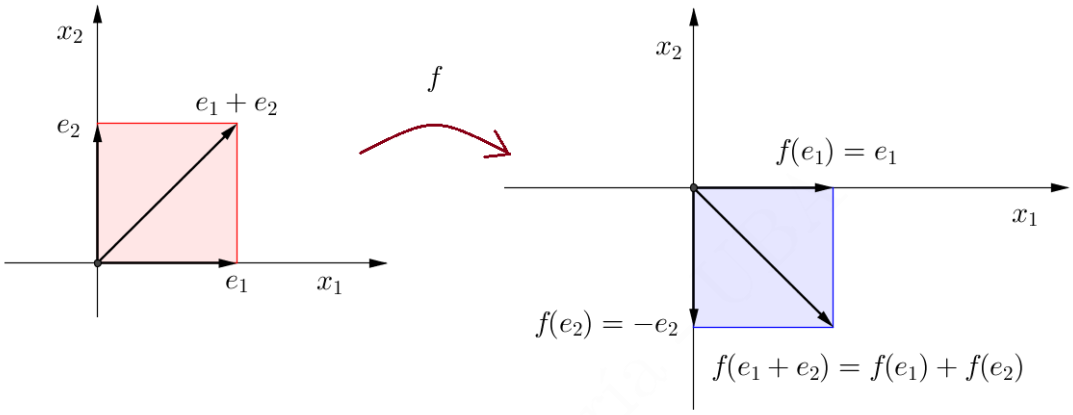


Figura 2.1: Transformación de $\mathbb{R}^2$ en $\mathbb{R}^2$

Para lograrlo, notemos lo siguiente: dado que f preserva las combinaciones lineales, alcanzará con describir cómo transforma a los vectores de la base canónica e_1 y e_2. En efecto, cualquier vector de $\mathbb{R}^2$ es de la forma

$$x = x_1 e_1 + x_2 e_2$$

y dado que f es lineal

$$f(x) = f(x_1 e_1 + x_2 e_2) = x_1 f(e_1) + x_2 f(e_2)$$

En este caso, una posibilidad es definir

$$f(e_1) = e_1 \quad \wedge \quad f(e_2) = -e_2$$

lo que lleva a la fórmula general

$$\begin{aligned} f[(x_1, x_2)] &= f(x_1 e_1 + x_2 e_2) \\ &= x_1 f(e_1) + x_2 f(e_2) \\ &= x_1 e_1 - x_2 e_2 \\ &= (x_1, -x_2) \end{aligned}$$

Ejercicio.

2.3 Definir una transformación lineal $f:\mathbb{R}^2 \to \mathbb{R}^2$ que transforme el cuadrado rojo de la figura 2.1 en:

- Un paralelogramo de vértices $(0,0), (1,0), (2,1), (1,1)$.
- El segmento de extremos $(0,0), (0,1)$.

En general, *una transformación lineal $f: V \to W$ quedará bien determinada si conocemos cómo se transforma una base del espacio V*. Esa es la idea que queremos destacar en el siguiente teorema.

Teorema 2.4 (Fundamental de las Transformaciones Lineales)

Sean V y W dos espacios vectoriales con $B = \{v_1, v_2, ..., v_n\}$ una base de V. Dados los vectores $w_1, w_2, ..., w_n$ de W (no necesariamente distintos), entonces existe una única transformación lineal $f : V \to W$ que verifica

$$f(v_i) = w_i \quad (i = 1, 2, ..., n)$$

La demostración formal del teorema es bastante técnica, pues implica probar que dicha transformación existe, que es lineal y que es única. Se pueden consultar los detalles en la bibliografía.

Ejemplo 2.7

Hallar la fórmula de la transformación lineal $f : \mathbb{R}^2 \to \mathbb{R}^3$ que verifica

$$f[(1,0)] = (0,1,0) \ \wedge \ f[(1,1)] = (0,1,0)$$

De acuerdo con el teorema, existe y es única la transformación lineal que verifica lo pedido, porque $B = \{(1,0),(1,1)\}$ es una base de $\mathbb{R}^2$. En este caso particular, las imágenes de los vectores de la base se repiten, por lo que podemos anticipar que $\dim(\text{Im}\, f) = 1$. Para hallar la fórmula de $f[(x_1, x_2)]$, tendremos en cuenta que contamos con la información de cómo se transforman los vectores $(1,0)$ y $(1,1)$. Luego, si logramos escribir al vector genérico (x_1, x_2) como combinación lineal de ellos

$$(x_1, x_2) = \alpha(1,0) + \beta(1,1)$$

para transformarlo sencillamente usamos la linealidad de f:

$$\begin{aligned} f[(x_1, x_2)] &= f\left[\alpha(1,0) + \beta(1,1)\right] \\ &= \alpha f[(1,0)] + \beta f[(1,1)] \\ &= \alpha(0,1,0) + \beta(0,1,0) \end{aligned}$$

Vale decir que nos falta conocer α y β, las coordenadas de (x_1, x_2) con respecto a la base B. Para ello planteamos

$$(x_1, x_2) = \alpha(1,0) + \beta(1,1)$$

de donde llegamos al sistema de ecuaciones

$$\begin{aligned} \alpha + \beta &= x_1 \\ \beta &= x_2 \end{aligned}$$

Entonces $\beta = x_2$ y $\alpha = x_1 - x_2$ y llegamos a

$$\begin{aligned} f[(x_1, x_2)] &= \alpha(0,1,0) + \beta(0,1,0) \\ &= (0, x_1, 0) \end{aligned}$$

2.4 Clasificación de las transformaciones lineales

Recordemos que para cualquier función $f : A \to B$, se dice que es *inyectiva* si verifica que

$$\forall x_1, x_2 \in A : f(x_1) = f(x_2) \Rightarrow x_1 = x_2$$

vale decir que a elementos distintos del dominio corresponden imágenes distintas. Si además dicha función es una transformación lineal, recibe un nombre especial.

Definición 2.5

Se dice que una transformación lineal inyectiva $f : V \to W$ es un ***monomorfismo***.

La propiedad siguiente caracteriza a los monomorfismos.

Teorema 2.5

La transformación lineal $f : V \to W$ es un monomorfismo si y sólo si $\operatorname{Nu} f = \{0_V\}$.

Demostración:
En primer lugar, asumamos que f es un monomorfismo. Sabemos que $f(0_V) = 0_W$ por ser una transformación lineal, luego $0_V \in \operatorname{Nu} f$. Como hemos supuesto que f es inyectiva, no puede haber otro vector que se transforme en cero. Por ende $\operatorname{Nu} f = \{0_V\}$.

Recíprocamente, supongamos que $\operatorname{Nu} f = \{0_V\}$. Necesitamos probar que f es inyectiva: de acuerdo con la definición de inyectividad, consideramos dos vectores $v_1, v_2 \in V$ tales que $f(v_1) = f(v_2)$. Entonces

$$\begin{aligned} f(v_1) = f(v_2) &\Rightarrow f(v_1) - f(v_2) = 0_W \\ &\Rightarrow f(v_1 - v_2) = 0_W \\ &\Rightarrow v_1 - v_2 \in \operatorname{Nu} f \end{aligned}$$

y como asumimos que el núcleo es trivial, se deduce que $v_1 - v_2 = 0_V$ como queríamos demostrar. □

En general, cualquier función $f : A \to B$, se dice *suryectiva* o *sobreyectiva* si alcanza todos los elementos del codominio. Es decir, si verifica que

$$f(A) = B$$

Si además se trata de una transformación lineal, adquiere un nombre especial.

Definición 2.6

Se dice que una transformación lineal sobreyectiva $f : V \to W$ es un ***epimorfismo***.
O sea que f es un epimorfismo si y sólo si

$$\operatorname{Im} f = W$$

Por último, cualquier función $f : A \to B$, se dice *biyectiva* si es simultáneamente inyectiva y sobreyectiva.

Definición 2.7

Se dice que una transformación lineal biyectiva $f : V \to W$ es un ***isomorfismo***..

Ejemplo 2.8

Consideremos la transformación lineal

$$f : \mathbb{R}^2 \to \mathbb{R}^3, f(x_1, x_2) = (x_1 + x_2, x_1 - x_2, 2x_1)$$

Calculamos en primer lugar el núcleo:

$$\begin{aligned}(x_1, x_2) \in \operatorname{Nu} f &\Leftrightarrow f(x_1, x_2) = (0,0,0)\\ &\Leftrightarrow (x_1 + x_2, x_1 - x_2, 2x_1) = (0,0,0)\\ &\Leftrightarrow x_1 = 0 \wedge x_2 = 0\end{aligned}$$

Entonces ya podemos afirmar que es un monomorfismo, puesto que $\operatorname{Nu} f = \{(0,0)\}$.

Por otra parte, planteando el teorema de la dimensión 2.3 tenemos para este caso

$$\underbrace{\dim\left(\mathbb{R}^2\right)}_{2} = \underbrace{\dim(\operatorname{Nu} f)}_{0} + \dim(\operatorname{Im} f)$$

de donde deducimos que $\dim(\operatorname{Im} f) = 2$ y por ende $\operatorname{Im} f \neq \mathbb{R}^3$, con lo cual f no es un epimorfismo (ni un isomorfismo).

Ejemplo 2.9

Consideremos la transformación lineal

$$g : \mathbb{R}^{2\times 2} \to \mathbb{R}^2, g\left(\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}\right) = (x_{11} + x_{12}, -x_{22})$$

Calculamos en primer lugar el conjunto imagen, reescribiendo la fórmula de g:

$$g\left(\begin{bmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{bmatrix}\right) = x_{11}(1,0) + x_{12}(1,0) + x_{22}(0,-1)$$

con lo cual

$$\operatorname{Im} g = \operatorname{gen}\{(1,0),(1,0),(0,-1)\}$$

y deducimos que $\dim\left(\operatorname{Im} g\right) = 2$, con lo cual $\operatorname{Im} g = \mathbb{R}^2$ y g es un epimorfismo.
Dejamos como ejercicio demostrar que g no es un monomorfismo.

Ejemplo 2.10

Consideremos la transformación lineal

$$h : \mathbb{R}^2 \to \mathscr{P}_1, h(a,b) = 2a - b\,t$$

Tal vez sea conveniente aclarar esta definición antes de seguir. Para ello, consideramos la imagen de $(2,-3)$:

$$h(2,-3) = 2\cdot 2 - (-3)\,t = 4+3t$$

Ahora sí, decidamos si h es monomorfismo. Para ello planteamos

$$\begin{aligned}(a,b) \in \operatorname{Nu} h &\Leftrightarrow h(a,b) = 0_{\mathscr{P}_1}\\ &\Leftrightarrow 2a - b\,t = 0_{\mathscr{P}_1}\\ &\Leftrightarrow a = 0 \wedge b = 0\end{aligned}$$

con lo cual $\operatorname{Nu} f = \{0_{\mathscr{P}_1}\}$ y se trata de un monomorfismo.
Por otra parte, al plantear el teorema de la dimensión 2.3 llegamos a

$$\underbrace{\dim\left(\mathbb{R}^2\right)}_{2} = \underbrace{\dim(\operatorname{Nu} f)}_{0} + \dim(\operatorname{Im} f)$$

de donde deducimos que $\dim(\operatorname{Im} f) = 2$ y por ende $\operatorname{Im} f = \mathscr{P}_1$. En este caso, h es un epimorfismo y un isomorfismo.

Ejercicio.

2.4 Considere una transformación lineal $f : V \to W$ entre espacios de dimensión finita.

- ¿Es posible que sea un isomorfismo si $\dim(V) \neq \dim(W)$?
- ¿Es posible que sea un monomorfismo si $\dim(V) > \dim(W)$?

La propiedad siguiente es una consecuencia directa del teorema de la dimensión 2.3 y dejamos su demostración como ejercicio.

Proposición 2.4

Sea $f : V \to W$ una transformación lineal entre espacios de igual dimensión n. Entonces f es monomorfismo si y sólo si f es epimorfismo si y sólo si f es isomorfismo.

Para terminar esta sección, profundizaremos en la relación que existe entre los vectores de un espacio y sus coordenadas con respecto a una base.

Sea entonces $B = \{v_1; v_2; \ldots v_n\}$ una base ordenada de un espacio vectorial V con escalares $\mathbb{K}$. Consideremos la aplicación que a cada vector del espacio le asigna sus coordenadas con respecto a la base B:

$$c_B : V \to \mathbb{K}^n, c_B(v) = [v]_B$$

Ya hemos demostrado en el teorema 1.12 que las coordenadas son lineales con respecto a la suma de vectores y la multiplicación por escalares. Por lo tanto c_B es una transformación lineal.

Estamos en condiciones de demostrar que además se trata de un isomorfismo. Veamos en primer lugar cuál es su núcleo:

$$\begin{aligned}v \in \operatorname{Nu} c_B &\Leftrightarrow c_B(v) = (0,0,\ldots,0)\\ &\Leftrightarrow v = 0v_1 + 0v_2 + \ldots + 0v_n\\ &\Leftrightarrow v = 0\end{aligned}$$

Luego el núcleo de c_B es trivial y se trata de un monomorfismo. Al aplicar el teorema de la dimensión 2.3 obtenemos $n = \dim(V) = \dim(\operatorname{Im} c_B)$, con lo cual $\operatorname{Im} c_B = \mathbb{K}^n$ y se trata de un epimorfismo. Agrupamos estas observaciones en un teorema.

Teorema 2.6

Dada una base ordenada $B = \{v_1; v_2; ...; v_n\}$ de un espacio vectorial V con escalares $\mathbb{K}$, la aplicación que a cada vector del espacio le asigna sus coordenadas con respecto a la base B

$$c_B : V \to \mathbb{K}^n, c_B(v) = [v]_B$$

es un isomorfismo, llamado ***isomorfismo de coordenadas***.

Queda establecida entonces una correspondencia lineal y biyectiva entre los vectores de un espacio y sus coordenadas con respecto a cierta base. La profunda identidad estructural que hay entre el espacio original y el espacio de coordenadas se hace evidente en resultados como el teorema 1.13, según el cual un conjunto de vectores es linealmente independiente si y sólo si sus coordenadas lo son.

2.5 Transformación Inversa

Al final de la sección anterior dejamos establecida una correspondencia uno a uno entre los vectores de un espacio vectorial y sus coordenadas con respecto a cierta base ordenada. Por tratarse de una biyección debe ser posible hacer el camino inverso, es decir: dadas las coordenadas del vector, recuperar el vector original. Manteniendo la notación establecida en el teorema 2.6, la función buscada es

$$(c_B)^{-1} : \mathbb{K}^n \to V, (c_B)^{-1}(\alpha_1, \alpha_2, ..., \alpha_n) = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_n v_n$$

En la figura 2.2 se aprecia la relación uno a uno entre los vectores y sus coordenadas.

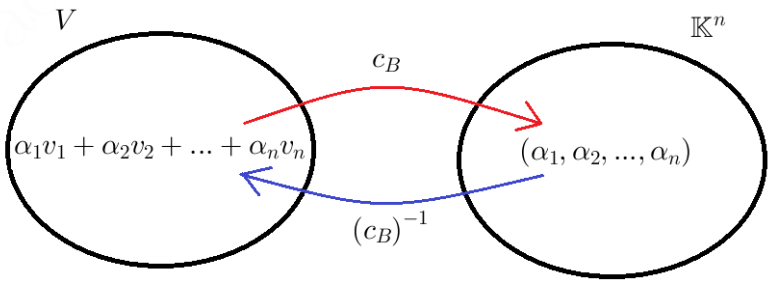


Figura 2.2: Isomorfismo de coordenadas y su inversa

Señalamos en primer lugar que la *composición* [2] de c_B y $(c_B)^{-1}$ resulta la transformación identidad en el espacio adecuado:

$$(c_B)^{-1} \circ c_B = I_V \quad \wedge \quad c_B \circ (c_B)^{-1} = I_{K^n}$$

[2] Recordemos que la composición $g \circ f$ de dos funciones $f : A \to B$ y $g : C \to D$ es posible siempre que $B \subseteq C$ y se define como $(g \circ f)(x) = g[f(x)]$.

Por lo tanto $(c_B)^{-1}$ es la *función inversa*[3] del isomorfismo de coordenadas c_B. Otra observación importante, cuya demostración dejamos como ejercicio, es que la inversa $(c_B)^{-1}$ es una transformación lineal.

Los comentarios anteriores se generalizan en el teorema siguiente.

Teorema 2.7

Sea $f : V \to W$ un isomorfismo. Entonces la función inversa $f^{-1} : W \to V$ existe y es también un isomorfismo.

Se verifica entonces que $f^{-1} \circ f = I_V$ y $f \circ f^{-1} = I_W$.

Demostración:
Por ser f una función biyectiva, sabemos que existe la función inversa $f^{-1} : W \to V$ y es biyectiva. Falta probar su linealidad. Sean entonces $w_1, w_2 \in W$: por ser f biyectiva existen y son únicos los $v_1, v_2 \in V$ tales que $f(v_1) = w_1$ y $f(v_2) = w_2$. Luego, usando la linealidad de f:

$$f^{-1}(w_1 + w_2) = f^{-1}\left(f(v_1) + f(v_2)\right) = f^{-1}\left(f(v_1 + v_2)\right) = \left(f^{-1} \circ f\right)(v_1 + v_2)$$

y como la composición $f^{-1} \circ f$ es la identidad en V resulta

$$f^{-1}(w_1 + w_2) = I_V(v_1 + v_2) = v_1 + v_2 = f^{-1}(w_1) + f^{-1}(w_2)$$

Esto prueba la linealidad de f^{-1} con respecto a la suma. La linealidad con respecto al producto por escalares es similar y se deja como ejercicio. □

Ejemplo 2.11

Sea la transformación lineal

$$f : \mathbb{R}^2 \to \mathscr{P}_1, f(a, b) = (a - b) + 2a\,t \tag{2.5}$$

Dejamos como ejercicio comprobar que se trata de un isomorfismo y que por ende existe la inversa.

Una forma de buscar la transformación inversa es estudiando cómo se transforma una base de $\mathbb{R}^2$:

$$f(e_1) = 1 + 2t \quad \wedge \quad f(e_2) = -1$$

Por lo tanto la transformación inversa $f^{-1} : \mathscr{P}_1 \to \mathbb{R}^2$ debe verificar

$$f^{-1}(1 + 2t) = e_1 \quad \wedge \quad f^{-1}(-1) = e_2$$

Dado que $\{1 + 2t, -1\}$ es una base de $\mathscr{P}_1$ y ya sabemos cómo se transforma, el teorema fundamental de las transformaciones lineales 2.4 garantiza que f^{-1} ya quedó bien definida.

Para hallar una expresión explícita de f^{-1} escribimos un polinomio genérico $a + bt \in \mathscr{P}_1$ como combinación lineal de los elementos de la base:

$$a + bt = \frac{b}{2}(1 + 2t) + \left(\frac{b}{2} - a\right)(-1)$$

[3] Recordemos que la función $f : A \to B$ admite inversa si y sólo si existe cierta función $f^{-1} : B \to A$ tal que $f^{-1} \circ f = I_A$ y $f \circ f^{-1} = I_B$. La función inversa existe si y sólo si la función es biyectiva.

y ahora usando la linealidad de f^{-1}

$$\begin{aligned} f^{-1}(a+bt) &= \frac{b}{2}f^{-1}(1+2t)+\left(\frac{b}{2}-a\right)f^{-1}(-1) \\ &= \frac{b}{2}e_1+\left(\frac{b}{2}-a\right)e_2 \\ &= \left(\frac{b}{2},\frac{b}{2}-a\right) \end{aligned}$$

Para comprobar la fórmula obtenida realizamos la composición

$$\begin{aligned} \left(f^{-1}\circ f\right)(a,b) &= f^{-1}\left[f(a,b)\right] \\ &= f^{-1}\left[(a-b)+2a\,t\right] \\ &= \left(\frac{2a}{2},\frac{2a}{2}-(a-b)\right) \\ &= (a,b) \end{aligned}$$

por lo que resulta efectivamente $\left(f^{-1}\circ f\right)=I_{\mathbb{R}^2}$

2.6 Matriz asociada a una transformación lineal

Al comienzo de este capítulo señalamos que cualquier función $f(x)=Ax$ con $A\in\mathbb{K}^{m\times n}$ define una transformación lineal de $\mathbb{K}^n\to\mathbb{K}^m$. Más aún, vimos en el ejemplo 2.1 cómo una transformación lineal de $\mathbb{K}^n\to\mathbb{K}^m$ se podía representar en la forma Ax. A esta matriz se la denomina *matriz de la transformación lineal.*

En esta sección veremos cómo cualquier transformación lineal entre espacios de dimensión finita se puede representar mediante la multiplicación por una matriz. Evidentemente, si tenemos una transformación lineal de $V\to W$ con V y W espacios arbitrarios, es posible que la multiplicación por matrices no esté definida. Por eso consideraremos los espacios de coordenadas asociados a V y W.

Para ilustrar estas ideas, retomamos el último ejemplo de la sección anterior.

Ejemplo 2.12

Consideremos entonces la transformación lineal definida en (2.5). Una vez más, la clave está en saber cómo se transforma una base. Por eso hacemos el siguiente manejo

$$f(x_1,x_2)=f(x_1e_1+x_2e_2)=x_1f(e_1)+x_2f(e_2)$$

La expresión anterior es una combinación lineal de $f(e_1)$ y $f(e_2)$, sus coeficientes son x_1 y x_2. Ahora bien, $f(e_1)$ y $f(e_2)$ son polinomios. Para transformar esta expresión en una combinación lineal de las columnas de una matriz, tomamos coordenadas con respecto a la base canónica de polinomios $E=\{1;t\}$:

$$\left[f(x_1,x_2)\right]_E=\left[f(x_1e_1+x_2e_2)\right]_E=x_1\left[f(e_1)\right]_E+x_2\left[f(e_2)\right]_E \tag{2.6}$$

Y ahora sí estamos en condiciones de expresar esto como combinación lineal de las columnas de una matriz, puesto

que

$$[f(e_1)]_E = \begin{bmatrix} 1 \\ 2 \end{bmatrix} \wedge [f(e_2)]_E = \begin{bmatrix} -1 \\ 0 \end{bmatrix}$$

Entonces la expresión (2.6) queda

$$[f(x_1, x_2)]_E = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$$

La matriz $\begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix}$ es la *matriz asociada a la transformación lineal* f (con respecto a las bases canónicas).

Para conocer el transformado del vector $(-2,3)$, sencillamente hacemos la cuenta indicada:

$$[f(-2,3)]_E = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} \begin{bmatrix} -2 \\ 3 \end{bmatrix} = \begin{bmatrix} -5 \\ -4 \end{bmatrix}$$

Donde el resultado se obtiene *en coordenadas* con respecto a la base E. El polinomio se recupera haciendo la combinación lineal que indican las coordenadas:

$$f(-2,3) = (-5) \cdot 1 + (-4) \cdot (t) = -5 - 4t$$

Destacamos del ejemplo anterior que una transformación lineal se puede representar con una matriz cuyas columnas son los transformados de los vectores de la base del dominio, en coordenadas con respecto a la base del codominio. La siguiente proposición generaliza lo que hemos señalado en el ejemplo. [4]

Proposición 2.5 (Matriz asociada a una transformación lineal)

Sea $f : V \to W$ una transformación lineal y sean

$$B = \{v_1; v_2; \ldots; v_n\} \wedge B' = \{w_1; w_2; \ldots; w_m\}$$

bases ordenadas de V y W respectivamente. La ***matriz asociada a f con respecto a las bases*** B ***y*** B' se denota $[f]_{BB'}$ y tiene por columnas las coordenadas con respecto a la base B' de los transformados de los vectores de la base B:

$$[f]_{BB'} = \begin{bmatrix} [f(v_1)]_{B'} & [f(v_2)]_{B'} & \cdots & [f(v_n)]_{B'} \end{bmatrix}$$

La matriz asociada a f verifica

$$[f]_{BB'}\,[x]_B = [f(x)]_{B'}$$

Convención

En el caso de que sea $f : V \to V$ y se considere una sola base B, adoptaremos la notación $[f]_B$ para la matriz asociada.

[4] La demostración formal no es difícil, aunque tiene una notación un poco exigente. Los lectores podrán buscarla en la bibliografía para repasar estos conceptos.

Ejemplo 2.13

Consideremos la transformación lineal

$$f:\mathbb{R}^2 \to \mathbb{R}^3, f(x_1,x_2) = (x_1+2x_2, x_1-x_2, x_2) \tag{2.7}$$

y las bases ordenadas

$$B = \{(0,1);(1,1)\} \ \wedge \ B' = \{(1,0,0);(1,1,0);(0,1,1)\}$$

Para construir la matriz asociada a f con respecto a estas bases, transformamos los vectores de la base B y los expresamos en coordenadas con respecto a la base B':

$$f(0,1) = (2,-1,1) = 4(1,0,0) + (-2)(1,1,0) + 1(0,1,1) \Rightarrow \left[f(0,1)\right]_{B'} = \begin{bmatrix} 4 \\ -2 \\ 1 \end{bmatrix}$$

$$f(1,1) = (3,0,1) = 4(1,0,0) + (-1)(1,1,0) + 1(0,1,1) \Rightarrow \left[f(0,1)\right]_{B'} = \begin{bmatrix} 4 \\ -1 \\ 1 \end{bmatrix}$$

con lo cual

$$[f]_{BB'} = \begin{bmatrix} 4 & 4 \\ -2 & -1 \\ 1 & 1 \end{bmatrix}$$

Para calcular $f[(3,5)]$, necesitamos expresar al vector en coordenadas con respecto a la base B:

$$(3,5) = 2(0,1) + 3(1,1)$$

luego

$$[(3,5)]_B = \begin{bmatrix} 2 \\ 3 \end{bmatrix}$$

y por lo tanto

$$\left[f(3,5)\right]_{B'} = \begin{bmatrix} 4 & 4 \\ -2 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 20 \\ -7 \\ 5 \end{bmatrix}$$

Para recuperar $f(3,5)$, debemos tener en cuenta que obtuvimos coordenadas con respecto a la base B', luego

$$f[(3,5)] = 20(1,0,0) + (-7)(1,1,0) + 5(0,1,1) = (13,-2,5)$$

Dejamos como ejercicio comprobar que si se eligen las base canónicas E_2 y E_3 la matriz asociada es

$$[f]_{E_2E_3} = \begin{bmatrix} 1 & 2 \\ 1 & -1 \\ 0 & 1 \end{bmatrix}$$

que es precisamente la matriz de la transformación que puede deducirse de la fórmula (2.7).

A continuación daremos una interpretación del rango de la matriz asociada a una transformación lineal. El rango de una matriz es el número de columnas linealmente independientes que tiene: en nuestro caso dichas columnas representan (en coordenadas) a los transformados de los vectores de la base, que son generadores de la imagen. Resumimos este argumento

en la propiedad siguiente.

Proposición 2.6

Sea $f : V \to W$ una transformación lineal entre espacios de dimensión finita. Sea $[f]_{BB'}$ la matriz asociada con respecto a ciertas bases. Entonces

$$\text{rango}\left([f]_{BB'}\right) = \dim\left(\text{Im}\, f\right)$$

(nótese que el rango no depende de las bases elegidas).

Señalemos que en el ejemplo 2.13, las dos matrices asociadas son de rango 2, por lo que la dimensión de la imagen es 2 y el teorema de la dimensión nos permite afirmar que el núcleo es trivial.

2.7 Composición de transformaciones lineales

Para presentar las ideas de esta sección retomamos (¡una vez más!) el ejemplo presentado en 2.11 y continuado en 2.12.

Ejemplo 2.14

Buscaremos la matriz asociada a $f^{-1} : \mathscr{P}_1 \to \mathbb{R}^2$ con respecto a las bases canónicas de $\mathscr{P}_1$, ($E = \{1, t\}$) y de $\mathbb{R}^2$ (E_2). Para ello deducimos, del trabajo hecho en ejemplo 2.11, que los transformados de los vectores de la base canónica de polinomios son (verifíquelo)

$$f^{-1}(1) = -e_2 \ \wedge \ f^{-1}(t) = \frac{1}{2}(e_1 + e_2)$$

con lo cual

$$[f^{-1}(1)]_{E_2} = \begin{bmatrix} 0 \\ -1 \end{bmatrix} \ \wedge \ [f^{-1}(t)]_{E_2} = \begin{bmatrix} 1/2 \\ 1/2 \end{bmatrix}$$

y entonces la matriz asociada es

$$[f^{-1}]_{EE_2} = \begin{bmatrix} 0 & 1/2 \\ -1 & 1/2 \end{bmatrix}$$

Dado que las coordenadas de un polinomio $a + bt$ con respecto a la base canónica son $[a\ b]^t$, la matriz nos permite obtener fácilmente la expresión de f^{-1}:

$$\begin{aligned} [f^{-1}(a+bt)]_E &= [f^{-1}]_{EE_2}\,[a+bt]_E \\ &= \begin{bmatrix} 0 & 1/2 \\ -1 & 1/2 \end{bmatrix} \begin{bmatrix} a \\ b \end{bmatrix} \\ &= \begin{bmatrix} b/2 \\ -a + b/2 \end{bmatrix} \end{aligned}$$

como ya conocíamos del ejemplo 2.11.

Ahora que ya conocemos las matrices $[f^{-1}]_{EE_2}$ y $[f]_{E_2E}$, estamos en condiciones de dar una interpretación de su

producto:

$$[f^{-1}]_{EE_2} \cdot [f]_{E_2E} = \begin{bmatrix} 0 & 1/2 \\ -1 & 1/2 \end{bmatrix} \cdot \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

$$[f]_{E_2E} \cdot [f^{-1}]_{EE_2} = \begin{bmatrix} 1 & -1 \\ 2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 0 & 1/2 \\ -1 & 1/2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} = I$$

¿Por qué las matrices asociadas son inversas? Pensemos qué sucede al multiplicar un vector $x \in \mathbb{R}^2$ por $[f]_{E_2E}$ y luego por $[f^{-1}]_{EE_2}$: el producto de las matrices $[f^{-1}]_{EE_2}$ y $[f]_{E_2E}$ da la matriz identidad porque esa es la matriz asociada a la composición $f^{-1} \circ f = I_{\mathbb{R}^2}$.

Las observaciones hechas en este ejemplo pueden extenderse al caso general sin mayores cambios. Queda como ejercicio demostrar las proposiciones que siguen.

Proposición 2.7

Sean

$$f : U \to V \quad g : V \to W$$

dos transformaciones lineales entre espacios de dimensión finita. Entonces

1. La composición de las transformaciones lineales f y g es una transformación lineal
$$(g \circ f) : U \to W, (g \circ f)(u) = g\left[f(u)\right]$$
2. Dadas B_1, B_2, B_3 bases de U, V, W respectivamente, la matriz asociada a la composición se obtiene haciendo el producto de matrices
$$[g \circ f]_{B_1B_3} = [g]_{B_2B_3} \cdot [f]_{B_1B_2}$$

La figura 2.3 ilustra la composición de las transformaciones lineales y las respectivas transformaciones de coordenadas.

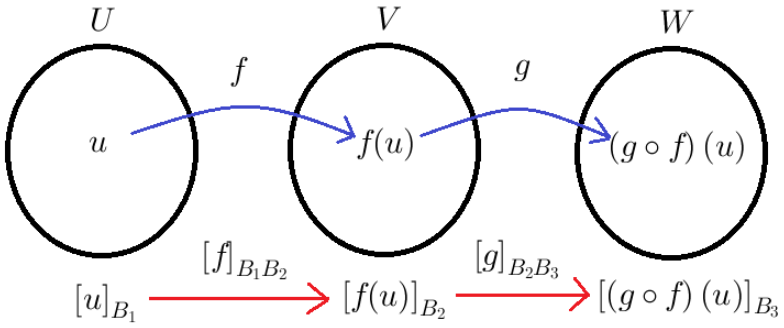


Figura 2.3: Composición de transformaciones lineales

Proposición 2.8

Sea $f : V \to W$ un isomorfismo entre espacios de dimensión finita. Sean B, B' bases de V y W respectivamente y sea $f^{-1} : W \to V$ la transformación inversa. Entonces las matrices asociadas $[f]_{BB'}$ y $[f^{-1}]_{B'B}$ son inversas. Es decir

$$[f^{-1}]_{B'B} = ([f]_{BB'})^{-1}$$

La figura 2.4 ilustra la composición de las transformaciones inversas y las respectivas transformaciones de coordenadas.

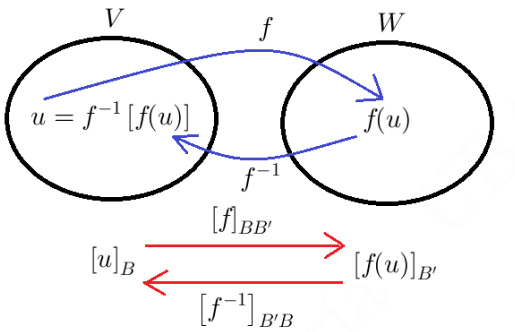


Figura 2.4: Composición de transformaciones inversas

Ejemplo 2.15

Sea f una transformación lineal definida mediante

$$f : \mathbb{R}^3 \to \mathbb{R}^3, f(x_1, x_2, x_3) = (x_1 + x_2, x_1 - x_3, x_3)$$

buscaremos (si es que existe) una fórmula para f^{-1}.

La matriz asociada con respecto a la base canónica de $\mathbb{R}^3$ es

$$A = [f]_E = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Como esta matriz es inversible, podemos asegurar que f es un isomorfismo (¿por qué?). De acuerdo con la proposición 2.8, la inversa de la matriz A será la matriz asociada a f^{-1} con respecto a la base canónica:

$$A^{-1} = ([f]_E)^{-1} = [f^{-1}]_E = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{bmatrix}$$

Luego la fórmula para f^{-1} se obtiene multiplicando un vector genérico (en coordenadas) por $[f^{-1}]_E$:

$$f^{-1} : \mathbb{R}^3 \to \mathbb{R}^3, f^{-1}\left([x_1\ x_2\ x_3]^t\right) = \begin{bmatrix} 0 & 1 & 1 \\ 1 & -1 & -1 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} x_2 + x_3 \\ x_1 - x_2 - x_3 \\ x_3 \end{bmatrix}$$

N A la luz de lo expuesto en la proposición 2.7, podemos ampliar lo estudiado en el teorema 1.14 acerca de la matriz de cambio de una base B a otra C. Ahora podemos interpretar que es la matriz asociada a la transformación identidad, aunque con respecto a las bases B y C. Es decir

$$C_{BC} = [I]_{BC}$$

Terminamos esta sección con un ejemplo en que se estudia cuál es la relación entre dos matrices asociadas a una misma transformación lineal, pero con respecto a bases diferentes. Este ejemplo es importante, porque nos acerca al problema de cuál es la representación matricial más adecuada: es decir, cuáles bases son más convenientes.

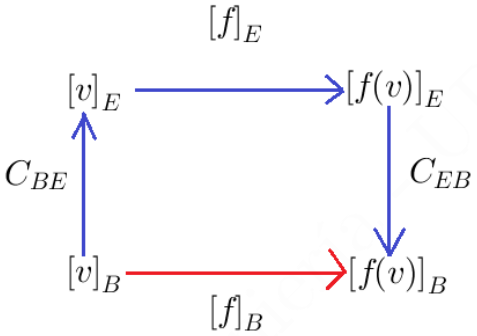


Figura 2.5: Cambio de representación matricial

Ejemplo 2.16

Sea $f : \mathbb{R}^2 \to \mathbb{R}^2$ una transformación lineal cuya matriz asociada con respecto a la base canónica es

$$[f]_E = \begin{bmatrix} 1 & 3 \\ 2 & 2 \end{bmatrix}$$

y buscamos $[f]_B$, siendo

$$B = \{(1,1), (3,-2)\}$$

Para ello, consideramos el esquema de la figura 2.5 donde se sugiere que realizar la transformación en coordenadas con respecto a la base B es equivalente a cambiar a coordenadas canónicas, transformar y volver a cambiar a coordenadas en base B. Teniendo en cuenta el orden en que la multiplicación de matrices representa la composición, es

$$[f]_B = C_{EB}\,[f]_E\,C_{BE} \tag{2.8}$$

De las dos matrices de cambio, la más fácil de construir es la que cambia de base B a canónica:

$$C_{BE} = \begin{bmatrix} 1 & 3 \\ 1 & -2 \end{bmatrix}$$

de donde obtenemos

$$C_{EB} = (C_{BE})^{-1} = \frac{1}{5}\begin{bmatrix} 2 & 3 \\ 1 & -1 \end{bmatrix}$$

Al realizar el cálculo planteado en (2.8) llegamos al resultado

$$[f]_B = \begin{bmatrix} 4 & 0 \\ 0 & 1 \end{bmatrix}$$

Notablemente, la matriz asociada a f con respecto a la base B es diagonal y puede simplificar muchos cálculos elegir esta base en lugar de la canónica.

Esta página fue dejada intencionalmente en blanco

3 Producto Interno

3.1 Introducción

Al comienzo de estas notas se definió el concepto de espacio vectorial basándose en algunas propiedades fundamentales que verifican los vectores ya conocidos de $\mathbb{R}^2$ y $\mathbb{R}^3$. Siguiendo en esta línea de argumentación, señalaremos algunas propiedades que tiene el producto escalar para los vectores del plano. Esto nos permitirá introducir una noción de producto interno en espacios vectoriales generales, a partir de la cual surgirán los conceptos geométricos de longitud, distancia, ángulo...

En el espacio vectorial $\mathbb{R}^2$ el producto interno (o escalar) se define para dos vectores cualesquiera mediante la fórmula

$$(x_1, x_2) \cdot (y_1, y_2) = x_1 y_1 + x_2 y_2 \tag{3.1}$$

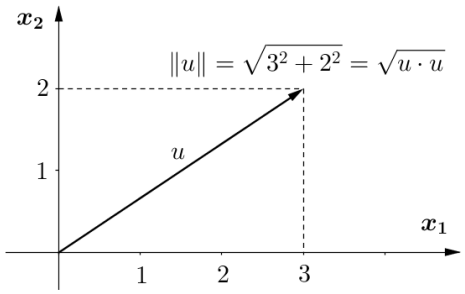


Figura 3.1: Longitud de vectores en $\mathbb{R}^2$

En la interpretación geométrica usual de los vectores de $\mathbb{R}^2$ como segmentos orientados ("flechas"), el teorema de Pitágoras adquiere una expresión sencilla en términos de las componentes de cada vector $u = (x_1, x_2)$ y su norma (o longitud) $\|u\|$:

$$\|u\| = \sqrt{x_1^2 + x_2^2}$$

La norma de los vectores puede también escribirse usando el producto escalar

$$\|u\| = \sqrt{u \cdot u} \tag{3.2}$$

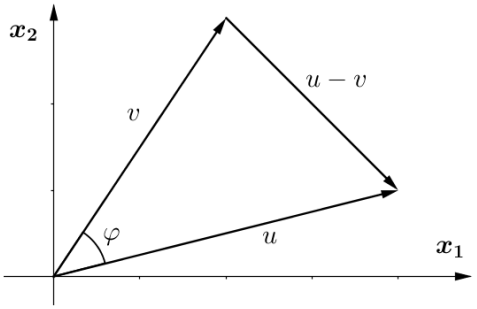


Figura 3.2: Teorema del coseno en $\mathbb{R}^2$

Es conocido en la geometría plana el teorema del coseno [1], que cuando se aplica a dos vectores y al triángulo que determinan adquiere la expresión

$$\|u - v\|^2 = \|u\|^2 + \|v\|^2 - 2\|u\|\,\|v\|\cos\varphi \tag{3.3}$$

Podemos despejar una fórmula que permite calcular el coseno del ángulo que forman dos vectores (no nulos) y a partir de esto obtener el ángulo. Para ello, destacamos tres propiedades que tiene el producto escalar de $\mathbb{R}^2$, cuya comprobación dejamos como ejercicio. Dados los vectores $u, v, w \in \mathbb{R}^2$ y dado $k \in \mathbb{R}$, se cumple que

$$\begin{aligned} u \cdot (v + w) &= u \cdot v + u \cdot w \\ u \cdot v &= v \cdot u \\ (ku) \cdot v &= k\,(u \cdot v) \end{aligned}$$

Considerando la relación (3.2), desarrollamos la expresión de la norma y aplicamos las propiedades precedentes:

$$\begin{aligned} \|u - v\|^2 &= (u - v) \cdot (u - v) \\ &= u \cdot u + u \cdot (-v) + (-v) \cdot u + (-v) \cdot (-v) \\ &= \|u\|^2 - u \cdot v - v \cdot u + v \cdot v \\ &= \|u\|^2 - u \cdot v - u \cdot v + \|v\|^2 \end{aligned}$$

Al reemplazar en la fórmula (3.3) y cancelar, se obtiene una expresión para el coseno del ángulo

$$\cos\varphi = \frac{u \cdot v}{\|u\|\ \|v\|} \tag{3.4}$$

La definición que sigue nos permitirá extender las nociones de norma y ángulo a espacios mucho más generales. Para ello, consideraremos como producto interno a una operación entre vectores que verifique propiedades como las ya señaladas en el producto de $\mathbb{R}^2$.

[1] En todo triángulo el cuadrado de la longitud de un lado es igual a la suma de los cuadrados de las longitudes de los otros dos lados, menos el doble del producto de dichas longitudes multiplicado por el coseno del ángulo que forman.

Definición 3.1

Un ***producto interno*** en un espacio vectorial real V es una función que a cada par de vectores $u, v \in V$ le asigna un número real (u, v), de modo tal que se verifican las siguientes condiciones para cualesquiera $u, v, w \in V$ y $k \in \mathbb{R}$:

1. $(u, v+w) = (u, v) + (u, w)$
2. $(ku, v) = k(u, v)$
3. $(u, v) = (v, u)$
4. $(u, u) \geq 0 \ \wedge \ (u, u) = 0 \Leftrightarrow u = 0$

Ejemplo 3.1 producto interno canónico de $\mathbb{R}^n$

El producto escalar definido en $\mathbb{R}^2$ mediante (3.1) se extiende naturalmente al espacio $\mathbb{R}^n$. El *producto interno canónico en* $\mathbb{R}^n$ de dos vectores arbitrarios $u = (x_1, ..., x_n)$ y $v = (y_1, ..., y_n)$ es

$$(u, v) = x_1 y_1 + ... + x_n y_n = \sum_{i=1}^{n} x_i y_i$$

La comprobación de las condiciones exigidas en la definición 3.1 de producto interno es inmediata, y la dejamos como ejercicio.

Teniendo en cuenta la identificación que hacemos entre las n-uplas de $\mathbb{R}^n$ y los vectores columna de $\mathbb{R}^{n\times 1}$, el producto interno canónico puede escribirse como el producto entre una matriz fila y una matriz columna:

$$(u, v) = u^t v \tag{3.5}$$

(donde el resultado es una matriz de tamaño 1×1 que identificamos con un escalar). Haremos amplio uso de esta notación a lo largo del texto.

La fórmula (3.5) se ha definido para matrices columna y resume el procedimiento de multiplicar las componentes respectivas y sumar todo. En el ejemplo que sigue lo extendemos a matrices de 2×2.

Ejemplo 3.2

En el espacio de matrices cuadradas $\mathbb{R}^{2\times 2}$ definimos un producto interno considerando para $A = [a_{ij}]$ y $B = [b_{ij}]$

$$(A, B) = a_{11}b_{11} + a_{21}b_{21} + a_{12}b_{12} + a_{22}b_{22} = \sum_{1 \leq i,j \leq 2} a_{ij} b_{ij}$$

La comprobación de que esta fórmula define un producto interno es inmediata y la dejamos como ejercicio.
Una forma de reescribir este producto consiste en inspeccionar el producto

$$A^t B = \begin{bmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{bmatrix} \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} a_{11}b_{11} + a_{21}b_{21} & a_{11}b_{12} + a_{21}b_{22} \\ a_{12}b_{11} + a_{22}b_{21} & a_{12}b_{12} + a_{22}b_{22} \end{bmatrix}$$

Notamos que en la diagonal se encuentran los productos entre los elementos respectivos de A y B. Entonces sólo falta

sumarlos para escribir el producto interno. Para ello calculamos la *traza*: [a]

$$\operatorname{tr}\left(A^t B\right) = a_{11}b_{11} + a_{21}b_{21} + a_{12}b_{12} + a_{22}b_{22}$$

[a] Recordemos que la traza de una matriz cuadrada se define como la suma de los elementos de su diagonal principal. La denotaremos tr.

Ejercicio.

3.1 En el espacio de matrices $\mathbb{R}^{m\times n}$ se define

$$(A, B) = \operatorname{tr}\left(A^t B\right)$$

Demuestre que es un producto interno.

Ejemplo 3.3

En el espacio $\mathscr{C}([a,b])$ de las funciones continuas en un intervalo, podemos proponer un producto interno obrando por analogía con el producto canónico de $\mathbb{R}^n$. Para dos funciones cualesquiera $f, g \in \mathscr{C}([a,b])$ definimos entonces:

$$(f,g) = \int_a^b f(t)g(t)\,dt \tag{3.6}$$

En primer lugar, señalamos que este producto está bien definido, en el sentido de que siempre puede puede calcularse para dos elementos cualesquiera de $\mathscr{C}([a,b])$. En efecto, por ser un espacio de funciones continuas, el producto será una función continua y siempre puede calcularse su integral definida en un intervalo cerrado y acotado. Falta ver que se cumplen las condiciones de la definición 3.1. Las condiciones ❶, ❷ y ❸ son una consecuencia inmediata de las propiedades de la integral y dejamos su comprobación formal como ejercicio. La condición ❹ exige probar que para cualquier función continua f

$$\int_a^b \left[f(t)\right]^2 dt \geq 0 \;\wedge\; \int_a^b \left[f(t)\right]^2 dt = 0 \Leftrightarrow f = 0$$

(donde $f = 0$ debe interpretarse como que f es la función nula). Señalemos en primer lugar que para cualquier función f es $\left[f(t)\right]^2 \geq 0$ y entonces la primera desigualdad es inmediata. Por otra parte, si una función continua y no negativa tiene su integral nula a lo largo de un intervalo $[a,b]$, es un resultado conocido del análisis que dicha función debe ser la función nula. El argumento para demostrar esto consiste en considerar una función g continua y no negativa en el intervalo: si fuera $g(t) > 0$ para cierto t del intervalo, entonces habría un entorno de dicho punto en que la función es positiva (¿por qué?) y por ende la integral no sería nula.

La fórmula (3.2) vincula, en $\mathbb{R}^2$, a la norma de los vectores con el producto interno interno canónico. Usaremos esta relación para extender la noción de norma.

Definición 3.2

Dado un espacio vectorial V con producto interno, para cada vector $u \in V$ definimos su *norma* como

$$\|u\| = \sqrt{(u,u)} \tag{3.7}$$

Como esta norma depende del producto interno que se haya definido, es usual nombrarla como *norma inducida por el producto interno.*

Ejercicio.

3.2

- Considerando el espacio de funciones continuas $\mathscr{C}([0,1])$ y el producto interno definido por la fórmula (3.6), calcule (f,g) y $\|g\|$ siendo $f(t)=1$ y $g(t)=t$.
- Repita los cálculos en el espacio $\mathscr{C}([-1,1])$.

Con la definición de producto interno que hemos adoptado, es posible definir distintos productos internos en un mismo espacio vectorial, que inducirán normas diferentes. Ilustramos esta afirmación exhibiendo un producto interno en $\mathbb{R}^2$ que no es el canónico.

Ejemplo 3.4

Para dos vectores cualesquiera de $\mathbb{R}^2$, $u=(x_1,x_2)$ y $v=(y_1,y_2)$ definimos el producto interno mediante

$$(u,v)=2x_1y_1+x_1y_2+x_2y_1+x_2y_2 \tag{3.8}$$

Para demostrar que esta fórmula define un producto interno, será de utilidad considerar a los vectores de $\mathbb{R}^2$ como columnas y reescribir la definición como un producto de matrices:

$$(u,v)=\begin{bmatrix} x_1 & x_2 \end{bmatrix}\begin{bmatrix} 2 & 1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

(¡comprobarlo!) Denominando a la matriz de 2×2 que aparece como A, la función definida puede expresarse como

$$(u,v)=u^tAv$$

Con esta última expresión y usando propiedades del producto de matrices, las condiciones 1, 2 y 3 se demuestran con facilidad. En efecto

$$\begin{aligned}(u,v+w)&=u^tA(v+w)=u^t(Av+Aw)=u^tAv+u^tAw=(u,v)+(u,w)\\(ku,v)&=(ku)^tAv=k\left(u^tAv\right)=k(u,v)\\(u,v)&=u^tAv=\left(u^tAv\right)^t=v^tA^tu=v^tAu=(v,u)\end{aligned}$$

En el último renglón, se usó que u^tAv es un escalar y que por ende es igual a su transpuesto, junto con la propiedad de la transposición $(AB)^t=B^tA^t$ y el hecho de que A es en este caso una matriz simétrica.
Para comprobar la condición 4, aplicamos la fórmula (3.8) a (u,u) y completamos cuadrados:

$$\begin{aligned}(u,u)&=2x_1^2+2x_1x_2+x_2^2\\&=x_1^2+x_1^2+2x_1x_2+x_2^2\\&=x_1^2+(x_1+x_2)^2\end{aligned}$$

La última expresión es una suma de cuadrados, de donde deducimos que para todo u es $(u,u)\geq 0$ y

$$(u,u)=0\Leftrightarrow x_1=0\wedge x_1+x_2=0\Leftrightarrow x_1=x_2=0\Leftrightarrow u=0$$

como queríamos.

Evidentemente, al cambiar la definición de producto interno en un espacio vectorial, cambiarán las nociones geométricas definidas a partir del producto interno. Si consideramos por ejemplo al vector de la figura 3.1, la norma inducida

por el producto interno canónico es $\sqrt{3^2+2^2} = \sqrt{13}$. En cambio para el producto definido por (3.8) la nueva norma resulta, según la definición dada en (3.7),

$$\|(3,2)\| = \sqrt{2\cdot 3^2 + 2\cdot 3\cdot 2 + 2^2} = \sqrt{34}$$

En lo que sigue introduciremos una definición de producto interno para $\mathbb{C}$ espacios vectoriales. A primera vista, parece que alcanzaría con copiar la definición dada en 3.1 para $\mathbb{R}$ espacios. Sin embargo, aceptar las condiciones de esa definición produciría contradicciones al permitir escalares complejos. Consideremos por ejemplo un vector no nulo $u \in V$ de un $\mathbb{C}$ espacio. Entonces $(u,u) > 0$. Ahora bien, consideremos el producto (iu, iu), que debería ser positivo de acuerdo con ❹. Sin embargo, al aplicar ❷ y ❸ resulta

$$(iu, iu) = i(u, iu) = i(iu, u) = ii(u,u) = (-1)(u,u) < 0$$

Dejamos como ejercicio para el lector comprobar que la definición siguiente evita el problema expuesto.

Definición 3.3

Un ***producto interno*** en un espacio vectorial complejo V es una función que a cada par de vectores $u, v \in V$ le asigna un número complejo (u,v) de modo tal que se verifican las siguientes condiciones para cualesquiera $u, v, w \in V$ y $z \in \mathbb{C}$:

❶ $(u, v+w) = (u,v) + (u,w)$

❷ $(zu, v) = \overline{z}\,(u,v)$

❸ $(u,v) = \overline{(v,u)}$

❹ $(u,u) \geq 0 \ \wedge \ (u,u) = 0 \Leftrightarrow u = 0$

N Esta última definición abarca a la anterior como un caso particular. Podría enunciarse para un $\mathbb{K}$ espacio vectorial y señalar que en el caso de los $\mathbb{R}$ espacios la conjugación no altera los escalares. Muchos de los enunciados que siguen se enunciarán para "un espacio vectorial con producto interno", dado que son válidos tanto si se trata de un $\mathbb{R}$ espacio como de un $\mathbb{C}$ espacio.

La fórmula (3.7) con que se definió la *norma inducida por el producto interno* se aplica de idéntica manera a los $\mathbb{C}$ espacios. La condición ❹ garantiza que (u,u) es un número real no negativo, para el que está bien definida $\sqrt{(u,u)}$.

N Existen muchos autores que definen la condición ❷ de una forma alternativa. El lector que los consulte deberá hacer las adaptaciones del caso.

Ejemplo 3.5 producto interno canónico de $\mathbb{C}^n$

La definición del *producto interno canónico en* $\mathbb{C}^n$ de dos vectores arbitrarios $u = (z_1, ..., z_n)$ y $v = (w_1, ..., w_n)$ es

$$(u, v) = \overline{z_1} w_1 + ... + \overline{z_n} w_n = \sum_{i=1}^{n} \overline{z_i} w_i \tag{3.9}$$

Considerando matrices columna de $\mathbb{C}^{n\times 1}$, el producto interno canónico puede escribirse como

$$(u, v) = u^H v \tag{3.10}$$

donde ^{H} indica la transposición del vector columna y la conjugación de cada elemento, es decir $u^H = \overline{u^t}$. Esta última expresión permite demostrar con sencillez las condiciones del producto interno:

$$\begin{aligned}
(u, v + w) &= \overline{u^t}(v + w) = \overline{u^t} v + \overline{u^t} w = (u, v) + (u, w) \\
(zu, v) &= \overline{(zu)^t} v = \overline{z}\,\overline{u^t} v = \overline{z}(u, v) \\
(u, v) &= \overline{u^t} v = \left(\overline{u^t} v\right)^t = v^t \overline{u} = \overline{\overline{v^t} u} = \overline{(v, u)} \\
(u, u) &= \overline{u^t} u \\
&= \begin{bmatrix} \overline{z_1} & \cdots & \overline{z_n} \end{bmatrix} \begin{bmatrix} z_1 \\ \vdots \\ z_n \end{bmatrix} \\
&= \sum_{i=1}^{n} \overline{z_i} z_i = \sum_{i=1}^{n} |z_i|^2 \geq 0
\end{aligned}$$

y por tratarse de una suma de cuadrados de módulos, será nula si y sólo si todas las z_i lo son.

De la última expresión se deduce que la norma, definida por la fórmula (3.7), se calcula mediante

$$\|u\| = \sqrt{\sum_{i=1}^{n} |z_i|^2}$$

Ejercicio.

3.3 En el espacio vectorial $\mathbb{C}^2$ con el producto interno canónico definido en (3.9) y (3.10) considere los vectores

$$u = \begin{bmatrix} 1+i \\ -2 \end{bmatrix} \quad v = \begin{bmatrix} -2i \\ 1-3i \end{bmatrix}$$

y calcule (u, v), $\|u\|$, $\|v\|$.

La notación ^{H} para referirse al transpuesto y conjugado de un vector columna puede extenderse a matrices de tamaño arbitrario.

Definición 3.4

Dada una matriz $A \in \mathbb{C}^{m\times n}$, denotamos como A^H a la matriz transpuesta y conjugada $\overline{A^t}$. Es decir, si $A = [a_{ij}] \in \mathbb{C}^{m\times n}$ entonces $A^H = [\overline{a_{ji}}] \in \mathbb{C}^{n\times m}$.

A continuación, presentamos algunas propiedades que se desprenden de la definición de producto interno. Las enunciamos y demostramos para un $\mathbb{K}$ espacio vectorial con producto interno. (En el caso de que sea $\mathbb{K} = \mathbb{R}$, la conjugación no altera los escalares).

Teorema 3.1

Sea V un espacio vectorial con producto interno. Entonces para cualesquiera $u, v, w \in V$ y $\alpha \in \mathbb{K}$ se verifican

1. $(u+v, w) = (u, w) + (v, w)$
2. $(u, \alpha v) = \alpha\,(u, v)$
3. $(u, 0) = (0, u) = 0$

Demostración:
Probamos 1:

$$(u+v, w) = \overline{(w, u+v)} = \overline{(w, u) + (w, v)} = \overline{(w, u)} + \overline{(w, v)} = (u, w) + (v, w)$$

Destacamos que en 2 el escalar de la segunda componente no se conjuga:

$$(u, \alpha v) = \overline{(\alpha v, u)} = \overline{\alpha}\overline{(v, u)} = \alpha\overline{(v, u)} = \alpha(u, v)$$

Dejamos como ejercicio la demostración de 3. (Sugerencia: escribir al vector nulo como $0v$ ó $v - v$). □

Ejercicio.

3.4 Demuestre que en un espacio vectorial V con producto interno la norma inducida verifica la llamada *identidad del paralelogramo*:

$$\|u+v\|^2 + \|u-v\|^2 = 2\,\|u\|^2 + 2\,\|v\|^2$$

para cualesquiera $u, v \in V$. Interprete geométricamente en $\mathbb{R}^2$.

La norma inducida por el producto interno verifica las propiedades que mencionamos en el siguiente teorema. Su demostración es inmediata a partir de la definición de norma y dejamos los detalles como ejercicio. La segunda propiedad se corresponde, en $\mathbb{R}^2$, con la interpretación usual del producto de un escalar por un vector "flecha".

Teorema 3.2

Sea V un espacio vectorial con producto interno y norma inducida $\|\bullet\|$. Entonces para cualesquiera $u \in V$ y $\alpha \in \mathbb{K}$ se verifican

1. $\|u\| \geq 0 \ \wedge\ \|u\| = 0 \Leftrightarrow u = 0$
2. $\|\alpha u\| = |\alpha|\ \|u\|$

El teorema siguiente establece una importante desigualdad que nos permitirá desarrollar el concepto de ángulo entre vectores. Remitimos a los lectores interesados en la demostración a la bibliografía.

Teorema 3.3 desigualdad de Cauchy-Schwarz

Sea V un espacio vectorial con producto interno y norma inducida $\|\bullet\|$. Entonces para cualesquiera $u, v \in V$ se cumple que

$$|(u,v)| \leq \|u\| \, \|v\|$$

Más aún, la igualdad se realiza si y sólo si u y v son linealmente dependientes.

La fórmula (3.4) tiene validez para vectores en el plano. Ahora la usaremos para extender el concepto de ángulo entre vectores en $\mathbb{R}$ espacios arbitrarios. Para ello, notemos que la desigualdad de Cauchy-Schwarz 3.3 garantiza, para dos vectores no nulos *de un* $\mathbb{R}$ *espacio vectorial*, que

$$-1 \leq \frac{(u,v)}{\|u\| \, \|v\|} \leq 1$$

Vale decir que el cociente $\frac{(u,v)}{\|u\| \, \|v\|}$ tiene el mismo rango de valores que la función coseno, lo cual permite la siguiente definición.

Definición 3.5

Sea V un $\mathbb{R}$ espacio vectorial con producto interno y norma inducida $\|\bullet\|$. Entonces para cualesquiera $u, v \in V$ no nulos el *ángulo* que forman se define como el único $\varphi \in [0, \pi]$ que verifica

$$\cos\varphi = \frac{(u,v)}{\|u\| \, \|v\|}$$

Ejemplo 3.6

Consideramos en $\mathbb{R}^2$ los vectores $u = (1,0)$ y $v = (1,1)$.

Si el producto interno es el canónico, el ángulo que forman u y v se deduce de

$$\cos\varphi = \frac{1}{\sqrt{2}} \Rightarrow \varphi = \frac{\pi}{4}$$

Si en cambio consideramos el producto interno del ejemplo 3.1, definido por la fórmula (3.8), obtenemos (comprobarlo)

$$\cos\varphi = \frac{3}{\sqrt{2}\sqrt{5}} \Rightarrow \varphi = \arccos\left(\frac{3}{\sqrt{10}}\right)$$

Convención

De aquí en más, siempre que se mencionen $\mathbb{R}^n$ y $\mathbb{C}^n$ como espacios con producto interno, se entenderá que es el *producto interno canónico* como se lo definió en 3.1 y 3.5 respectivamente, a menos que se indique explícitamente otra cosa.

A continuación presentamos una desigualdad con una interpretación geométrica inmediata para el caso de vectores de $\mathbb{R}^2$.

Teorema 3.4 desigualdad triangular

Sea V un espacio vectorial con producto interno y norma inducida $\|\bullet\|$. Entonces para cualesquiera $u, v \in V$ se verifican

$$\|u+v\| \leq \|u\| + \|v\|$$

Demostración:
Desarrollamos la definición de norma:

$$\|u+v\|^2 = (u+v, u+v) = (u,u) + (u,v) + (v,u) + (v,v)$$

Teniendo en cuenta que $(v,u) = \overline{(u,v)}$ y que para cualquier número complejo z es $z + \overline{z} = 2\operatorname{Re} z$, la expresión anterior queda:

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 + 2\operatorname{Re}(u,v)$$

Por otra parte, para cualquier número complejo z es $(\operatorname{Re} z)^2 \leq (\operatorname{Re} z)^2 + (\operatorname{Im} z)^2 = |z|^2$. Por lo tanto $\operatorname{Re} z \leq |\operatorname{Re} z| \leq |z|$. Con esto podemos escribir

$$\|u+v\|^2 \leq \|u\|^2 + \|v\|^2 + 2|(u,v)|$$

Aplicando ahora la desigualdad de Cauchy Schwarz 3.3 obtenemos

$$\|u+v\|^2 \leq \|u\|^2 + \|v\|^2 + 2\|u\|\ \|v\| = (\|u\| + \|v\|)^2$$

de donde se deduce inmediatamente la desigualdad buscada. □

En la geometría del plano $\mathbb{R}^2$, para calcular la distancia entre dos puntos medimos el vector que ellos determinan. En la figura 3.2, la distancia entre dos vectores u y v es $\|u-v\|$. Generalizamos esta idea:

Definición 3.6

Sea V un espacio vectorial con norma $\|\bullet\|$. Entonces para cualesquiera $u, v \in V$ definimos la *distancia* entre ellos como

$$\mathrm{d}(u,v) = \|u-v\|$$

Ejercicio.

3.5 Demuestre que la distancia verifica las siguientes propiedades para cualesquiera $u, v, w \in V$. Interprete geométricamente en $\mathbb{R}^2$.

- $\mathrm{d}(u,v) \geq 0 \ \wedge\ \mathrm{d}(u,v) = 0 \Leftrightarrow u = v$
- $\mathrm{d}(u,v) = \mathrm{d}(v,u)$
- $\mathrm{d}(u,v) \leq \mathrm{d}(u,w) + \mathrm{d}(w,v)$

De la definición de ángulo dada en 3.5 se desprende que si el ángulo que forman dos vectores es de $\frac{\pi}{2}$, necesariamente su producto interno será nulo, por eso la definición que sigue.

Definición 3.7

Sea V un espacio vectorial con producto interno.

- Dos vectores $u, v \in V$ se dicen *ortogonales* si $(u, v) = 0$. Notación: $u \perp v$.
- Dado un conjunto $S \subset V$, la notación $u \perp S$ indica que u es ortogonal a cada elemento de S. Vale decir

$$u \perp S \Leftrightarrow \forall s \in S : (u, s) = 0$$

Ejercicio.

- Dé un ejemplo de vectores de $\mathbb{R}^2$ que sean ortogonales para el producto interno canónico pero no para el producto interno definido en (3.8).
- Dé un ejemplo de vectores de $\mathbb{R}^2$ que sean ortogonales para el producto interno definido en (3.8) pero no para el producto interno canónico.
- ¿Existe un par de vectores de $\mathbb{R}^2$ no nulos que sean ortogonales para ambos productos internos?

El concepto de ortogonalidad nos permite extender el teorema de Pitágoras de la geometría plana a los espacios con producto interno.

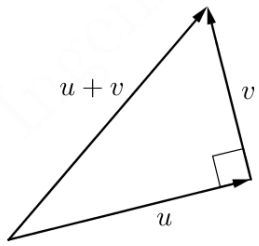


Figura 3.3: Teorema de Pitágoras

Teorema 3.5 de Pitágoras

Sea V un espacio vectorial con producto interno y norma inducida $\|\bullet\|$. Entonces para cualesquiera $u, v \in V$ tales que $u \perp v$ resulta

$$\|u + v\|^2 = \|u\|^2 + \|v\|^2$$

La demostración consiste en desarrollar la definición de norma y aplicar el hecho de que $(u, v) = 0$. Dejamos los detalles como ejercicio.

¿Es válida la implicación recíproca del teorema? En el caso de la geometría plana el enunciado recíproco es: "si en un triángulo el cuadrado del lado mayor es igual a la suma de los cuadrados de los otros dos lados, entonces el triángulo es rectángulo".

Ejercicio.

3.6 Demuestre que *en un* $\mathbb{R}$ *espacio* con producto interno, si dos vectores $u, v \in V$ verifican la relación $\|u + v\|^2 = \|u\|^2 + \|v\|^2$, entonces $u \perp v$.

Entre las propiedades que tienen las bases canónicas, tanto de $\mathbb{R}^n$ como de $\mathbb{C}^n$, destacamos que sus vectores son ortogonales y de norma uno. En lo que sigue extenderemos estas nociones a conjuntos más generales de vectores.

Definición 3.8

Sea V un espacio vectorial con producto interno. El conjunto de vectores $\{v_1, v_2, ..., v_r\} \subset V$ se dice un *conjunto ortogonal* si sus elementos son ortogonales de a pares. Esto es, si para todo $i \neq j$

$$\left(v_i, v_j\right) = 0$$

Ejemplo 3.7

En $\mathbb{R}^3$, el conjunto

$$\{(1,1,-1),(1,-1,0),(1,1,2),(0,0,0)\}$$

es ortogonal, como se puede comprobar por un cálculo directo.

Ejercicio. En $\mathscr{C}([-1,1])$, con el producto definido por la fórmula (3.6)

$$(f,g) = \int_{-1}^{1} f(t)g(t)dt$$

muestre que el conjunto $\left\{1, t, t^2 - \frac{1}{3}\right\}$ es ortogonal.

Teorema 3.6

Sea V un espacio vectorial con producto interno. Si el conjunto de vectores $\{v_1, v_2, ..., v_r\} \subset V$ es ortogonal y ninguno de sus elementos es el vector nulo, entonces es un conjunto linealmente independiente.

Demostración:
Buscamos una combinación lineal de los vectores que dé el cero:

$$\alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_r v_r = 0$$

A continuación, hacemos el producto con un vector genérico v_i donde i puede ser cualquier valor entre 1 y r. Usando las propiedades del producto interno y teniendo en cuenta que v_i es ortogonal al resto de los vectores del conjunto:

$$\left(v_i, \alpha_1 v_1 + ... + \alpha_i v_i + ... + \alpha_r v_r\right) = (v_i, 0)$$
$$\alpha_1 \underbrace{(v_i, v_1)}_{0} + ... + \alpha_i (v_i, v_i) + ... + \alpha_r \underbrace{(v_i, v_r)}_{0} = 0$$

Queda entonces

$$\alpha_i (v_i, v_i) = 0$$

y teniendo en cuenta que $(v_i, v_i) > 0$ por ser v_i no nulo, deducimos que $\alpha_i = 0$ para todo $i = 1, ..., r$. □

Ejemplo 3.8

Destacamos que el teorema anterior es válido *para cualquier producto interno* definido en el espacio vectorial. Si consideramos en $\mathbb{R}^2$ el producto interno definido en el ejemplo 3.4, el conjunto $\{(2,-1),(-1,3)\}$ es ortogonal (¡comprobarlo!), y por el teorema anterior, resulta linealmente independiente.

Finalizamos esta sección agregando un poco de terminología.

Definición 3.9

Sea V un espacio vectorial con producto interno. El conjunto de vectores $\{v_1, v_2, ..., v_r\} \subset V$ se dice un *conjunto ortonormal* si es un conjunto ortogonal y todos sus elementos tienen norma uno.

Ejercicio. Demuestre que todo conjunto ortonormal es linealmente independiente.

Definición 3.10

Sea V un espacio vectorial con producto interno.

- El conjunto de vectores $B = \{v_1, v_2, ..., v_n\}$ se dice una *base ortogonal* si es un conjunto ortogonal y base del espacio.
- El conjunto de vectores $B = \{v_1, v_2, ..., v_n\}$ se dice una *base ortonormal* si es un conjunto ortonormal y base del espacio.

Ejercicio.

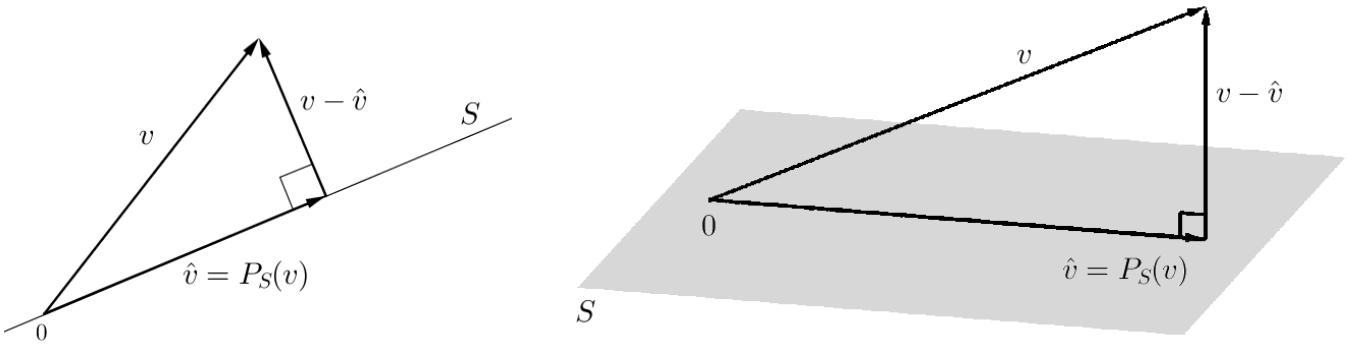
3.7 Demuestre que el conjunto $\{1, t, t^2 - \frac{1}{3}\}$ es una base ortogonal del espacio $\mathscr{P}_2$ considerando el producto interno definido mediante

$$(p,q) = \int_{-1}^{1} p(t)q(t)\,dt$$

¿Es una base ortonormal?

3.2 Proyección ortogonal

Para los casos más conocidos de vectores en un plano o un espacio tridimensional, la proyección de un vector v sobre un subespacio S tiene una interpretación geométrica clara.



(a) Proyección sobre una recta

(b) Proyección sobre un plano

Figura 3.4: Proyecciones ortogonales

Consiste en construir una recta perpendicular al subespacio S, que pase por el extremo del vector v. La intersección de la perpendicular con el subespacio S es el extremo del vector proyección $\hat{v} = P_S(v)$. La proyección así obtenida es el vector del subespacio S más próximo al vector v. En este capítulo presentaremos diversas situaciones en que se busca aproximar un vector con elementos de cierto subespacio. Para resolverlas, necesitaremos generalizar el concepto de proyección.

Notemos que en ambos esquemas de la figura 3.4 el vector $v - \hat{v}$ es ortogonal al subespacio sobre el que se proyecta. Esta propiedad es suficiente para caracterizar a la proyección ortogonal.

Definición 3.11

Sea V un espacio vectorial con producto interno. Dados un subespacio S y un vector $v \in V$, decimos que $\hat{v}$ es la *proyección ortogonal* del vector v sobre el subespacio S si verifica

$$\hat{v} \in S \ \wedge \ v - \hat{v} \perp S$$

Notación: $\hat{v} = P_S(v)$.

Es importante señalar que, en principio, nada garantiza que para cada vector de cierto espacio exista su proyección ortogonal sobre cualquier subespacio. Veremos en primer lugar que si la proyección existe, entonces es única. Más adelante probaremos que en los espacios de dimensión finita siempre es posible encontrar la proyección e indicaremos un método para hacerlo.

Teorema 3.7

Sea V un espacio vectorial con producto interno. Sean $v \in V$ y S un subespacio de V. Entonces si existe $P_S(v)$ es única.

Demostración:
Suponemos que existen dos vectores $\hat{v}_1$ y $\hat{v}_2$ proyección. Demostraremos que $\hat{v}_1 - \hat{v}_2 = 0$ probando que su norma vale cero:

$$\begin{aligned} \|\hat{v}_1 - \hat{v}_2\|^2 &= (\hat{v}_1 - \hat{v}_2, \hat{v}_1 - \hat{v}_2) \\ &= (\hat{v}_1 - \hat{v}_2, \hat{v}_1 - v + v - \hat{v}_2) \\ &= (\hat{v}_1 - \hat{v}_2, \hat{v}_1 - v) + (\hat{v}_1 - \hat{v}_2, v - \hat{v}_2) \end{aligned}$$

Ahora bien, $\hat{v}_1 - \hat{v}_2$ es un elemento de S por serlo $\hat{v}_1$ y $\hat{v}_2$. A su vez $\hat{v}_1$ y $\hat{v}_2$ son proyecciones de v, de modo que $\hat{v}_1 - v$ y $v - \hat{v}_2$ son ortogonales a cualquier elemento de S, en particular a $\hat{v}_1 - \hat{v}_2$. Esto prueba que $\|\hat{v}_1 - \hat{v}_2\|^2 = 0$ como queríamos. □

Ejemplo 3.9

Consideremos en $\mathbb{R}^3$ el subespacio $S = \text{gen}\{(1,1,-2),(1,-1,0)\}$. Buscaremos la proyección $\hat{v}$ del vector $v = (1,0,0)$ sobre S. Adoptando la notación

$$v_1 = (1,1,-2) \quad v_2 = (1,-1,0)$$

y teniendo en cuenta cómo se definió la proyección en 3.11, debemos hallar un vector en S, es decir de la forma $\hat{v} = \alpha v_1 + \beta v_2$, para el cual se cumpla que $v - \hat{v} \perp S$. Esta última exigencia equivale a que $v - \hat{v}$ sea ortogonal a los dos generadores de S (¿por qué?). Comenzamos planteando la ortogonalidad con v_1:

$$0 = (v_1, v - \hat{v}) = \left(v_1, v - \alpha v_1 - \beta v_2\right) = (v_1, v) - \alpha (v_1, v_1) - \beta (v_1, v_2)$$

Teniendo en cuenta que v_1 y v_2 son ortogonales y no nulos, resulta

$$0 = (v_1, v) - \alpha (v_1, v_1) \Rightarrow \alpha = \frac{(v_1, v)}{(v_1, v_1)} = \frac{1}{6}$$

Dejamos como ejercicio comprobar que

$$\beta = \frac{(v_2, v)}{(v_2, v_2)} = \frac{1}{2}$$

con lo cual llegamos a

$$P_S(v) = \hat{v} = \frac{1}{6}(1,1,-2) + \frac{1}{2}(1,-1,0) = \left(\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}\right)$$

Como los coeficientes resultaron únicos, podemos afirmar que en este caso la proyección existe y es única.

El ejemplo precedente señala el camino para calcular la proyección ortogonal sobre un subespacio *siempre que se conozca una base ortogonal* del mismo.

Teorema 3.8

Sea V un espacio vectorial con producto interno. Sea S un subespacio de V con una base ortogonal $B = \{v_1; v_2; ...; v_r\}$. Entonces para cualquier $v \in V$ existe $P_S(v)$ y es

$$P_S(v) = \frac{(v_1, v)}{(v_1, v_1)} v_1 + \frac{(v_2, v)}{(v_2, v_2)} v_2 + ... + \frac{(v_r, v)}{(v_r, v_r)} v_r = \sum_{i=1}^{r} \frac{(v_i, v)}{(v_i, v_i)} v_i \tag{3.11}$$

Demostración:
Buscamos un vector del subespacio S, es decir un vector de la forma $\hat{v} = \alpha_1 v_1 + \alpha_2 v_2 + ... + \alpha_r v_r$, para el cual se verifique la relación $v - \hat{v} \perp S$. Esto último equivale a que $v - \hat{v} \perp v_i$ con $i = 1, ..., r$. Teniendo en cuenta que se trata de una base ortogonal

resulta

$$\begin{aligned} v - \hat{v} \perp v_i &\Leftrightarrow (v_i, v - \hat{v}) = 0 \\ &\Leftrightarrow (v_i, v - \alpha_1 v_1 - \alpha_2 v_2 - \ldots - \alpha_r v_r) = 0 \\ &\Leftrightarrow (v_i, v) - \alpha_1 (v_i, v_1) - \alpha_2 (v_i, v_2) - \ldots - \alpha_r (v_i, v_r) = 0 \\ &\Leftrightarrow (v_i, v) - \alpha_i (v_i, v_i) = 0 \\ &\Leftrightarrow \alpha_i = \frac{(v_i, v)}{(v_i, v_i)} \end{aligned}$$

vale decir que los coeficientes de $\hat{v}$ están unívocamente determinados y conducen a la fórmula (3.11). Cabe señalar que los α_i de la demostración se obtuvieron sin conjugar porque se despejaron de la segunda componente del producto interno, aplicando la propiedad 2 del teorema 3.1. □

N La fórmula (3.11) debe usarse con especial cuidado al aplicarla en $\mathbb{C}$ espacios. En efecto, si se alterara el orden de los coeficientes (v_i, v) por (v, v_i), se obtendrían los conjugados de los valores correctos.

El ejemplo 3.9 es un caso particular de aplicación de la fórmula de proyección (3.11), puesto que se conoce una base ortogonal del espacio sobre el que se pretende proyectar. Planteamos ahora el problema de proyectar sobre un subespacio del que no se conoce una base ortogonal.

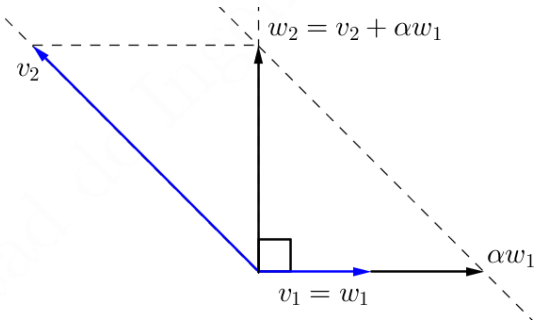


Figura 3.5: Construcción de una base ortogonal

Ejemplo 3.10

Consideremos en $\mathbb{R}^3$ el subespacio $S = \text{gen}\{(1,1,-2),(1,0,1)\}$. Buscaremos la proyección $\hat{v}$ del vector $v = (1,2,0)$ sobre S. Una base del subespacio S es $B = \{v_1, v_2\}$, donde

$$v_1 = (1,1,-2) \quad v_2 = (1,0,1)$$

Dado que no es una base ortogonal, nos proponemos en primer lugar hallar una base $B' = \{w_1, w_2\}$ de S que sí lo sea. Para ello, definimos en primer lugar

$$w_1 = v_1$$

La elección de w_2 no es tan sencilla (de hecho, en el subespacio S existen infinitos vectores ortogonales a v_1). Basándonos en el esquema de la figura 3.5 planteamos que $w_2 = v_2 + \alpha w_1$ es ortogonal a w_1 para una elección adecuada de α:

$$0 = (w_1, v_2 + \alpha w_1) = (w_1, v_2) + \alpha (w_1, w_1) \Rightarrow \alpha = -\frac{(w_1, v_2)}{(w_1, w_1)}$$

Por lo tanto el vector que completa la base ortogonal buscada es

$$w_2 = v_2 - \frac{(w_1, v_2)}{(w_1, w_1)} w_1 \tag{3.12}$$

Dejamos como ejercicio comprobar que la base ortogonal buscada resulta entonces

$$B' = \left\{(1,1,-2); \left(\frac{7}{6}, \frac{1}{6}, \frac{2}{3}\right)\right\}$$

y que a partir de ella la fórmula de proyección (3.11) arroja

$$P_S(v) = \left(\frac{16}{11}, \frac{7}{11}, -\frac{5}{11}\right)$$

Destacamos del ejemplo 3.10 que la fórmula (3.12) para calcular el segundo vector de la base ortogonal consiste en restarle al vector v_2 su proyección sobre el subespacio que genera w_1:

$$w_2 = v_2 - P_{\text{gen}\{w_1\}}(v_2)$$

En efecto, por ser $\text{gen}\{w_1\}$ un espacio de dimensión uno, $\{w_1\}$ es una base ortogonal y se puede aplicar la fórmula de proyección (3.11). Es la misma definición de proyección ortogonal la que garantiza que el vector $v_2 - P_{\text{gen}\{w_1\}}(v_2)$ será ortogonal a $\text{gen}\{w_1\}$, como se aprecia en la figura 3.6. Esta idea nos permitirá demostrar, en el próximo teorema, que en todo espacio de dimensión finita puede construirse una base ortogonal a partir de una base dada.

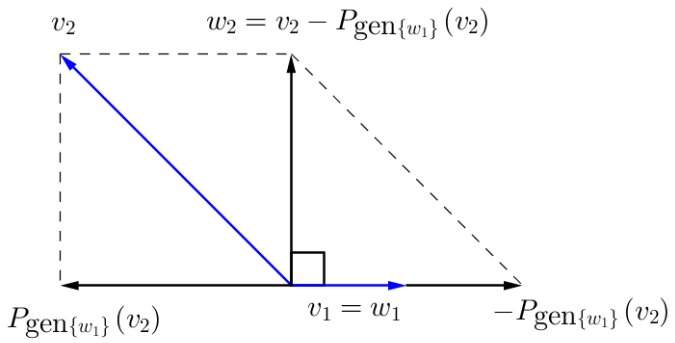


Figura 3.6: Construcción de una base ortogonal usando proyecciones

Teorema 3.9 método de Gram-Schmidt

Sea V un espacio vectorial con producto interno y con una base $B = \{v_1; v_2; ...; v_n\}$. Entonces el conjunto $B' = \{w_1; w_2; ...; w_n\}$ es una base ortogonal de V si sus elementos se definen mediante

$$\begin{aligned} w_1 &= v_1 \\ w_2 &= v_2 - \frac{(w_1, v_2)}{(w_1, w_1)} w_1 = v_2 - P_{\text{gen}\{w_1\}}(v_2) \\ &\vdots \\ w_n &= v_n - \frac{(w_1, v_n)}{(w_1, w_1)} w_1 - ... - \frac{(w_{n-1}, v_n)}{(w_{n-1}, w_{n-1})} w_{n-1} = v_n - P_{\text{gen}\{w_1,...,w_{n-1}\}}(v_n) \end{aligned}$$

Más aun, los subespacios que generan los primeros k vectores de cada base coinciden. Esto es

$$\text{gen}\{v_1, ..., v_k\} = \text{gen}\{w_1, ..., w_k\} \quad \forall k = 1...n$$

Demostración:
En el caso de que la base B tenga un solo vector, es decir $n = 1$, el teorema es inmediatamente verdadero. Para probar el caso general de n vectores, usaremos un argumento de inducción, construyendo la base paso por paso. El primer paso ya esta dado definiendo $w_1 = v_1$. A continuación presentamos una fórmula general para el paso k:

$$w_k = v_k - \frac{(w_1, v_k)}{(w_1, w_1)} w_1 - ... - \frac{(w_{k-1}, v_k)}{(w_{k-1}, w_{k-1})} w_{k-1} \tag{3.13}$$

(Por supuesto, esta fórmula requiere que se hayan construido previamente los vectores $w_1, ..., w_{k-1}$). Así, por ejemplo, la fórmula para el caso $k = 2$ resulta

$$w_2 = v_2 - \frac{(w_1, v_2)}{(w_1, w_1)} w_1$$

Dado que $\{w_1\}$ es una base ortogonal de $\text{gen}\{w_1\}$ podemos afirmar que se ha definido

$$w_2 = v_2 - P_{\text{gen}\{w_1\}}(v_2)$$

y la definición de proyección garantiza que w_2 será ortogonal a $\text{gen}\{w_1\}$. Por lo tanto ya tenemos un conjunto ortogonal $\{w_1, w_2\}$. Además, ninguno de sus vectores es nulo: $w_1 = v_1$ no lo es por ser elemento de una base, mientras que si w_2 fuera nulo resultaría v_2 un elemento de $\text{gen}\{w_1\}$, lo cual es absurdo (¿por qué?). Por el teorema 3.6, el conjunto $\{w_1, w_2\}$ resulta linealmente independiente. Dado que sus elementos son combinaciones lineales de v_1 y v_2, genera el mismo subespacio que $\{v_1, v_2\}$. Este argumento se puede repetir con $w_3, w_4, ...$. Para garantizar que se puede repetir hasta cubrir los n vectores, supondremos que lo hemos aplicado $k-1$ veces hasta obtener un conjunto ortogonal $\{w_1, w_2, ..., w_{k-1}\}$ que genera el mismo subespacio que $\{v_1, v_2, ..., v_{k-1}\}$. Entonces la fórmula (3.13) equivale a

$$w_k = v_k - P_{\text{gen}\{w_1,...,w_{k-1}\}}(v_k)$$

y una vez más la definición de proyección garantiza que el vector w_k será ortogonal a $\text{gen}\{w_1, ..., w_{k-1}\}$. Razonando como antes, w_k no puede ser nulo porque en tal caso sería v_k un elemento de $\text{gen}\{w_1, ..., w_{k-1}\} = \text{gen}\{v_1, ..., v_{k-1}\}$. Este argumento tiene generalidad: si $k = 2$, muestra cómo podemos construir $\{w_1, w_2\}$ a partir de $\{w_1\}$; si $k = 3$, $\{w_1, w_2, w_3\}$ a partir de $\{w_1, w_2\}$ ... y seguir construyendo paso a paso, hasta completar una base ortogonal. □

N Ya sabemos (por el teorema 1.4) que los espacios finitamente generados admiten una base. Además, hemos mostrado cómo obtener una base ortogonal, cómo realizar la proyección y que dicha proyección es única. Por lo tanto, podemos afirmar que ***en todo espacio vectorial con producto interno, la proyección ortogonal de un vector v sobre un subespacio de dimensión finita existe y es única***.

Una consecuencia inmediata del método de Gram Schmidt es que para cualquier espacio de dimensión finita puede obtenerse una base *ortonormal*. En efecto, basta con aplicar el método, calcular una base ortogonal y luego dividir a cada vector por su norma.

Ejemplo 3.11

En el espacio vectorial $\mathbb{R}^4$ se considera el subespacio S determinado por la base $B = \{v_1; v_2; v_3\}$, siendo

$$v_1 = (1,1,1,-1) \quad v_2 = (2,-1,-1,1) \quad v_3 = (0,2,3,-3)$$

La aplicación del método de Gram Schmidt arroja (comprobarlo) la base ortogonal $B' = \{w_1; w_2; w_3\}$ con

$$w_1 = (1,1,1,-1) \quad w_2 = \left(\frac{9}{4}, -\frac{3}{4}, -\frac{3}{4}, \frac{3}{4}\right) \quad w_3 = \left(0, -\frac{2}{3}, \frac{1}{3}, -\frac{1}{3}\right)$$

La base ortonormal buscada es $B'' = \left\{\frac{w_1}{\|w_1\|}; \frac{w_2}{\|w_2\|}; \frac{w_3}{\|w_3\|}\right\} = \{u_1; u_2; u_3\}$:

$$B'' = \left\{\frac{1}{2}(1,1,1,-1); \frac{2}{3\sqrt{3}}\left(\frac{9}{4}, -\frac{3}{4}, -\frac{3}{4}, \frac{3}{4}\right); \sqrt{\frac{3}{2}}\left(0, -\frac{2}{3}, \frac{1}{3}, -\frac{1}{3}\right)\right\}$$

Queremos señalar que cuando se usa una base ortonormal para calcular la proyección de un vector, la fórmula (3.11) adquiere una escritura particularmente sencilla. En nuestro ejemplo, para proyectar un vector $v \in \mathbb{R}^4$ sobre S es

$$P_S(v) = (u_1, v)\, u_1 + (u_2, v)\, u_2 + (u_3, v)\, u_3 = \sum_{i=1}^{3} (u_i, v)\, u_i$$

Dejamos como ejercicio demostrar que $P_S(1,0,1,0) = \left(1, 0, \frac{1}{2}, -\frac{1}{2}\right)$.

Finalizamos esta sección demostrando que la proyección de un vector sobre un subespacio es la mejor aproximación que se puede obtener de dicho vector con elementos del subespacio. La figura 3.7 ilustra esta afirmación: de todos los vectores en el subespacio S, el vector más próximo a v es $P_S(v)$. En este contexto, "la mejor aproximación" se entiende en el sentido de la distancia definida en 3.6. Vale decir que de todos los vectores $u \in S$, el valor de $\|v - u\|$ se hace mínimo si $u = P_S(v)$.

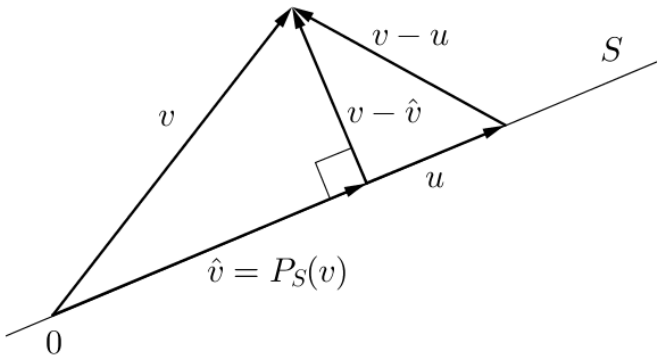


Figura 3.7: Mejor aproximación de v con elementos de S

Formalizamos esta afirmación con el teorema siguiente.

Teorema 3.10

Sean V un espacio vectorial con producto interno y S un subespacio. Sea $v \in V$ para el cual existe la proyección ortogonal $\hat{v} = P_S(v)$. Entonces

$$\|v-\hat{v}\| \leq \|v-u\| \quad \forall u \in S \tag{3.14}$$

Además, la igualdad se da solamente para $u = \hat{v}$.

Demostración:
La figura 3.7 sugiere considerar el triángulo rectángulo de hipotenusa $v-u$. Para ello escribimos

$$v-u = (v-\hat{v}) + (\hat{v}-u)$$

y señalamos que $(v-\hat{v}) \perp S$ mientras que $(\hat{v}-u) \in S$. Por ende se trata de vectores ortogonales y se puede aplicar el teorema de Pitágoras 3.5:

$$\|v-u\|^2 = \|v-\hat{v}\|^2 + \|\hat{v}-u\|^2$$

Por tratarse de cantidades no negativas, es inmediato que

$$\|v-u\|^2 \geq \|v-\hat{v}\|^2$$

y la igualdad se dará si y sólo si $\|\hat{v}-u\| = 0$, es decir si y sólo si $u = \hat{v}$. □

Con esto cobra sentido la siguiente definición.

Definición 3.12

Sean V un espacio vectorial con producto interno y S un subespacio. Sea $v \in V$ para el cual existe la proyección ortogonal $\hat{v} = P_S(v)$. Entonces la *distancia* de v al subespacio S es

$$\mathrm{d}(v,S) = \|v-P_S(v)\|$$

Esta propiedad de la proyección nos permite aproximar funciones eventualmente difíciles de calcular mediante otras más sencillas. El ejemplo siguiente nos brinda una aproximación de la función exponencial mediante una función polinómica de grado uno. La figura 3.8 muestra la recta que mejor aproxima al gráfico de la función exponencial en el intervalo $[-1,1]$.

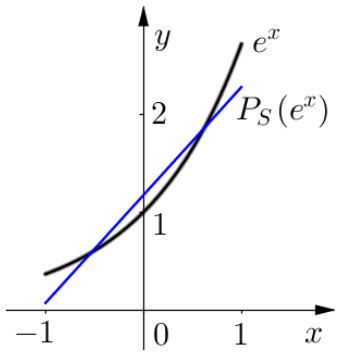


Figura 3.8: Aproximación lineal de e^x

Ejemplo 3.12

En el espacio de funciones continuas $\mathscr{C}([-1,1])$ consideramos el producto interno como se definió en la fórmula (3.6). Buscaremos la proyección de la función $f(x) = e^x$ sobre el subespacio de polinomios $S = \mathscr{P}_1 = \text{gen}\{1, t\}$. Dado que $\{1, t\}$ es una base ortogonal para el producto definido (verificarlo), aplicamos la fórmula de proyección (3.11):

$$\begin{aligned} P_S(f) &= \frac{(1,f)}{(1,1)}1 + \frac{(t,f)}{(t,t)}t \\ &= \frac{1}{2}\left(e - \frac{1}{e}\right) + \frac{3}{e}t \end{aligned}$$

La distancia de la función exponencial al subespacio $\mathscr{P}_1$ es

$$d(f,S) = \|f - P_S(f)\| = \sqrt{1 - \frac{7}{e^2}} \approx 0{,}229$$

Ejercicio.

3.8 En el espacio $\mathbb{R}^3$ considere el subespacio $S = \{x \in \mathbb{R}^3 : x_1 + x_2 + x_3 = 0\}$ y el vector $v = (1,0,0)$. Encuentre el vector de S que mejor aproxima a v y calcule $\text{d}(v,S)$.

3.3 El complemento ortogonal de un subespacio

Ahora que ya hemos dejado asentada la existencia de la proyección ortogonal sobre espacios de dimensión finita, comentaremos algunas propiedades considerando un espacio V con producto interno y un subespacio S de dimensión finita:

1. La proyección ortogonal, entendida como una aplicación que a cada vector $v \in V$ le asigna su proyección $P_S(v)$, es una transformación lineal. Dejamos como ejercicio verificarlo: la comprobación es inmediata usando la fórmula de proyección (3.11) y las propiedades del producto interno; también se puede demostrar aplicando directamente la definición de proyección.

2. Para todo vector $v \in S$ es $P_S(v) = v$. En efecto, es el propio vector v el que verifica la definición de proyección 3.11: $v \in S$ y $v - v \perp S$.

3. Como consecuencia de lo anterior, la composición de proyecciones P_S sobre un mismo subespacio S equivale siempre a la misma proyección.

4. Si un vector v es ortogonal a S, entonces su proyección sobre S es el vector nulo, porque $0 \in S$ y $v - 0 \perp S$. Recíprocamente, si la proyección de un vector v sobre S es el vector nulo, significa que $v - 0 \perp S$.

Resumimos estas observaciones en la siguiente proposición.

Proposición 3.1

Sea V un espacio vectorial con producto interno y sea S un subespacio de dimensión finita. Entonces se cumplen

1. $P_S : V \to V$ es una transformación lineal
2. $P_S(v) = v \Leftrightarrow v \in S$
3. $P_S \circ P_S = P_S$
4. $P_S(v) = 0 \Leftrightarrow v \perp S$

Dado que la proyección es una transformación lineal, prestaremos atención a a su imagen y núcleo.

1. De acuerdo con la definición de proyección, para cualquier $v \in V$ es $P_S(v) \in S$, con lo cual $\operatorname{Im} P_S \subset S$. Usando además la propiedad 2, resulta que el conjunto imagen de la proyección es todo el subespacio S.

2. Según la propiedad 4, el núcleo de la proyección P_S está constituido por los vectores que son ortogonales a todo elemento de S.

Reunimos estas observaciones en una proposición.

Proposición 3.2

Sea V un espacio vectorial con producto interno y sea S un subespacio de dimensión finita. Entonces la proyección ortogonal $P_S : V \to V$ es una transformación lineal que verifica

1. $\operatorname{Im} P_S = S$
2. $\operatorname{Nu} P_S = \{v \in V : v \perp S\}$

Este conjunto de vectores ortogonales a un subespacio dado será de gran interés (aún para casos de dimensión infinita) y lleva nombre propio.

Definición 3.13

Sea V un espacio vectorial con producto interno. Dado un subespacio S, su *complemento ortogonal* es el conjunto de vectores ortogonales a S. Con la notación $S^\perp$ queda

$$S^\perp = \{v \in V : v \perp S\} = \{v \in V : (v, u) = 0 \ \forall u \in S\}$$

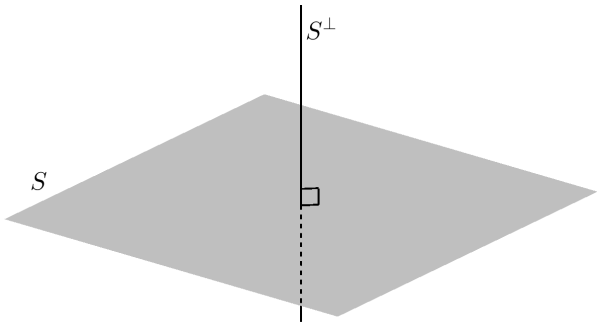


Figura 3.9: Complemento ortogonal de un subespacio S

Ejercicio. Demuestre que en un espacio con producto interno V, dado un subespacio $S = \text{gen}\{v_1, v_2, ..., v_r\}$, un vector arbitrario es ortogonal al subespacio S si y sólo si es ortogonal a sus generadores $v_1, v_2, ..., v_r$. Es decir

$$v \perp S \Leftrightarrow (v, v_1) = (v, v_2) = ... = (v, v_r) = 0$$

Ejercicio. En el espacio $\mathbb{R}^3$

- calcule el complemento ortogonal del subespacio $\text{gen}\{(1,2,1),(-1,2,0)\}$
- calcule el complemento ortogonal del subespacio $\text{gen}\{(2,1,-4)\}$

Ejercicio. Dado un espacio vectorial con producto interno V, demuestre que

- Para cualquier subespacio S su complemento ortogonal es un subespacio de V.
- $V^\perp = \{0_V\}$
- $\{0_V\}^\perp = V$

La figura 3.9 sugiere que un espacio puede descomponerse como la suma directa de un subespacio y su complemento ortogonal. Veremos que esto es así *en el caso de subespacios de dimensión finita.*

Teorema 3.11

Sean V un espacio vectorial con producto interno y S un subespacio de dimensión finita. Entonces

$$V = S \oplus S^\perp$$

Demostración:
Por tratarse de un subespacio de dimensión finita, sabemos que existe la proyección ortogonal P_S. Luego escribimos, para cualquier vector $v \in V$

$$v = \underbrace{P_S(v)}_{\in S} + \underbrace{(v - P_S(v))}_{\in S^\perp}$$

con lo cual $V = S + S^\perp$. Para ver que la suma es directa, según el teorema 1.19 basta con probar que la intersección es el vector nulo. Sea entonces un vector $v \in S \cap S^\perp$. Por estar en $S^\perp$, debe ser ortogonal a todo vector de S, en particular a sí mismo.

Luego

$$(v, v) = 0 \Rightarrow \|v\|^2 = 0 \Rightarrow v = 0$$

□

Otra propiedad que sugiere la figura 3.9 es que el complemento del complemento es el subespacio original. Una vez más, esto será cierto *en el caso de subespacios de dimensión finita.*

Teorema 3.12

Sean V un espacio vectorial con producto interno y S un subespacio de dimensión finita. Entonces

$$\left(S^{\perp}\right)^{\perp} = S$$

Demostración:
Probaremos la doble inclusión de los conjuntos. En primer lugar, consideramos un vector $v \in S$. Por definición de complemento ortogonal

$$\left(S^{\perp}\right)^{\perp} = \left\{v \in V : (v, u) = 0 \ \forall u \in S^{\perp}\right\}$$

Dado que $v \in S$ verifica esta definición, concluimos que $S \subset \left(S^{\perp}\right)^{\perp}$ (notemos que esta inclusión es válida sin necesidad de imponer ninguna restricción sobre la dimensión del subespacio S).

Para demostrar la otra inclusión, consideramos un vector $v \in \left(S^{\perp}\right)^{\perp}$. Aplicando el teorema 3.11, podemos descomponer al espacio como $V = S \oplus S^{\perp}$. Vale decir que existen únicos $v_1 \in S$ y $v_2 \in S^{\perp}$ tales que $v = v_1 + v_2$. Probaremos que $v_2 = 0$ y que por lo tanto $v = v_1 \in S$, lo cual demostrará la inclusión $\left(S^{\perp}\right)^{\perp} \subset S$. Para esto, señalamos que $(v, v_2) = 0$, puesto que $v \in \left(S^{\perp}\right)^{\perp}$ y $v_2 \in S^{\perp}$. Luego

$$0 = (v, v_2) = (v_1 + v_2, v_2) = (v_1, v_2) + (v_2, v_2) = 0 + \|v_2\|^2$$

de donde $v_2 = 0$ como queríamos demostrar. □

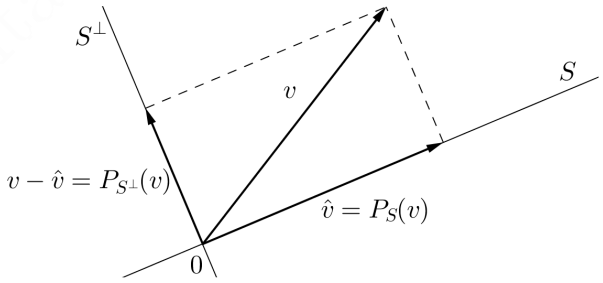


Figura 3.10: Descomposicion de un vector en sus componentes de S y $S^{\perp}$

De acuerdo con los teoremas anteriores, todo vector $v \in V$ puede descomponerse (de manera única) como la suma de un vector en S y otro en $S^{\perp}$: $v = P_S(v) + (v - P_S(v))$. Afirmamos que $(v - P_S(v)) = P_{S^{\perp}}(v)$. En efecto, verifica la definición de proyección sobre $S^{\perp}$:

- $(v - P_S(v)) \in S^{\perp}$ (por definición de P_S)
- $v - (v - P_S(v)) = P_S(v) \in S = \left(S^{\perp}\right)^{\perp}$

Con esto hemos probado el siguiente teorema.

Teorema 3.13

Sean V un espacio vectorial con producto interno y S un subespacio de dimensión finita. Entonces para cualquier $v \in V$ es

$$v = P_S(v) + P_{S^\perp}(v)$$

N Existen ocasiones en que la proyección sobre un subespacio es más dificultosa que la proyección sobre su complemento ortogonal. Como sugiere la figura 3.10, puede ser conveniente entonces calcular la proyección sobre el complemento ortogonal y luego despejar la proyección buscada. Para calcular entonces la distancia de un vector v al subespacio S como se definió en 3.12, se puede usar que $\|v - P_S(v)\| = \left\|P_{S^\perp}(v)\right\|$.

Ejemplo 3.13

En $\mathbb{R}^4$ consideramos el subespacio $S = \{x \in \mathbb{R}^4 : x_1 + x_2 + x_3 - x_4 = 0\}$ de dimensión 3. Para calcular la proyección del vector $v = [1\ 1\ 2\ -1]^t$ sobre S habría que encontrar una base ortogonal de S, presumiblemente aplicando el método de Gram Schmidt, para recién entonces usar la fórmula de proyección (3.11). Si en cambio consideramos que $S^\perp$ es de dimensión 1, alcanzará con encontrar un generador: será inmediatamente una base ortogonal con la que podremos realizar la proyección. Para obtener $S^\perp$ reescribimos la definición de S:

$$S = \left\{ x \in \mathbb{R}^4 : \begin{bmatrix} 1 & 1 & 1 & -1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix} = 0 \right\}$$

Con esto se pone en evidencia que, considerando la matriz

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

resulta que los vectores de S son aquellos vectores ortogonales a las filas de A, es decir que $S = (\text{Fil}\, A)^\perp$. Aplicando el teorema 3.12 obtenemos

$$S^\perp = \text{Fil}\, A = \text{gen}\{[1\ 1\ 1\ -1]^t\}$$

(Recordemos que en $\mathbb{R}^3$ esto permite detectar el vector normal a un plano). La proyección de v sobre $S^\perp$ es, aplicando la fórmula (3.11):

$$P_{S^\perp}(v) = \left[\frac{5}{4}\ \frac{5}{4}\ \frac{5}{4}\ -\frac{5}{4}\right]^t$$

con lo cual

$$P_S(v) = v - P_{S^\perp}(v) = \left[-\frac{1}{4}\ -\frac{1}{4}\ \frac{3}{4}\ \frac{1}{4}\right]^t$$

y

$$\text{d}(v, S) = \|v - P_S(v)\| = \left\|P_{S^\perp}(v)\right\| = \frac{5}{2}$$

Ejemplo 3.14

Calculamos en $\mathbb{R}^3$ el complemento ortogonal del subespacio

$$S = \text{gen}\left\{\begin{bmatrix} 1 \\ 1 \\ -1 \end{bmatrix}, \begin{bmatrix} 2 \\ 2 \\ 0 \end{bmatrix}\right\}$$

Bastará con encontrar los vectores $v = [x_1\ x_2\ x_3]^t$ ortogonales a los generadores:

$$\begin{aligned} S^\perp &= \{v \in \mathbb{R}^3 : x_1 + x_2 - x_3 = 0 \ \wedge\ 2x_1 + 2x_2 = 0\} \\ &= \left\{ v \in \mathbb{R}^3 : \begin{bmatrix} 1 & 1 & -1 \\ 2 & 2 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\} \end{aligned}$$

Vale decir que considerando la matriz

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

el complemento ortogonal de S es el espacio nulo de la matriz A, formado por los vectores ortogonales a sus filas:

$$S^\perp = \text{Nul}\, A = (\text{Fil}\, A)^\perp = \text{gen}\left\{\begin{bmatrix} 1 \\ -1 \\ 0 \end{bmatrix}\right\}$$

Las observaciones hechas en los ejemplos acerca de los espacios fundamentales de una matriz se pueden extender sin dificultades al caso general. Dejamos los detalles como ejercicio.

Proposición 3.3

Dada una matriz $A \in \mathbb{R}^{m\times n}$, los subespacios fundamentales verifican las siguientes relaciones (en los espacios $\mathbb{R}^m$ o $\mathbb{R}^n$ según corresponda):

1. $(\text{Fil}\, A)^\perp = \text{Nul}\, A$
2. $(\text{Col}\, A)^\perp = \text{Nul}\, A^t$

Ejercicio.

3.9 Adaptar la proposición anterior para el caso de matrices de $\mathbb{C}^{m\times n}$.

3.4 Descomposición QR

En lo que sigue veremos que toda matriz $A \in \mathbb{K}^{m\times n}$ con rango $A = n$, es decir cuyas columnas forman un conjunto linealmente independiente, se puede factorizar como producto de dos matrices, una de ellas de $m \times n$ tal que sus columnas forman un conjunto ortonormal y otra de $n \times n$ que es triangular superior y con números positivos en la diagonal (y por tanto inversible).

Tal factorización, que se denomina *descomposición QR* de A, es utilizada por varios algoritmos numéricos para la resolución de problemas diversos, incluyendo sistemas de ecuaciones lineales.

Definición 3.14

Dada $A \in \mathbb{K}^{m\times n}$ con rango $A = n$, una *descomposición QR* es una factorización

$$A = QR \quad \text{con } Q \in \mathbb{K}^{m\times n} \text{ y } R \in \mathbb{K}^{n\times n}$$

tales que $Q^H Q = I$ y R es triangular superior con números positivos en la diagonal principal.

En principio, la definición no garantiza que pueda obtenerse la descomposición QR de cualquier matriz, ni cómo calcularla efectivamente. Veremos que cualquier matriz $A \in \mathbb{K}^{m\times n}$ con rango $A = n$ admite una descomposición QR y mostraremos un método para calcularla. Para ello, comenzamos con un teorema que brinda algunas características de la descomposición.

Teorema 3.14

Sea $A \in \mathbb{K}^{m\times n}$ con rango $A = n$ tal que $A = QR$ es una descomposición QR. Entonces las columnas de Q forman una base ortonormal de Col A.

Demostración:
Denotamos como q_i a la i-ésima columna de Q. De acuerdo con la definición 3.14

$$I = Q^H Q = \begin{bmatrix} q_1^H \\ q_2^H \\ \vdots \\ q_n^H \end{bmatrix} \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} = \begin{bmatrix} q_1^H q_1 & q_1^H q_2 & \cdots & q_1^H q_n \\ q_2^H q_1 & q_2^H q_2 & \cdots & q_2^H q_n \\ \vdots & \vdots & \ddots & \vdots \\ q_n^H q_1 & q_n^H q_2 & \cdots & q_n^H q_n \end{bmatrix}$$

Al igualar esta última matriz con la identidad concluimos que

$$q_i^H q_j = \begin{cases} 0 & \text{si } i \neq j \\ 1 & \text{si } i = j \end{cases}$$

de modo que $\{q_1, q_2, ..., q_n\}$ es un conjunto ortonormal. Por lo tanto constituye una base de Col Q.

Para terminar, probaremos que Col Q = Col A: dado que ambos subespacios tienen dimensión n, alcanzará con probar una inclusión. Sea entonces $y \in$ Col A, es decir que y es combinación lineal de las columnas de A. Por lo tanto debe existir cierto $x \in \mathbb{K}^n$ tal que $y = Ax$. Usando la descomposición QR, tenemos que $y = QRx = Q(Rx)$. Ahora bien, Rx es un vector de K^n, con lo cual y es una combinación lineal de las columnas de Q. Luego, Col $A \subset$ Col Q. □

Teniendo en cuenta que en una descomposición $A = QR$ la matriz R es triangular superior, podemos afirmar que la primera columna de A es combinación lineal de la primera columna de Q, la segunda columna de A es combinación lineal de las primeras dos columnas de Q y así sucesivamente. Como además las columnas de Q son una base ortonormal, podemos concluir que la descomposición QR explica cómo se relacionan dos bases de un mismo espacio, una de ellas ortonormal. Veremos entonces cómo el método de Gram-Schmidt 3.9 nos permite construir la descomposición.

Ejemplo 3.15

Hallar una descomposición QR de

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

Llamando v_i a la i-ésima columna de A, aplicamos el método de Gram-Schmidt para obtener una base ortogonal $\{w_1; w_2; w_3\}$ de Col A:

$$w_1 = v_1 = \begin{bmatrix} 1 \\ 1 \\ 0 \end{bmatrix}$$

$$w_2 = v_2 - \frac{w_1^H v_2}{w_1^H w_1} w1 = \begin{bmatrix} 1/2 \\ -1/2 \\ 1 \end{bmatrix}$$

$$w_3 = v_3 - \frac{w_1^H v_3}{w_1^H w_1} w1 - \frac{w_2^H v_3}{w_2^H w_2} w2 = \begin{bmatrix} -2/3 \\ 2/3 \\ 2/3 \end{bmatrix}$$

De acuerdo con el teorema 3.14, formamos la matriz Q normalizando a los vectores $\{w_1; w_2; w_3\}$:

$$Q = \begin{bmatrix} \sqrt{2}/2 & \sqrt{6}/6 & -\sqrt{3}/3 \\ \sqrt{2}/2 & -\sqrt{6}/6 & \sqrt{3}/3 \\ 0 & \sqrt{6}/3 & \sqrt{3}/3 \end{bmatrix}$$

Dado que $A = QR$ (probaremos en el próximo teorema que la descomposición efectivamente existe) y por ser $Q^H Q = I$, podemos calcular R

$$A = QR \Rightarrow Q^H A = Q^H QR = R$$

Luego

$$R = Q^H A = \begin{bmatrix} \sqrt{2} & \sqrt{2}/2 & \sqrt{2}/2 \\ 0 & \sqrt{6}/2 & \sqrt{6}/6 \\ 0 & 0 & 2\sqrt{3}/3 \end{bmatrix}$$

Teorema 3.15

Sea $A \in \mathbb{K}^{m \times n}$ con rango $A = n$. Entonces existe una descomposición QR de A.

Demostración:
Denominemos $v_1, v_2, ..., v_n$ a las columnas de A. Por hipótesis, $B = \{v_1; v_2; ...; v_n\}$ es un conjunto linealmente independiente.

Aplicando el método de Gram-Schmidt al conjunto B obtenemos una base ortogonal $\{w_1; w_2; ...; w_n\}$ de Col A que satisface

$$\begin{aligned} w_1 &= v_1 \\ w_2 &= v_2 - \alpha_{12} w_1 \\ w_3 &= v_3 - \alpha_{13} w_1 - \alpha_{23} w_2 \\ &\vdots \\ w_n &= v_n - \alpha_{1n} w_1 - ... - \alpha_{n-1,n} w_{n-1} \end{aligned}$$

donde $\alpha_{ij} = \frac{w_i^H v_j}{w_i^H w_i}$. Despejando cada v_i:

$$\begin{aligned} v_1 &= w_1 \\ v_2 &= \alpha_{12} w_1 + w_2 \\ v_3 &= \alpha_{13} w_1 + \alpha_{23} w_2 + w_3 \\ &\vdots \\ v_n &= \alpha_{1n} w_1 + ... + \alpha_{n-1,n} w_{n-1} + w_n \end{aligned}$$

lo cual puede escribirse en forma matricial

$$A = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} = \begin{bmatrix} w_1 & w_2 & \cdots & w_n \end{bmatrix} \begin{bmatrix} 1 & \alpha_{12} & \alpha_{13} & \cdots & \alpha_{1n} \\ 0 & 1 & \alpha_{23} & \cdots & \alpha_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix}$$

Esta factorización es muy parecida a la buscada. Falta lograr que los vectores columna del primer factor sean de norma uno. Para ello, definimos

$$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \text{ con } q_i = \frac{w_i}{\|w_i\|} \quad (i = 1, ..., n)$$

y modificamos adecuadamente el segundo factor

$$A = \begin{bmatrix} q_1 & q_2 & \cdots & q_n \end{bmatrix} \begin{bmatrix} \|w_1\| & \alpha_{12}\|w_1\| & \alpha_{13}\|w_1\| & \cdots & \alpha_{1n}\|w_1\| \\ 0 & \|w_2\| & \alpha_{23}\|w_2\| & \cdots & \alpha_{2n}\|w_2\| \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \|w_n\| \end{bmatrix}$$

con lo que llegamos a la descomposición buscada. □

N Mencionamos, sin demostración, que en las condiciones de la definición 3.14 la descomposición QR es única.

Ejercicio.

3.10 Calcule la descomposición QR de la matriz $A = \begin{bmatrix} -1 & 0 & 1 \\ 1 & 2 & -1 \\ 0 & 1 & 1 \end{bmatrix}$.

3.5 Cuadrados Mínimos

Dada una matriz $A \in \mathbb{R}^{m\times n}$, consideraremos una vez más la ecuación matricial

$$Ax = b \tag{3.15}$$

donde $b \in \mathbb{R}^m$ y $x \in \mathbb{R}^n$ es la incógnita. En la sección 1.5 habíamos señalado que el sistema es compatible si y sólo si existe una combinación lineal de las columnas de A que resulte ser el vector b. Dicho de otra forma: el sistema es compatible si y sólo si $b \in \operatorname{Col} A$. En las aplicaciones, es muy común que un modelo se describa mediante ecuaciones lineales en la forma de (3.15) y que, debido a múltiples factores como por ejemplo errores de medición, no se pueda obtener un x que verifique exactamente las ecuaciones. En esta sección estudiaremos una alternativa que se presenta para el caso en que el sistema es incompatible pero pretendemos obtener, al menos, una solución aproximada.

Definición 3.15

Dado el sistema de ecuaciones (3.15), el vector $\hat{x} \in \mathbb{R}^n$ es una *solución por cuadrados mínimos* si $A\hat{x}$ minimiza la distancia a b, es decir si

$$\|b - A\hat{x}\| \leq \|b - Ax\| \quad \forall x \in \mathbb{R}^n$$

N Dado que estamos considerando la norma usual, el cálculo de $\|b - Ax\|^2$ es la suma de los cuadrados de las componentes de un vector de $\mathbb{R}^m$. Buscamos $\hat{x}$ para que dicha suma sea mínima: de allí el nombre "cuadrados mínimos".

Ejercicio. Demuestre que en el caso de que el sistema (3.15) sea compatible, cualquier solución en el sentido usual es también solución por cuadrados mínimos.

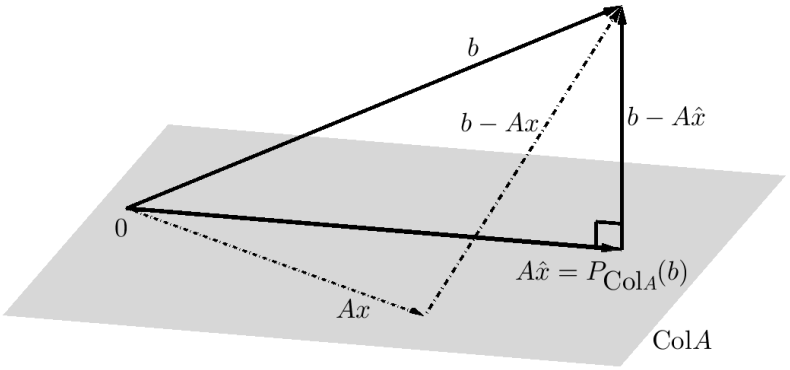


Figura 3.11: Mejor aproximación de b con elementos de Col A

Ejemplo 3.16

Hallaremos la solución por cuadrados mínimos $\hat{x} = [\hat{x}_1\ \hat{x}_2]^t$ del sistema de ecuaciones

$$Ax = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$

La figura 3.11 ilustra este caso, en el que se tiene una matriz $A \in \mathbb{R}^{3\times 2}$, con lo cual $b \in \mathbb{R}^3$ y $x \in \mathbb{R}^2$. Por lo que se ve, el vector b no está en Col A y el sistema es incompatible. Basándonos en el teorema de mejor aproximación 3.10, sabemos que el elemento de Col A más próximo a b es la proyección ortogonal de b sobre Col A. Ahora bien, para conocer la proyección apelamos al ejemplo 3.10, en que proyectábamos un vector $[1\ 2\ 0]^t$ sobre un subespacio S que coincide con Col A. Habíamos obtenido

$$P_{\text{Col}\,A}(b) = \frac{1}{11}\begin{bmatrix} 16 \\ 7 \\ -5 \end{bmatrix}$$

Falta entonces encontrar la solución por cuadrados mínimos $\hat{x}$. *La solución por cuadrados mínimos $\hat{x}$ no debe confundirse con la proyección de b sobre* Col A. El vector $\hat{x}$ es un elemento de $\mathbb{R}^2$ tal que $A\hat{x} = P_{\text{Col}\,A}b$:

$$\begin{bmatrix} 1 & 1 \\ 1 & 0 \\ -2 & 1 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \frac{1}{11}\begin{bmatrix} 16 \\ 7 \\ -5 \end{bmatrix}$$

Este sistema sí es compatible, porque la proyección de b es un elemento de Col A. La solución es

$$\hat{x} = \frac{1}{11}\begin{bmatrix} 7 \\ 9 \end{bmatrix}$$

Siguiendo con la línea de argumentación del ejemplo 3.16, pasamos a considerar el caso general del sistema (3.15). Una vez más, el teorema de mejor aproximación 3.10 indica que el elemento de Col A más próximo a b es la proyección ortogonal de b sobre Col A. Por eso una forma de buscar la solución por cuadrados mínimos es resolver el sistema

$$A\hat{x} = P_{\text{Col}\,A}(b)$$

como ya hicimos en el ejemplo. Destacamos que este sistema es compatible, porque consiste en encontrar una combinación lineal de las columnas de A que genere un elemento de $\text{Col}\,A$. Sin embargo, este procedimiento implica calcular una proyección ortogonal y, eventualmente, obtener una base ortogonal de $\text{Col}\,A$. Presentamos una alternativa más eficaz: para ello, consideramos que

$$A\hat{x} = P_{\text{Col}\,A}(b) \Leftrightarrow b - A\hat{x} \perp \text{Col}\,A \Leftrightarrow b - A\hat{x} \in (\text{Col}\,A)^{\perp}$$

Teniendo en cuenta que $(\text{Col}\,A)^{\perp} = \text{Nul}\,A^t$ (ver la proposición 3.3) resulta

$$A\hat{x} = P_{\text{Col}\,A}(b) \Leftrightarrow b - A\hat{x} \in \text{Nul}\,A^t \Leftrightarrow A^t(b - A\hat{x}) = 0$$

y llegamos a que la ecuación $A^t A\hat{x} = A^t b$ es equivalente al planteo original; con la ventaja de que no hay que calcular explícitamente la proyección. Resumimos el método propuesto en el siguiente teorema.

Teorema 3.16

Dado el sistema de ecuaciones (3.15), el vector $\hat{x} \in \mathbb{R}^n$ es una solución por cuadrados mínimos si y sólo si verifica la *ecuación normal*

$$A^t A\hat{x} = A^t b \tag{3.16}$$

Además, la ecuación normal siempre tiene al menos una solución.

Basándonos en la figura 3.11, podemos definir al error como la distancia entre el vector original y su aproximación con un elemento de $\text{Col}\,A$.

Definición 3.16

Dado el sistema de ecuaciones (3.15) y una solución por cuadrados mínimos $\hat{x} \in \mathbb{R}^n$, el *error por cuadrados mínimos* se define como la distancia entre el vector b y su proyección sobre $\text{Col}\,A$, es decir

$$E = \|b - A\hat{x}\|$$

Ejercicio. Encuentre la solución por cuadrados mínimos del sistema del ejemplo 3.16 planteando la ecuación normal. Calcule el error de la aproximación por cuadrados mínimos.

Llegados a este punto, conviene mencionar una aplicación de la descomposición QR para encontrar soluciones por cuadrados mínimos. Considerando una matriz $A \in \mathbb{K}^{m\times n}$ con $\text{rango}\,A = n$ y descomposición QR como se definió en 3.14, la ecuación normal 3.16 resulta

$$A^t A\hat{x} = A^t b \Leftrightarrow (QR)^t (QR)\hat{x} = (QR)^t b \Leftrightarrow R^t R\hat{x} = R^t Q^t b$$

donde usamos que $Q^t Q = I$. Como además R es una matriz inversible, también lo es R^t (¿por qué?), luego podemos multiplicar a izquierda por $\left(R^t\right)^{-1}$ para llegar a una ecuación equivalente a la ecuación normal

$$R\hat{x} = Q^t b$$

con la ventaja de que es un sistema cuya matriz de coeficientes es triangular superior y reduce la cantidad de cálculos necesarios.

Ejercicio. Encuentre la solución por cuadrados mínimos del sistema del ejemplo 3.16 usando la descomposición QR de la matriz de coeficientes del sistema.

Ejemplo 3.17

Resolvamos el siguiente sistema de ecuaciones

$$A = \begin{bmatrix} 1 & 2 \\ 1 & 2 \\ 1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix}$$

Como visiblemente es un sistema incompatible, planteamos la ecuación normal (3.16):

$$\begin{bmatrix} 3 & 6 \\ 6 & 12 \end{bmatrix} \begin{bmatrix} \hat{x}_1 \\ \hat{x}_2 \end{bmatrix} = \begin{bmatrix} 4 \\ 8 \end{bmatrix}$$

de donde se obtienen infinitas soluciones de la forma

$$\hat{x} = \begin{bmatrix} 4/3 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} -2 \\ 1 \end{bmatrix} \quad \alpha \in \mathbb{R}$$

El producto entre A y *cualquiera* de las soluciones por cuadrados mínimos da la proyección de b sobre Col A:

$$P_{\text{Col}\,A}(b) = A\left(\begin{bmatrix} 4/3 \\ 0 \end{bmatrix} + \alpha \begin{bmatrix} -2 \\ 1 \end{bmatrix} \right) = A \begin{bmatrix} 4/3 \\ 0 \end{bmatrix} = \begin{bmatrix} 4/3 \\ 4/3 \\ 4/3 \end{bmatrix}$$

El error cometido es

$$\left\| \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} - \begin{bmatrix} 4/3 \\ 4/3 \\ 4/3 \end{bmatrix} \right\| = \sqrt{\frac{2}{3}}$$

El último ejemplo pone de manifiesto que la ecuación normal (3.16) no sólo es compatible sino que podría tener infinitas soluciones. Como ya habíamos señalado, la ecuación normal es equivalente a $A\hat{x} = P_{\text{Col}\,A}b$. En esta ecuación es claro que si las columnas de A son linealmente independientes, habrá una única combinación lineal para generar cada elemento de Col A, con lo cual $\hat{x}$ será única. Si en cambio las columnas de A son dependientes, habrá un conjunto $\hat{x}_p + \text{Nul}\,A$ de infinitas soluciones (donde $\hat{x}_p$ es una solución particular).

A continuación probaremos algunos resultados más acerca de las soluciones por cuadrados mínimos. Para ello, necesitamos establecer previamente una propiedad.

Teorema 3.17

Para toda matriz $A \in \mathbb{R}^{m\times n}$ se verifica que

$$\text{Nul}\,A = \text{Nul}\,A^tA$$

y por lo tanto

$$\text{rango}\,A = \text{rango}\,A^tA$$

Demostración:
Probaremos la doble inclusión de los conjuntos Nul A y Nul A^tA. En primer lugar, consideramos

$$x \in \text{Nul}\,A \Rightarrow Ax = 0 \Rightarrow A^tAx = 0 \Rightarrow x \in \text{Nul}\,A^tA$$

Para la segunda inclusión, consideremos un $x \in \text{Nul}\, A^t A$. Entonces $A^t Ax = 0$. Multiplicando a izquierda por x^t obtenemos

$$x^t A^t Ax = 0 \Rightarrow (Ax)^t Ax = 0 \Rightarrow \|Ax\|^2 = 0 \Rightarrow Ax = 0 \Rightarrow x \in \text{Nul}\, A$$

La igualdad de los rangos se prueba teniendo en cuenta que para cualquier matriz $B \in \mathbb{R}^{m\times n}$ vale la relación

$$\dim(\text{Nul}\, B) + \text{rango}\, B = n$$

como consecuencia del teorema de la dimensión 2.3 (justifíquelo). Entonces

$$\dim(\text{Nul}\, A) + \text{rango}\, A = n = \dim(\text{Nul}\, A^t A) + \text{rango}\, A^t A$$

y se deduce que rango A = rango $A^t A$. □

En el caso de que una matriz $A \in \mathbb{R}^{m\times n}$ tenga sus columnas linealmente independientes, la ecuación normal (3.16) tendrá una matriz $A^t A$ de rango n y por lo tanto inversible. En tal caso, la única solución por cuadrados mínimos será

$$\hat{x} = \left(A^t A\right)^{-1} A^t b$$

Más aún, al multiplicar a izquierda por A obtendremos la proyección de b sobre Col A:

$$A\hat{x} = P_{\text{Col}\, A} b = A\left(A^t A\right)^{-1} A^t b$$

con lo cual $A\left(A^t A\right)^{-1} A^t$ es la matriz asociada a la proyección sobre Col A. Agrupamos estas observaciones en un teorema.

Teorema 3.18

Dado el sistema de ecuaciones (3.15), si la matriz $A \in \mathbb{R}^{m\times n}$ tiene sus columnas linealmente independientes, entonces

1. Existe una única solución por cuadrados mínimos $\hat{x}$ y se obtiene como

   $$\hat{x} = \left(A^t A\right)^{-1} A^t b$$

   (A la matriz $A^\sharp = \left(A^t A\right)^{-1} A^t$ se la denomina *matriz seudoinversa* de A.)

2. La proyección ortogonal de b sobre Col A se puede calcular mediante

   $$P_{\text{Col}\, A} b = A\left(A^t A\right)^{-1} A^t b$$

3. La matriz $A\left(A^t A\right)^{-1} A^t = AA^\sharp$ es la matriz asociada a la proyección sobre Col A, con respecto a la base canónica de $\mathbb{R}^m$:

   $$A\left(A^t A\right)^{-1} A^t = AA^\sharp = [P_{\text{Col}\, A}]_E$$

Ejemplo 3.18

Retomamos el ejemplo 3.9. En este caso, buscamos una fórmula para la proyección ortogonal P_S. Por supuesto que podemos usar la fórmula de proyección (3.11). Sin embargo, optaremos por una opción más breve que se desprende del teorema 3.18. Como nuestra intención es proyectar sobre el subespacio S, construiremos una matriz A que tenga por columnas una base de ese mismo subespacio: de acuerdo con 3, la matriz asociada a la proyección sobre Col $A = S$ será

$$[P_S]_E = A\left(A^t A\right)^{-1} A^t = AA^\sharp$$

Ahora bien, *si elegimos la base de modo que sea ortonormal*, será $A^tA = I$ y la expresión anterior se reducirá a

$$[P_S]_E = AA^t$$

En nuestro caso podemos normalizar la base y presentarla como columnas de la matriz A. Con esto, la matriz asociada a la proyección es

$$[P_S]_E = \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{2} \\ 1/\sqrt{6} & -1/\sqrt{2} \\ -2/\sqrt{6} & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix}$$

Para obtener una fórmula:

$$P_S(x) = [P_S]_E \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2 & -1 & -1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \frac{1}{3} \begin{bmatrix} 2x_1 - x_2 - x_3 \\ -x_1 + 2x_2 - x_3 \\ -x_1 - x_2 + 2x_3 \end{bmatrix}$$

El método usado en el ejemplo para obtener la matriz asociada a la proyección se generaliza sin dificultad.

Teorema 3.19

Sea $B = \{u_1; u_2; \ldots; u_r\}$ una *base ortonormal* del subespacio S de $\mathbb{R}^m$. Si $A \in \mathbb{R}^{m\times r}$ tiene por columnas a los vectores de la base B, entonces

$$[P_S]_E = AA^t$$

Ejercicio.

3.11 Considere en $\mathbb{R}^4$ el subespacio $S = \text{gen}\{[1\ 0\ -1\ 0]^t, [0\ 1\ 1\ 0]^t\}$. Calcule $[P_S]_E$.

Terminamos esta sección con un ejemplo de ajuste de datos. Ya mencionamos que un problema usual en las aplicaciones es adaptar una colección de datos a un modelo lineal.

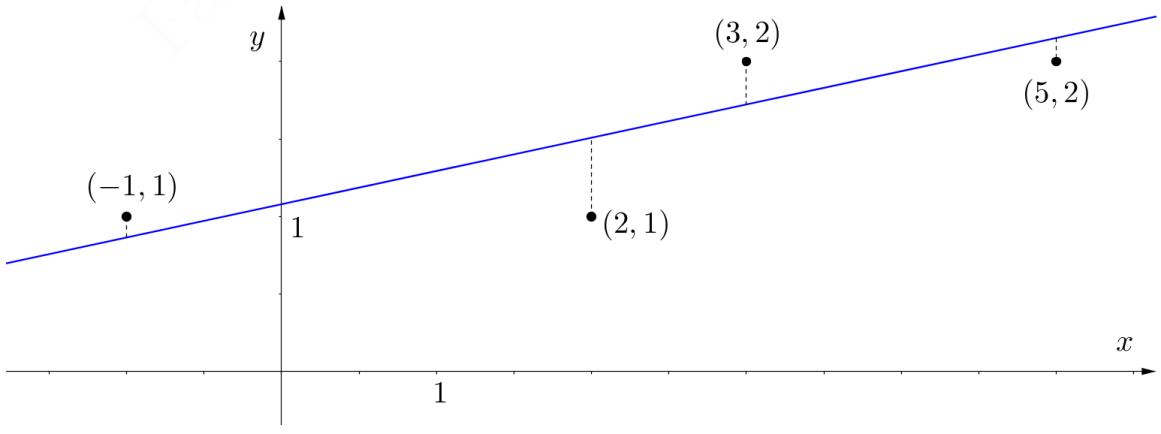


Figura 3.12: Ajuste de datos por cuadrados mínimos

Ejemplo 3.19 Ajuste de datos

Consideremos en $\mathbb{R}^2$ los puntos

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

que podemos interpretar como mediciones de cierto fenómeno. Aunque no están exactamente alineados, buscaremos la recta que "mejor ajuste" los datos en algún sentido. Para eso, consideramos que la forma general de dicha recta será $y = ax + b$ con a, b a determinar. Si hubiera un ajuste exacto, todos los puntos verificarían la ecuación de la recta:

$$\begin{aligned} 1 &= a(-1) + b \\ 1 &= a2 + b \\ 2 &= a3 + b \\ 2 &= a5 + b \end{aligned}$$

lo cual escrito matricialmente queda

$$\underbrace{\begin{bmatrix} -1 & 1 \\ 2 & 1 \\ 3 & 1 \\ 5 & 1 \end{bmatrix}}_{A} \begin{bmatrix} a \\ b \end{bmatrix} = \underbrace{\begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix}}_{b}$$

y para este sistema incompatible buscamos una solución por cuadrados mínimos. Dejamos como ejercicio obtener el resultado

$$\hat{x} = \begin{bmatrix} \hat{a} \\ \hat{b} \end{bmatrix} = \begin{bmatrix} 14/75 \\ 27/25 \end{bmatrix}$$

La recta de ecuación

$$y = \frac{14}{75}x + \frac{27}{25}$$

es entonces la *recta de ajuste por cuadrados mínimos*. Sabemos, por la definición de cuadrados mínimos 3.15, que el valor del error $\|b - A\hat{x}\|$ es mínimo comparado con cualquier otra elección de a, b. En nuestro ejemplo, el cuadrado de este error es

$$\begin{aligned} \|b - A\hat{x}\|^2 &= \left\| \begin{bmatrix} 1 \\ 1 \\ 2 \\ 2 \end{bmatrix} - \begin{bmatrix} 67/75 \\ 109/75 \\ 41/25 \\ 151/75 \end{bmatrix} \right\|^2 \\ &= \left(1 - \frac{67}{75}\right)^2 + \left(1 - \frac{109}{75}\right)^2 + \left(2 - \frac{41}{25}\right)^2 + \left(2 - \frac{151}{75}\right)^2 \\ &= \frac{26}{75} \approx 0{,}347 \end{aligned}$$

vale decir que cada componente del vector $b - A\hat{x}$ mide la diferencia entre la altura verdadera del punto y la que alcanza la recta. Es la suma de los cuadrados de estas diferencias lo que la recta minimiza, como se aprecia en la figura 3.12.

Ejercicio.

3.12 Ajuste los mismos datos del ejemplo 3.19 con una función cuadrática $y = ax^2 + bx + c$.

N Hemos optado por presentar el tema de cuadrados mínimos en el contexto de los $\mathbb{R}$ espacios vectoriales. Sin embargo, la adaptación al caso de $\mathbb{C}$ espacios no presenta mayor dificultad. La ecuación normal (3.16), por ejemplo, debe plantearse como $A^H A\hat{x} = A^H b$.

3.6 Aplicaciones geométricas

Dedicaremos esta sección a algunos aspectos geométricos de las transformaciones lineales y los espacios con producto interno. Aunque muchas de las ideas que expondremos pueden extenderse a espacios más generales, limitaremos nuestra atención al espacio $\mathbb{R}^n$ con el producto interno canónico.

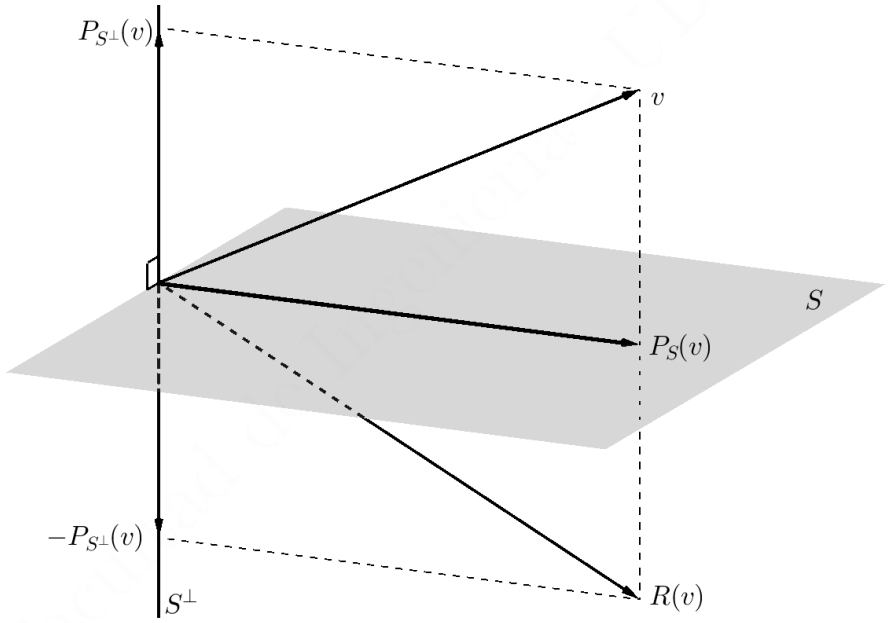


Figura 3.13: Reflexión de un vector con respecto a un subespacio

Consideramos el problema de lograr en una reflexión (simetría) con respecto a un subespacio S. En la figura 3.13 se aprecia un esquema general. Dado un vector v es posible descomponerlo como $v = P_S(v) + P_{S^\perp}(v)$; su reflexión con respecto a S surge de preservar la componente en S e invertir la componente en $S^\perp$.

Definición 3.17

Sea S un subespacio de $\mathbb{R}^n$. La *reflexión* del vector v con respecto al subespacio S es

$$R(v) = P_S(v) - P_{S^\perp}(v) \tag{3.17}$$

Por tratarse de una suma de proyecciones, resulta inmediato que la reflexión $R : \mathbb{R}^n \to \mathbb{R}^n$ es una transformación lineal.

Ejercicio.

3.13 Demuestre que la reflexión de un vector $v \in \mathbb{R}^n$ con respecto a cierto subespacio S verifica

- v y $R(v)$ determinan una recta perpendicular a S.
- v y $R(v)$ determinan un segmento cuyo punto medio es $P_S(v)$.
- v y $R(v)$ equidistan del subespacio de reflexión.

Ejemplo 3.20 Reflexión con respecto a un plano

Consideremos en $\mathbb{R}^3$ el subespacio $S = \text{gen}\{[1\ 1\ -2]^t, [1\ 0\ 1]^t\}$ que ya se estudió en el ejemplo 3.10. Usaremos nuevamente al vector $v = [1\ 2\ 0]^t$, aunque esta vez para calcular su reflexión con respecto al plano S. Como ya habíamos calculado su proyección sobre S resulta

$$v = P_S(v) + (v - P_S(v)) = \left[\frac{16}{11}\ \frac{7}{11}\ -\frac{5}{11}\right]^t + \left[-\frac{5}{11}\ \frac{15}{11}\ \frac{5}{11}\right]^t$$

y de acuerdo con la fórmula (3.17) obtenemos

$$R(v) = P_S(v) - (v - P_S(v)) = \left[\frac{21}{11}\ -\frac{8}{11}\ -\frac{10}{11}\right]^t$$

La complicación del método utilizado es que para calcular $P_S(v)$ hubo que construir una base ortogonal de S. A continuación, exponemos una alternativa que se basa en la proyección sobre $S^\perp$, con la ventaja de que es un subespacio de dimensión uno. Para ello, reescribimos la fórmula (3.17):

$$R(v) = \left(v - P_{S^\perp}(v)\right) - P_{S^\perp}(v) = v - 2P_{S^\perp}(v)$$

Dado que $u = \frac{1}{\sqrt{11}}[-1\ 3\ 1]^t$ constituye una base ortonormal de $S^\perp$, aplicamos la fórmula de proyección del teorema 3.19, según la cual $\left(uu^t\right)$ es la matriz asociada a la proyección y resulta

$$P_{S^\perp}(v) = \left(uu^t\right)v = \frac{1}{11}\begin{bmatrix} 1 & -3 & -1 \\ -3 & 9 & 3 \\ -1 & 3 & 1 \end{bmatrix} v$$

Entonces la reflexión es

$$\begin{aligned} R(v) &= v - 2\left(uu^t\right)v \\ &= \left(I - 2\left(uu^t\right)\right)v \\ &= \frac{1}{11}\begin{bmatrix} 9 & 6 & 2 \\ 6 & -7 & -6 \\ 2 & -6 & 9 \end{bmatrix} v \end{aligned}$$

Al evaluar esta fórmula en $v = [1\ 2\ 0]^t$ se obtiene nuevamente la reflexión ya calculada.

Los argumentos empleados en el ejercicio anterior pueden generalizarse sin dificultad: $\left(I - 2\left(uu^t\right)\right)$ es la matriz asociada a una reflexión con respecto a un subespacio, cuyo complemento ortogonal está generado por un vector normalizado u.

Definición 3.18

Para cada $u \in \mathbb{R}^n$ tal que $\|u\| = 1$ asociamos una *matriz de Householder*

$$H = I - 2\left(uu^t\right) \tag{3.18}$$

con la que a su vez definimos la *transformación de Householder*

$$T : \mathbb{R}^n \to \mathbb{R}^n, T(x) = Hx = \left(I - 2\left(uu^t\right)\right)x \tag{3.19}$$

Ejercicio. Demostrar que la matriz de Householder es involutiva ($H^2 = I$) e interpretar gráficamente en $\mathbb{R}^3$.

Ejemplo 3.21 Reflexión con respecto a una recta

Hallaremos la fórmula de la transformación lineal $R : \mathbb{R}^3 \to \mathbb{R}^3$ que es la reflexión con respecto a la recta $S = \text{gen}\left\{[1\ 1\ -2]^t\right\}$.

Como ya se indicó en la fórmula (3.17), para calcular $R(v)$ debemos descomponer al vector en sus proyecciones sobre S y $S^\perp$. En este caso, es más fácil proyectar sobre S puesto que es de dimensión uno:

$$R(v) = P_S(v) - P_{S^\perp}(v) = P_S(v) - (v - P_S(v)) = 2P_S(v) - v$$

Si consideramos un generador normalizado de S, $u = \frac{1}{\sqrt{6}}[1\ 1\ -2]^t$, la matriz asociada a la proyección sobre S es uu^t, con lo cual

$$\begin{aligned} R(v) &= (2P_S - I)\,v = \left(2uu^t - I\right)v \\ &= \frac{1}{3}\begin{bmatrix} -2 & 1 & -2 \\ 1 & -2 & -2 \\ -2 & -2 & 1 \end{bmatrix} v \end{aligned}$$

Debido a la naturaleza geométrica del problema, existen muchos recursos para validar los resultados obtenidos. Una verificación (parcial) de la fórmula consiste en elegir un vector $v \in \mathbb{R}^3$, transformarlo y comprobar que $\frac{1}{2}(v + R(v)) \in S$. Si elegimos, por ejemplo, $v = [3\ 5\ -1]^t$, resulta

$$R(v) = \frac{1}{3}\begin{bmatrix} 1 \\ -5 \\ -17 \end{bmatrix}$$

$$\frac{1}{2}(v + R(v)) = \frac{1}{3}\begin{bmatrix} 5 \\ 5 \\ -10 \end{bmatrix} \in S$$

Esta página fue dejada intencionalmente en blanco

4 Autovalores y autovectores

4.1 Introducción

Comenzamos esta sección retomando el ejemplo 3.20. Allí habíamos obtenido la fórmula

$$R(v) = \frac{1}{11}\begin{bmatrix} 9 & 6 & 2 \\ 6 & -7 & -6 \\ 2 & -6 & 9 \end{bmatrix} v$$

para definir la reflexión con respecto al plano

$$S = \text{gen}\left\{[1\ 1\ -2]^t, [1\ 0\ 1]^t\right\}$$

La matriz que aparece en la fórmula anterior es la matriz asociada a la reflexión $R : \mathbb{R}^3 \to \mathbb{R}^3$ con respecto a la base canónica (de acuerdo con la definición 3.18, se trata de una matriz de Householder):

$$[R]_E = \frac{1}{11}\begin{bmatrix} 9 & 6 & 2 \\ 6 & -7 & -6 \\ 2 & -6 & 9 \end{bmatrix} \tag{4.1}$$

Ahora bien, la matriz asociada a una transformación tiene por columnas a los transformados de los vectores de una base, expresados en coordenadas con respecto a cierta base (eventualmente la misma). En el caso de $[R]_E$, sus columnas son los transformados de los vectores de la base canónica. Ya habíamos mencionado que un problema importante es encontrar una base del espacio para la que la matriz asociada sea "mejor" en algún sentido. Las características geométricas de la transformación R nos hacen prestarle atención a los subespacios S y $S^\perp$. Destacamos que los vectores de S permanecerán fijos, mientras que los de $S^\perp$ se transformarán en sus opuestos: esto es todo lo que se necesita para describir la reflexión. En otras palabras: es mucho más simple describir la reflexión explicando cómo se transforman los vectores de S y $S^\perp$ que describiendo cómo se transforma la base canónica. Definiendo una base ordenada de $\mathbb{R}^3$ con vectores de S y $S^\perp$

$$B = \left\{[1\ 1\ -2]^t, [1\ 0\ 1]^t, [-1\ 3\ 1]^t\right\}$$

resulta

$$\begin{aligned} R\left([1\ 1\ -2]^t\right) &= 1 \cdot [1\ 1\ -2]^t \\ R\left([1\ 0\ 1]^t\right) &= 1 \cdot [1\ 0\ 1]^t \\ R\left([-1\ 3\ 1]^t\right) &= (-1) \cdot [-1\ 3\ 1]^t \end{aligned}$$

y las coordenadas de los transformados con respecto a la base B son sencillamente $[1\,0\,0]^t$, $[0\,1\,0]^t$ y $[0\,0\,-1]^t$. La matriz asociada resulta entonces mucho más simple

$$[R]_B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \tag{4.2}$$

y más fácil de interpretar en términos geométricos que $[R]_E$.

Ejercicio.

4.1

- Interprete geométricamente las potencias

$$([R]_B)^n = \underbrace{[R]_B \,.\, [R]_B \ldots [R]_B}_{n \text{ veces}}$$

- Es inmediato que $\det [R]_B = -1$. Explique por qué $\det [R]_E = -1$ también.

Las observaciones anteriores ponen de manifiesto que la clave para lograr una matriz asociada diagonal es encontrar *una base de vectores que se transformen en múltiplos de ellos mismos.*

Definición 4.1

Dada una matriz $A \in \mathbb{K}^{n\times n}$, un escalar $\lambda \in \mathbb{K}$ es un *autovalor* de la matriz si existe un vector no nulo $x \in \mathbb{K}^{n\times 1}$ tal que

$$Ax = \lambda x \tag{4.3}$$

Todo vector no nulo que verifica la relación (4.3) se dice un *autovector* de la matriz A asociado al autovalor λ.

Una propiedad que se desprende inmediatamente de la definición es que los autovectores asociados a un autovalor, junto con el vector nulo, constituyen un subespacio. Dejamos la demostración formal como ejercicio.

Proposición 4.1

Dada una matriz $A \in \mathbb{K}^{n\times n}$, el conjunto de autovectores asociados a un autovalor, junto con el vector nulo, es decir

$$S_\lambda = \{x \in \mathbb{K}^{n\times 1} : Ax = \lambda x\}$$

es un subespacio de $\mathbb{K}^{n\times 1}$. Lo denominamos *autoespacio* asociado al autovalor λ.

Ejemplo 4.1

Dada la matriz

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

buscaremos sus autovalores y autovectores. Para ello, planteamos la ecuación (4.3) que los define y buscamos alguna

equivalencia que nos permita avanzar:

$$Ax = \lambda x \Leftrightarrow Ax - \lambda Ix = 0 \Leftrightarrow (A - \lambda I)x = 0 \tag{4.4}$$

De acuerdo con la última expresión, el problema de hallar autovectores equivale a encontrar elementos no triviales de $\text{Nul}(A - \lambda I)$. En principio, nada garantiza que existan tales elementos: podría ser $\text{Nul}(A - \lambda I) = \{0\}$. Por lo tanto, debemos elegir valores adecuados de λ, para los cuales sea $\text{Nul}(A - \lambda I) \neq \{0\}$, y esto equivale a que el determinante de la matriz $(A - \lambda I)$ se anule. Ahora bien,

$$\begin{aligned}\det(A - \lambda I) &= \det \begin{bmatrix} -1/3 - \lambda & 2/3 & 2/3 \\ 2/3 & -1/3 - \lambda & 2/3 \\ 2/3 & 2/3 & -1/3 - \lambda \end{bmatrix} \\ &= -\lambda^3 - \lambda^2 + \lambda + 1 \\ &= -(\lambda - 1)(\lambda + 1)^2\end{aligned}$$

Por lo tanto, sólo eligiendo $\lambda = 1$ ó $\lambda = -1$ la ecuación (4.4) tendrá soluciones no triviales.

Si elegimos $\lambda = 1$, los autovectores serán las soluciones no triviales de

$$(A - 1I)v = 0 \Leftrightarrow \begin{bmatrix} -4/3 & 2/3 & 2/3 \\ 2/3 & -4/3 & 2/3 \\ 2/3 & 2/3 & -4/3 \end{bmatrix} v = 0$$

con lo cual el autoespacio asociado al autovalor $\lambda = 1$ resulta

$$S_1 = \text{Nul}(A - 1I) = \text{gen}\left\{[1\ 1\ 1]^t\right\}$$

De manera similar, al elegir $\lambda = -1$ los autovectores serán las soluciones no triviales de

$$(A + 1I)v = 0 \Leftrightarrow \begin{bmatrix} 2/3 & 2/3 & 2/3 \\ 2/3 & 2/3 & 2/3 \\ 2/3 & 2/3 & 2/3 \end{bmatrix} v = 0$$

En este caso, la dimensión del espacio nulo es dos:

$$S_{-1} = \text{Nul}(A + 1I) = \text{gen}\left\{[-1\ 1\ 0]^t, [-1\ 0\ 1]^t\right\}$$

Si consideramos la transformación lineal que se asocia con la matriz

$$T : \mathbb{R}^3 \to \mathbb{R}^3,\ T(x) = Ax$$

notamos que todos los vectores de S_1 permanecen fijos, mientras que los de su complemento ortogonal S_{-1} se transforman en su opuesto. Se trata, por lo tanto, de una reflexión con respecto al eje S_1.

El argumento usado en el ejemplo 4.1 para encontrar los autovalores y autovectores tiene generalidad. Basándonos entonces en la ecuación (4.4), agrupamos las observaciones hechas.

Proposición 4.2

Dada una matriz $A \in \mathbb{K}^{n\times n}$, son equivalentes

1. $\lambda \in \mathbb{K}$ es un autovalor de A.
2. $\text{Nul}(A-\lambda I) \neq \{0\}$
3. $\text{rango}(A-\lambda I) < n$
4. $\det(A-\lambda I) = 0$

Como ya se vio en el ejemplo 4.1, $\det(A-\lambda I)$ es un polinomio que tiene a λ por indeterminada. Más aun, de la definición de determinante se desprende que es un polinomio de grado n. Por ser de gran interés, recibe nombre propio.

Definición 4.2

Dada una matriz $A \in \mathbb{K}^{n\times n}$, su *polinomio característico* [a] es

$$p_A(t) = \det(A - tI)$$

[a] Lo escribiremos usualmente en la indeterminada t para evitar confusión con los autovalores.

Podemos ahora enunciar y demostrar algunas propiedades interesantes sobre los autovalores de una matriz a partir de su polinomio característico.

Teorema 4.1

Sea $A \in \mathbb{C}^{n\times n}$. Entonces

1. Los autovalores de A son las raíces de su polinomio característico $p_A(t)$.
2. Siempre es posible factorizar al polinomio característico como

   $$p_A(t) = (-1)^n (t-\lambda_1)(t-\lambda_2)\dots(t-\lambda_n)$$

   donde los $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$ son los autovalores de A (eventualmente repetidos).
3. A tiene a lo sumo n autovalores distintos.

Demostración:
El enunciado 1 es una consecuencia inmediata de la proposición 4.2.

De las propiedades del determinante se desprende que

$$p_A(t) = \det[(-1)(tI - A)] = (-1)^n \det(tI - A)$$

y de acuerdo con el teorema fundamental del álgebra, todo polinomio con coeficientes en $\mathbb{C}$ y grado $n \geq 1$ tiene n raíces en $\mathbb{C}$ (contando multiplicidades). Por lo tanto siempre es posible factorizar al polinomio característico como

$$p_A(t) = (-1)^n (t-\lambda_1)(t-\lambda_2)\dots(t-\lambda_n)$$

donde los $\lambda_1, \lambda_2, \dots, \lambda_n \in \mathbb{C}$ son los autovalores de A. Entonces A tiene a lo sumo n autovalores distintos. □

Ejemplo 4.2

Consideremos la matriz asociada a una rotación de ángulo $\frac{\pi}{3}$ con centro en $(0,0)$. En la figura 4.1 se aprecian los transformados de los vectores de la base canónica. Luego la matriz asociada para esta base es

$$A = \begin{bmatrix} \cos\frac{\pi}{3} & -\operatorname{sen}\frac{\pi}{3} \\ \operatorname{sen}\frac{\pi}{3} & \cos\frac{\pi}{3} \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}$$

Antes de hacer cálculos, podemos anticipar un dato sobre los autovalores basándonos en las características geométricas de la transformación: ningún vector (distinto del nulo) se transformará en un múltiplo real de sí mismo al rotar un ángulo de $\frac{\pi}{3}$. Luego los autovalores deben ser números complejos. En efecto, el polinomio característico es

$$p_A(t) = t^2 - t + 1 = \left(t - \tfrac{1+i\sqrt{3}}{2}\right)\left(t - \tfrac{1-i\sqrt{3}}{2}\right)$$

Llamando

$$\lambda_1 = \frac{1+i\sqrt{3}}{2} \qquad \lambda_2 = \frac{1-i\sqrt{3}}{2}$$

los autoespacios correspondientes son

$$S_{\lambda_1} = \operatorname{gen}\left\{\begin{bmatrix} i \\ 1 \end{bmatrix}\right\} \qquad S_{\lambda_2} = \operatorname{gen}\left\{\begin{bmatrix} -i \\ 1 \end{bmatrix}\right\}$$

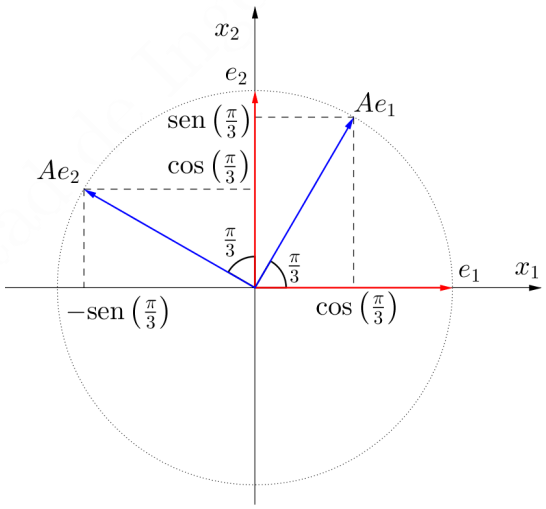


Figura 4.1: Rotación de ángulo $\frac{\pi}{3}$

En este último ejemplo, las raíces del polinomio característico son simples. Si en cambio observamos la factorización del polinomio característico para la reflexión del ejemplo 4.1, detectamos que $\lambda = 1$ es una raíz simple mientras que $\lambda = -1$ es una raíz doble. Esto se vincula con el hecho de que hay una recta (dimensión 1) de vectores que permanecen fijos, mientras que hay un plano (dimensión 2) de vectores que se transforman en sus opuestos. A lo largo del capítulo estudiaremos la relación que hay entre la multiplicidad de las raíces del polinomio característico y la dimensión del autoespacio asociado. Comenzamos con una definición.

Definición 4.3

Dada una matriz $A \in \mathbb{K}^{n\times n}$, para cada autovalor definimos su *multiplicidad algebraica* como su multiplicidad en cuanto raíz de $p_A(t)$.

Notación: si λ es autovalor de A, denotamos a su multiplicidad algebraica como m_λ.

Ejercicio.

4.2 Considere la matriz

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

Encuentre los autovalores, indicando su multiplicidad algebraica y los autoespacios asociados.
Interprete geométricamente considerando la transformación $f : \mathbb{R}^2 \to \mathbb{R}^2, f(x) = Ax$.

A continuación, indagaremos el polinomio característico de una matriz $A = [a_{ij}] \in \mathbb{K}^{2\times 2}$ para establecer una relación con sus raíces $\lambda_1, \lambda_2 \in \mathbb{C}$, es decir con los autovalores de A.

$$\begin{aligned} p_A(t) &= \det(A - tI) \\ &= \det\begin{pmatrix} a_{11} - t & a_{12} \\ a_{21} & a_{22} - t \end{pmatrix} \\ &= t^2 - (a_{11} + a_{22})\, t + (a_{11}a_{22} - a_{12}a_{21}) \\ &= t^2 - (\operatorname{tr} A)\, t + \det A \end{aligned}$$

(Esta expresión de $p_A(t)$ se puede constatar en el ejemplo 4.2).
Por otra parte, desarrollando la expresión factorizada obtenemos

$$\begin{aligned} p_A(t) &= (t - \lambda_1)(t - \lambda_2) \\ &= t^2 - (\lambda_1 + \lambda_2)\, t + \lambda_1\lambda_2 \end{aligned}$$

De esta forma, igualando las dos expresiones obtenidas para $p_A(t)$, podemos establecer una relación entre los autovalores de una matriz y los coeficientes de su polinomio característico.

Proposición 4.3

Dada una matriz $A \in \mathbb{K}^{2\times 2}$ con autovalores $\lambda_1, \lambda_2 \in \mathbb{C}$

1. Su polinomio característico es $p_A(t) = t^2 - (\operatorname{tr} A)\, t + \det A$
2. $\det A = \lambda_1\lambda_2$
3. $\operatorname{tr} A = \lambda_1 + \lambda_2$

Una pregunta que surge naturalmente en este contexto es si estas relaciones pueden extenderse de alguna forma a matrices de mayor tamaño. La respuesta es que, efectivamente, *para cualquier matriz $A \in \mathbb{K}^{n\times n}$ la traza es la suma de los autovalores y el determinante es el producto de los autovalores.*

Proposición 4.4

Dada una matriz $A \in \mathbb{K}^{n\times n}$ con autovalores $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{C}$

1. $\det A = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$
2. $\operatorname{tr} A = \lambda_1 + \lambda_2 + ... + \lambda_n$

La demostración del caso general usa las mismas ideas que para $A \in \mathbb{K}^{2\times 2}$, aunque es más exigente en cuanto a notación y la dejamos como ejercicio para los lectores interesados.

Ejercicio. Dada una matriz $A \in \mathbb{K}^{n\times n}$ con autovalores $\lambda_1, \lambda_2, ..., \lambda_n \in \mathbb{C}$ y adoptando para su polinomio característico la notación $p_A(t) = (-1)^n t^n + a_{n-1} t^{n-1} + ... + a_1 t + a_0$, demuestre que

1. $a_{n-1} = (-1)^{n-1} \operatorname{tr} A$
2. $a_0 = \det A$
3. $\operatorname{tr} A = \lambda_1 + \lambda_2 + ... + \lambda_n$
4. $\det A = \lambda_1 \cdot \lambda_2 \cdots \lambda_n$

Finalizamos esta sección mencionando un concepto que abarca al de autoespacio como caso particular.

Definición 4.4

Dados un subespacio S de $\mathbb{K}^{n\times 1}$ y una matriz $A \in \mathbb{K}^{n\times n}$, se dice que S es un *subespacio invariante* por A si para todo vector $x \in S$ se cumple que $Ax \in S$.

Ejercicio. Demuestre que todo autoespacio de una matriz $A \in \mathbb{K}^{n\times n}$ es un subespacio invariante de $\mathbb{K}^{n\times 1}$.

El ejemplo siguiente ilustra la diferencia que existe entre las nociones de autoespacio y subespacio invariante.

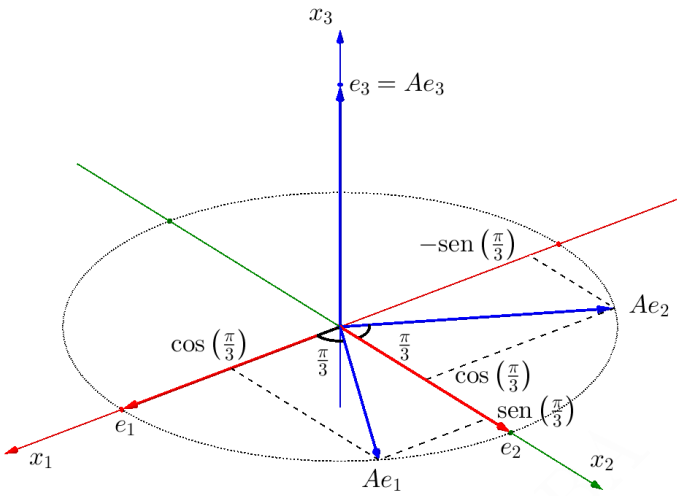


Figura 4.2: Rotación de ángulo $\frac{\pi}{3}$ en torno del eje x_3

Ejemplo 4.3

En $\mathbb{R}^3$ consideramos una rotación de $\frac{\pi}{3}$ en torno del eje x_3. Razonando como en el ejemplo 4.2, deducimos que la matriz que representa este movimiento con respecto a la base canónica es

$$A = \begin{bmatrix} \cos\frac{\pi}{3} & -\text{sen}\,\frac{\pi}{3} & 0 \\ \text{sen}\,\frac{\pi}{3} & \cos\frac{\pi}{3} & 0 \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 & 0 \\ \sqrt{3}/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Antes de hacer ningún cálculo, podemos anticipar que el eje x_3 será un autoespacio asociado al autovalor uno. El polinomio característico es [a]

$$p_A(t) = -(t-1)\left(t^2 - t + 1\right) = -(t-1)\left(t - \tfrac{1+i\sqrt{3}}{2}\right)\left(t - \tfrac{1-i\sqrt{3}}{2}\right)$$

Llamando

$$\lambda_1 = \frac{1+i\sqrt{3}}{2} \qquad \lambda_2 = \frac{1-i\sqrt{3}}{2} \qquad \lambda_3 = 1$$

los autoespacios correspondientes son

$$S_{\lambda_1} = \text{gen}\left\{\begin{bmatrix} i \\ 1 \\ 0 \end{bmatrix}\right\} \quad S_{\lambda_2} = \text{gen}\left\{\begin{bmatrix} -i \\ 1 \\ 0 \end{bmatrix}\right\} \quad S_{\lambda_3} = \text{gen}\left\{\begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}\right\}$$

Es claro que el plano x_1x_2 no puede ser autoespacio, porque ninguno de sus vectores (distinto del nulo) se transforma en múltiplo de sí mismo. Sin embargo, es un subespacio invariante. En efecto, llamando S al plano x_1x_2, tenemos que $S = \text{gen}\{e_1, e_2\}$. Sus elementos son de la forma $\alpha e_1 + \beta e_2$ con $\alpha, \beta \in \mathbb{R}$. Luego sus transformados son elementos del mismo subespacio:

$$A(\alpha e_1 + \beta e_2) = \begin{bmatrix} 1/2 & -\sqrt{3}/2 & 0 \\ \sqrt{3}/2 & 1/2 & 0 \\ 0 & 0 & 1 \end{bmatrix}\left(\alpha\begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} + \beta\begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix}\right)$$

$$= \alpha\begin{bmatrix} 1/2 \\ -\sqrt{3}/2 \\ 0 \end{bmatrix} + \beta\begin{bmatrix} -\sqrt{3}/2 \\ 1/2 \\ 0 \end{bmatrix}$$

El eje x_3, en cambio, es autoespacio y por ende subespacio invariante.

[a] Se trata de un polinomio con coeficientes reales. Por lo tanto, si un número complejo es raíz, también lo es su conjugado: con esto se deduce que al menos una raíz es un número real. Esto señala una notable diferencia con los movimientos en $\mathbb{R}^2$.

Ejercicio. Demuestre que la suma de autoespacios de una matriz $A \in \mathbb{K}^{n\times n}$ es un subespacio invariante por A. ¿Es siempre esta suma un autoespacio?

Ejercicio. Demuestre que si A es una matriz *con elementos en* $\mathbb{R}$ y $\lambda \in \mathbb{C}$ es un autovalor con un autovector asociado $v \in \mathbb{C}^n$, entonces $\overline{\lambda}$ es un autovalor con un autovector asociado $\overline{v}$ (es decir, con todos sus elementos conjugados).

4.2 Diagonalización de matrices

Al comienzo de la sección anterior habíamos considerado la reflexión con respecto a un subespacio y elegido una base B adecuada para que la matriz asociada $[R]_B$ fuera diagonal. Dicha base contiene lo que llamamos autovectores. Con los cambios de base, la relación entre las matrices (4.1) y (4.2) es

$$[R]_E = C_{BE}\,[R]_B\,C_{EB} = C_{BE}\,[R]_B\,(C_{BE})^{-1} \tag{4.5}$$

En esta sección generalizaremos esta idea, buscando para cada matriz A una matriz inversible P de modo tal que resulte $A = PDP^{-1}$ siendo D una matriz diagonal. Veremos en qué condiciones es posible lograr esto. Para ello, daremos previamente algunas definiciones.

Definición 4.5

Dadas las matrices $A \in \mathbb{K}^{n\times n}$ y $B \in \mathbb{K}^{n\times n}$, se dice que son *semejantes* si existe una matriz inversible $P \in \mathbb{K}^{n\times n}$ tal que

$$A = PBP^{-1}$$

N Evidentemente, las matrices $[R]_E$ y $[R]_B$ son semejantes: basta con observar la ecuación (4.5) y considerar $P = C_{BE}$. No es difícil extender este razonamiento para concluir que dos matrices $[f]_{B_1}$ y $[f]_{B_2}$ asociadas a una misma transformación lineal f con respecto a dos bases distintas son semejantes. En otras palabras: *podemos interpretar la semejanza de matrices como dos representaciones matriciales de una misma transformación.*

A la luz de este último comentario, es de esperar que las matrices semejantes tengan muchas propiedades y características en común: las llamaremos *propiedades invariantes por semejanza*. Recogemos algunas de ellas en el teorema siguiente.

Teorema 4.2

Sean $A \in \mathbb{K}^{n\times n}$ y $B \in \mathbb{K}^{n\times n}$ matrices semejantes, es decir que existe una matriz inversible $P \in \mathbb{K}^{n\times n}$ tal que $A = PBP^{-1}$. Entonces se cumple que

1. $p_A(t) = p_B(t)$
2. A y B tienen los mismos autovalores con igual multiplicidad algebraica.
3. $\det A = \det B$
4. $\operatorname{tr} A = \operatorname{tr} B$
5. $\operatorname{rango} A = \operatorname{rango} B$

Demostración:

Para probar la parte 1 usamos propiedades del determinante:

$$\begin{aligned} p_A(t) &= \det(A - tI) = \det(PBP^{-1} - tI) \\ &= \det(PBP^{-1} - tPP^{-1}) = \det\left[P(B - tI)P^{-1}\right] \\ &= \det(P)\det(B - tI)\det\left(P^{-1}\right) \\ &= \det(B - tI) = p_B(t) \end{aligned}$$

De la igualdad de $p_A(t)$ y $p_B(t)$ se desprende inmediatamente 2. Como en particular es $p_A(0) = p_B(0)$, resulta $\det(A - 0I) = \det(B - 0I)$, con lo cual se deduce 3. En cuanto a 4, dado que $p_A(t) = p_B(t)$ y usando la expresión que se da en la proposición 4.3, se deduce la igualdad de las trazas. La propiedad es válida para matrices de $\mathbb{K}^{n\times n}$: para demostrarla se usa la expresión del polinomio característico enunciada al final de la sección 4.1.
Para demostrar 5 tenemos en cuenta que A es la matriz asociada, en base canónica, a la transformación $T(x) = Ax$. A su vez B, por ser semejante, es la matriz asociada a T con respecto a otra base. Por ende, $\text{rango}\, A = \dim(\text{Im}\, T) = \text{rango}\, B$.
□

Ejercicio.

4.3 Considere las matrices

$$A = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \qquad \text{y} \qquad I = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

- Compruebe que verifican todas las propiedades enunciadas en el teorema 4.2.
- Decida si son semejantes.

La definición que sigue refleja la idea con que comenzamos este capítulo: buscar una base para la que la representación matricial de una transformación sea diagonal.

Definición 4.6

Una matriz $A \in \mathbb{K}^{n\times n}$ se dice *diagonalizable* si es semejante a una matriz diagonal.

Como primer ejemplo de matriz diagonalizable, retomamos la que ya fuera estudiada en el ejemplo 4.1.

Ejemplo 4.4

La matriz

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

representa, en coordenadas con respecto a la base canónica, una reflexión. Consideramos una base formada por autovectores asocidados, respectivamente, a los autovalores $\lambda_1 = 1$, $\lambda_2 = \lambda_3 = -1$:

$$B = \left\{[1\ 1\ 1]^t, [-1\ 1\ 0]^t, [-1\ 0\ 1]^t\right\}$$

Siguiendo la ecuación (4.5), proponemos una matriz P que sea el cambio de coordenadas de la base B a la base

canónica y una matriz diagonal D que representa a la reflexión con respecto a la base de autovectores:

$$P = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \wedge D = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix}$$

Con estos datos se comprueba por un cálculo directo la semejanza de A y D:

$$PDP^{-1} = \begin{bmatrix} 1 & -1 & -1 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \frac{1}{3} \begin{bmatrix} 1 & 1 & 1 \\ -1 & 2 & -1 \\ -1 & -1 & 2 \end{bmatrix} = A$$

Ejercicio.

4.4 Demuestre que la matriz A del ejemplo 4.2 es diagonalizable. Encuentre una matriz inversible P y una matriz diagonal D de modo tal que sea $A = PDP^{-1}$. ¿Son únicas?

El ejemplo y el ejercicio anteriores involucran matrices diagonalizables. Llegados a este punto nos preguntamos entonces si toda matriz $A \in \mathbb{K}^{n\times n}$ es diagonalizable. La respuesta será negativa. Sin embargo, tenemos hasta ahora pocos elementos para asegurar que no se puede diagonalizar cierta matriz: habría que probar que no es semejante a ninguna matriz diagonal. Para simplificar esto, estableceremos una importante caracterización de las matrices diagonalizables.

Teorema 4.3

Una matriz $A \in \mathbb{K}^{n\times n}$ es diagonalizable si y sólo si existe una base de $\mathbb{K}^{n\times 1}$ formada por autovectores de A.

Demostración:
Debemos probar una doble implicación: comenzamos por el caso que ya se vio en los ejemplos, es decir, suponemos que existe una base $B = \{v_1, v_2, ..., v_n\}$ formada por autovectores de A. Denotamos como λ_i a los autovalores correspondientes. Es decir $Av_i = \lambda_i v_i$ para $i = 1, 2, ..., n$. Proponemos entonces las matrices

$$P = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \quad D = \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix}$$

La matriz $P \in \mathbb{K}^{n\times n}$ es de rango n (¿por qué?) y por lo tanto inversible. Si logramos probar que $AP = PD$, multiplicando a derecha por P^{-1} obtendremos $A = PDP^{-1}$ como queremos.

$$\begin{aligned} AP &= A\begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \\ &= \begin{bmatrix} Av_1 & Av_2 & \cdots & Av_n \end{bmatrix} \\ &= \begin{bmatrix} \lambda_1 v_1 & \lambda_2 v_2 & \cdots & \lambda_n v_n \end{bmatrix} \\ &= \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \begin{bmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_n \end{bmatrix} \\ &= PD \end{aligned}$$

Recíprocamente, suponemos ahora que la matriz A es diagonalizable, es decir que $A = PDP^{-1}$ para cierta matriz inversible

$$P = \begin{bmatrix} p_1 & p_2 & \cdots & p_n \end{bmatrix}$$

y cierta matriz diagonal

$$D = \begin{bmatrix} \alpha_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \alpha_n \end{bmatrix}$$

Probaremos que las columnas de P constituyen la base de autovectores buscada. En primer lugar, dichas columnas son una base de $\mathbb{K}^{n\times 1}$, puesto que P es inversible. Falta ver que son autovectores de A. Para ello deducimos de $A = PDP^{-1}$ que $AP = PD$, con lo cual

$$\begin{bmatrix} Ap_1 & Ap_2 & \cdots & Ap_n \end{bmatrix} = \begin{bmatrix} \alpha_1 p_1 & \alpha_2 p_2 & \cdots & \alpha_n p_n \end{bmatrix}$$

vale decir que $Ap_i = \alpha_i p_i$ para $i = 1, 2, ..., n$ como queríamos. □

Usaremos este teorema para nuestro primer ejemplo de matriz no diagonalizable.

Ejemplo 4.5

Consideremos la matriz

$$A = \begin{bmatrix} 4 & 1 \\ 0 & 4 \end{bmatrix}$$

Su polinomio característico es $p_A(t) = (t-4)^2$, de modo que su único autovalor es $\lambda_1 = 4$ con multiplicidad algebraica dos. Ahora bien, el autoespacio asociado es

$$S_2 = \text{gen}\left\{ \begin{bmatrix} 1 \\ 0 \end{bmatrix} \right\}$$

Por lo tanto no existe una base de autovectores y la matriz no es diagonalizable.

Destacamos del ejemplo anterior que el autoespacio asociado al autovalor $\lambda_1 = 4$ es de dimensión uno: como no hay más autovalores, será imposible construir una base de autovectores. Dicho de otra manera: dado que la multiplicidad algebraica de $\lambda_1 = 4$ es dos, habríamos necesitado que el autoespacio asociado tuviera la misma dimensión para poder diagonalizar. Esta observación se puede generalizar: *una matriz es diagonalizable si y sólo si la multiplicidad algebraica de cada autovalor coincide con la dimensión del autoespacio asociado.*

Demostrar esta última afirmación no es inmediato y lo haremos en varios pasos. Aunque algunas demostraciones son un poco exigentes, creemos que serán un buen repaso de los conceptos sobre autovectores y autovalores. Comenzamos con una definición.

Definición 4.7

Dada una matriz $A \in \mathbb{K}^{n\times n}$, para cada autovalor definimos su *multiplicidad geométrica* como la dimensión del autoespacio asociado.

Notación: si λ es autovalor de A, denotamos a su multiplicidad geométrica como μ_λ.

Teorema 4.4

Dado un autovalor λ de una matriz $A \in \mathbb{K}^{n\times n}$, su multiplicidad geométrica es menor o igual que su multiplicidad algebraica:

$$\mu_\lambda \leq m_\lambda$$

Demostración:
Como es usual, denotamos mediante S_λ al autoespacio asociado al autovalor λ. Abreviamos como μ su multiplicidad geométrica: $\mu = \dim(S_\lambda)$. Consideramos una base $B_\lambda = \{v_1, v_2, ..., v_\mu\}$ del autoespacio S_λ y la completamos para obtener una base de todo el espacio $\mathbb{K}^{n\times 1}$:

$$B = \{v_1, v_2, ..., v_\mu, ..., v_n\}$$

La idea de la demostración es pensar que la matriz A representa una transformación lineal $T : \mathbb{K}^n \to \mathbb{K}^n$ con respecto a la base canónica, es decir $T(x) = Ax$. Estudiaremos la representación matricial M de *la misma transformación* con respecto a la base B, o sea

$$M = [T]_B$$

por lo que A y M son matrices semejantes. La matriz M, de acuerdo con la proposición 2.5, tiene por columnas las coordenadas con respecto a la base B de los transformados de los vectores de esa base. En el caso de los autovectores tenemos, para $i = 1, ..., \mu$:

$$T(v_i) = \lambda v_i \Rightarrow [T(v_i)]_B = \lambda [v_i]_B$$

donde $[v_i]_B$ es una columna de ceros, salvo en la i-ésima posición, donde hay un uno. De esta forma, las primeras μ columnas de M son

$$\begin{bmatrix} \lambda & 0 & \cdots & 0 \\ 0 & \lambda & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda \\ 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 \end{bmatrix}$$

y la matriz M resulta

$$M = \begin{bmatrix} \lambda I & C \\ 0 & D \end{bmatrix}$$

donde el primer bloque $\lambda I \in \mathbb{K}^{\mu\times\mu}$ es una matriz diagonal con valor λ en la diagonal. El resto de los bloques son matrices con los tamaños adecuados [1]. Por ser A y M semejantes, tienen igual polinomio característico (teorema 4.2) y por ende

$$\begin{aligned} p_A(t) = p_M(t) &= \det(M - tI) \\ &= \det \begin{bmatrix} (\lambda - t)I & C \\ 0 & D - tI \end{bmatrix} \end{aligned}$$

(El símbolo I debe interpretarse como una matriz identidad del tamaño adecuado). Se puede deducir de la definición de determinante (cualquiera que se haya estudiado), que en este caso equivale al producto de los determinantes de los bloques diagonales:

$$\begin{aligned} p_A(t) &= \det[(\lambda - t)I] \det[D - tI] \\ &= (\lambda - t)^\mu \det[D - tI] \end{aligned}$$

[1] $C \in \mathbb{K}^{\mu\times(n-\mu)}$, $D \in \mathbb{K}^{(n-\mu)\times(n-\mu)}$ y $0 \in \mathbb{K}^{(n-\mu)\times\mu}$

La última expresión muestra al polinomio característico de A como un producto de polinomios: el segundo de ellos se desconoce, por lo tanto la multiplicidad de λ como raíz de $p_A(t)$ es *al menos* igual a μ, como queríamos probar. □

Teorema 4.5

Sea $A \in \mathbb{K}^{n\times n}$ y sean $v_1, v_2, ..., v_r$ autovectores correspondientes a autovalores distintos. Entonces $v_1, v_2, ..., v_r$ son linealmente independientes.

Demostración:
Llamamos $\lambda_1, \lambda_2, ..., \lambda_r$ a los autovalores (todos distintos) de modo tal que

$$Av_i = \lambda_i v_i \quad i = 1, ..., r$$

En primer lugar, el conjunto $\{v_1\}$ es linealmente independiente, porque v_1 no es nulo. Probaremos que se pueden agregar sucesivamente todos los autovectores $v_2, ..., v_r$ obteniendo en cada paso un conjunto linealmente independiente. Para ello, y razonando inductivamente, probaremos que agregar uno de estos vectores siempre produce un nuevo conjunto linealmente independiente. Consideramos entonces que el conjunto $\{v_1, ..., v_h\}$ es linealmente independiente (donde h podría ser $1, 2, ..., r-1$); mostraremos que al agregarle un autovector se mantiene independiente. Planteamos una combinación lineal del conjunto $\{v_1, ..., v_h, v_{h+1}\}$ que da cero

$$\alpha_1 v_1 + ... + \alpha_h v_h + \alpha_{h+1} v_{h+1} = 0 \tag{4.6}$$

Multiplicando por λ_{h+1} queda

$$\alpha_1 \lambda_{h+1} v_1 + ... + \alpha_h \lambda_{h+1} v_h + \alpha_{h+1} \lambda_{h+1} v_{h+1} = 0$$

Por otra parte, multiplicando (4.6) por la matriz A resulta

$$\alpha_1 \lambda_1 v_1 + ... + \alpha_h \lambda_h v_h + \alpha_{h+1} \lambda_{h+1} v_{h+1} = 0$$

y restando miembro a miembro las dos últimas ecuaciones obtenemos

$$\alpha_1 (\lambda_{h+1} - \lambda_1) v_1 + ... + \alpha_h (\lambda_{h+1} - \lambda_h) v_h = 0$$

Como supusimos que $\{v_1, ..., v_h\}$ es independiente, debe ser

$$\alpha_i (\lambda_{h+1} - \lambda_i) = 0 \ \ \forall i = 1, ..., h$$

y por tratarse de autovalores distintos, deducimos que

$$\alpha_i = 0 \ \ \forall i = 1, ..., h$$

lo cual reemplazado en (4.6) arroja

$$\alpha_{h+1} v_{h+1} = 0$$

de donde, al ser los autovectores no nulos,

$$\alpha_{h+1} = 0$$

con lo cual hemos probado que la única solución de la ecuación (4.6) es la trivial $\alpha_1 = ... = \alpha_h = \alpha_{h+1} = 0$, de modo que $\{v_1, ..., v_h, v_{h+1}\}$ es independiente. □

Ejercicio. Demuestre que una matriz $A \in \mathbb{C}^{n\times n}$ con todos sus autovalores de multiplicidad algebraica uno es diagonalizable.

Si reunimos los resultados de los últimos teoremas 4.3, 4.4 y 4.5 resulta que para poder diagonalizar una matriz $A \in \mathbb{K}^{n\times n}$ es condición necesaria y suficiente encontrar una base de autovectores. Ahora bien, si la matriz tiene r autovalores distintos λ_i, sus multiplicidades algebraicas m_{λ_i} verifican

$$m_{\lambda_1} + m_{\lambda_2} + \ldots + m_{\lambda_r} = n$$

Sabemos que a cada autovalor λ_i corresponderán μ_{λ_i} autovectores independientes (que también serán independientes del resto de los autovectores). Como $\mu_{\lambda_i} \leq m_{\lambda_i}$ resulta

$$\mu_{\lambda_1} + \mu_{\lambda_2} + \ldots + \mu_{\lambda_r} \leq n$$

y la igualdad se dará si $\mu_{\lambda_i} = m_{\lambda_i}$ para todo $i = 1, \ldots, r$. Esta es la única forma de que exista una base de n autovectores. Resumimos este argumento en el teorema siguiente.

Teorema 4.6

Una matriz $A \in \mathbb{C}^{n\times n}$ es diagonalizable si y sólo para cada autovalor coinciden sus multiplicidades algebraica y geométrica.

Ejercicio.

4.5 Considere la matriz

$$A = \begin{bmatrix} 2 & 5 & 3 \\ 5 & 2 & c \\ 0 & 0 & -3 \end{bmatrix}$$

Decida para qué valores de $c \in \mathbb{R}$ es diagonalizable. Encuentre, en tales casos, una matriz inversible P y una matriz diagonal D tales que $A = PDP^{-1}$.

4.3 Potencias de matrices y polinomios matriciales

Comenzamos esta sección retomando el ejemplo 4.2 de la rotación en $\mathbb{R}^2$. Vamos a indagar qué ocurre al componer rotaciones y cómo resulta la representación matricial.

Ejemplo 4.6

Consideremos la rotación de ángulo $\frac{\pi}{3}$ con centro en $(0,0)$. Si denotamos por $f : \mathbb{R}^2 \to \mathbb{R}^2$ a esta transformación lineal, la matriz asociada con respecto a la base canónica es, como ya habíamos visto,

$$A = [f]_E = \begin{bmatrix} \cos\frac{\pi}{3} & -\operatorname{sen}\frac{\pi}{3} \\ \operatorname{sen}\frac{\pi}{3} & \cos\frac{\pi}{3} \end{bmatrix} = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix}$$

La composición de esta rotación consigo misma es

$$(f \circ f)(x) = f[f(x)]$$

y si usamos la representación matricial resulta

$$\left(f \circ f\right)(x) = A\left(Ax\right) = \left(A \cdot A\right)x$$

Como ya habíamos señalado en el capítulo 2, el producto de las matrices se corresponde con la composición de las transformaciones que representan. Por lo tanto, hemos de esperar que $A.A$ sea la matriz que representa una rotación de ángulo $\frac{\pi}{3} + \frac{\pi}{3} = 2\frac{\pi}{3}$:

$$A.A = \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} \begin{bmatrix} 1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & 1/2 \end{bmatrix} = \begin{bmatrix} -1/2 & -\sqrt{3}/2 \\ \sqrt{3}/2 & -1/2 \end{bmatrix} = \begin{bmatrix} \cos\left(2\frac{\pi}{3}\right) & -\operatorname{sen}\left(2\frac{\pi}{3}\right) \\ \operatorname{sen}\left(2\frac{\pi}{3}\right) & \cos\left(2\frac{\pi}{3}\right) \end{bmatrix}$$

Para representar con generalidad esta clase de producto, adoptamos la notación más obvia.

Definición 4.8

Dada una matriz $A \in \mathbb{K}^{n\times n}$ definimos para cualquier $n \in \mathbb{N}$ su potencia n-ésima mediante

$$A^0 = I \qquad\qquad A^n = \underbrace{A \cdot A \cdots A}_{n \text{ veces}}$$

Ejercicio.

4.6 Considere la matriz que representa una rotación de ángulo θ

$$A = \begin{bmatrix} \cos\theta & -\operatorname{sen}\theta \\ \operatorname{sen}\theta & \cos\theta \end{bmatrix}$$

y encuentre la forma general de A^n. Interprete geométricamente.

El cálculo de las potencias se simplifica mucho para el caso de las matrices diagonalizables. En efecto, dada una matriz diagonalizable $A \in \mathbb{K}^{n\times n}$, sabemos que existen P inversible y D diagonal tales que $A = PDP^{-1}$, con lo cual

$$A^n = \left(PDP^{-1}\right)^n = \underbrace{\left(PDP^{-1}\right)\left(PDP^{-1}\right)\cdots\left(PDP^{-1}\right)}_{n \text{ veces}} = PD^nP^{-1} \tag{4.7}$$

y en la última expresión la matriz D^n es una matriz diagonal con las potencias n-ésimas de los autovalores de A.

Ya mencionamos que las potencias de una matriz se asocian con la composición de transformaciones lineales. En el ejemplo siguiente veremos cómo estas potencias pueden usarse en las aplicaciones para representar la repetición de cierto proceso. Los lectores interesados encontrarán más ejemplos de aplicaciones en [6] y [5].

Ejemplo 4.7

Consideremos que dos empresas lácteas abastecen de leche a cierta población. Con el tiempo, parte de los clientes cambian su preferencia por distintas razones (publicidad o costo entre otras). Buscamos modelar y analizar el movimiento de clientes entre los proveedores asumiendo, por simplicidad, que el número total de clientes permanece constante. Supongamos que al comenzar nuestro modelo las dos empresas, E_1 y E_2, controlan $a_0 = \frac{2}{3}$ y $b_0 = \frac{1}{3}$ (res-

pectivamente) del mercado. Definimos entonces un *vector de estado inicial*

$$x_0 = \begin{bmatrix} a_0 \\ b_0 \end{bmatrix} = \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix}$$

Estamos suponiendo que ninguna otra productora de leche interviene en el mercado, por eso se cumple que $a_0 + b_0 = 1$. Supongamos que después de un mes la empresa E_1 ha logrado mantener el 80 % de sus propios clientes y ha atraído al 30 % de los clientes de E_2. Por otra parte, la empresa E_2 ha logrado atraer al 20 % de los clientes de E_1 y conserva el 70 % de su propia clientela. Teniendo en cuenta las consideraciones anteriores, pueden obtenerse las fracciones a_1 y b_1 de clientes que tiene cada una de las empresas al cabo de un mes:

$$\begin{aligned} a_1 &= 0,8a_0 + 0,3b_0 \\ b_1 &= 0,2a_0 + 0,7b_0 \end{aligned}$$

En forma matricial

$$\underbrace{\begin{bmatrix} a_1 \\ b_1 \end{bmatrix}}_{x_1} = \underbrace{\begin{bmatrix} 0,8 & 0,3 \\ 0,2 & 0,7 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} a_0 \\ b_0 \end{bmatrix}}_{x_0}$$

La matriz A contiene la información sobre el intercambio de clientes entre ambas empresas. Si asumimos que esta forma de intercambiar clientes se mantiene estable, podemos obtener una sucesión de vectores en $\mathbb{R}^2$

$$\begin{aligned} x_1 &= Ax_0 \\ x_2 &= Ax_1 \\ &\vdots \\ x_k &= Ax_{k-1} \end{aligned}$$

donde cada x_k es el vector de estado al cabo de k meses. Esta sucesión puede expresarse en términos de las potencias de A y el vector de estado inicial x_0:

$$\begin{aligned} x_1 &= Ax_0 \\ x_2 &= Ax_1 = AAx_0 = A^2x_0 \\ &\vdots \\ x_k &= Ax_{k-1} = AAx_{k-2} = \cdots = A^kx_0 \end{aligned}$$

Calculemos las potencias de A aplicando diagonalización. Dejamos a cargo del lector verificar que

$$A = PDP^{-1} \quad \text{con} \quad P = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix} \quad \text{y} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 0,5 \end{bmatrix}$$

Usando la fórmula (4.7) podemos calcular sus potencias como

$$A^k = PD^kP^{-1} = \begin{bmatrix} 3 & 1 \\ 2 & -1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 0 & (0,5)^k \end{bmatrix} \begin{bmatrix} 0,2 & 0,2 \\ 0,4 & -0,6 \end{bmatrix}$$

El aspecto más interesante es la posibilidad de predecir el comportamiento de un sistema (en este caso el comportamiento del mercado lácteo) a largo plazo. Esto significa preguntarnos qué ocurre con x_k cuando consideramos un número grande de períodos de tiempo (meses en este caso). Cuando k tiende a infinito, la matriz D^k tiende [a] a

$$D^* = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}$$

y por lo tanto podemos anticipar el comportamiento de la sucesión de matrices A^k:

$$\lim_{k\to\infty} A^k = P \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix} P^{-1} = \begin{bmatrix} 0,6 & 0,6 \\ 0,4 & 0,4 \end{bmatrix}$$

Por lo tanto resulta

$$x^* = \lim_{k\to\infty} x_k = \lim_{k\to\infty} A^k x_0 = \begin{bmatrix} 0,6 & 0,6 \\ 0,4 & 0,4 \end{bmatrix} \begin{bmatrix} 2/3 \\ 1/3 \end{bmatrix} = \begin{bmatrix} 0,60 \\ 0,40 \end{bmatrix}$$

El vector límite x^* indica que, si se cumplen las suposiciones que hicimos, las empresas E_1 y E_2 tenderán a controlar el 60 % y el 40 % del mercado respectivamente.

[a] Entendemos que este límite es intuitivamente claro y omitimos su definición formal. Tan sólo mencionamos que se puede usar la norma de matrices que induce el producto interno del ejemplo 3.2, con la cual el límite $D^k \to D^*$ significa que $\left\| D^k - D^* \right\| \to 0$ conforme $k \to \infty$.

Ejercicio. Compruebe que en el ejemplo anterior el vector $\lim_{k\to\infty} x_k$ es independiente de cuál sea el vector de estado inicial.

Sugerencia: considere $x_0 = \begin{bmatrix} a \\ a-1 \end{bmatrix}$ donde $0 \le a \le 1$

A continuación indagamos si es siempre posible encontrar un límite para una sucesión de matrices o de vectores.

Ejemplo 4.8

Consideramos

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

Se trata de una matriz diagonalizable:

$$A = PDP^{-1} \quad \text{con} \quad P = \begin{bmatrix} 1 & 2 \\ 0 & 1 \end{bmatrix} \quad \text{y} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 1/2 \end{bmatrix}$$

Luego

$$D^k = \begin{bmatrix} 2^k & 0 \\ 0 & (1/2)^k \end{bmatrix}$$

Como no existe el límite de 2^k cuando $k \to \infty$, no podemos hallar un límite para la sucesión de matrices A^k. Observamos cómo *la convergencia de una sucesión de matrices depende de los autovalores de A*.

Dado un vector $v \in \mathbb{R}^2$, ¿qué ocurre con la sucesión de vectores $A^k v$? Consideremos en primer lugar el caso de los autovectores de la matriz:

- Si v es un autovector asociado al autovalor 2, no existe el límite de la sucesión de vectores. En efecto, para este caso es $A^k v = 2^k v$.
- Si en cambio v es un autovector asociado al autovalor $1/2$ resulta $\lim_{k\to\infty} A^k v = \lim_{k\to\infty} (1/2)^k v = 0$.
- En el caso general, los vectores $v \in \mathbb{R}^2$ pueden escribirse en la forma

$$v = \alpha \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

para α y β adecuados. Por lo tanto

$$A^k v = \alpha A^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta A^k \begin{bmatrix} 2 \\ 1 \end{bmatrix} = \alpha 2^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} + \beta \left(\frac{1}{2}\right)^k \begin{bmatrix} 2 \\ 1 \end{bmatrix}$$

Entonces, la sucesión de vectores $A^k v$ será convergente si y sólo $\alpha = 0$, es decir si v pertenece al autoespacio asociado al $1/2$.

Ejercicio.

4.7 Encuentre, si es posible, $\lim_{k\to\infty} A^k v$ para $v \in \mathbb{R}^2$ y $A = \begin{bmatrix} 2/3 & 1 \\ 0 & 1 \end{bmatrix}$.

La definición que sigue contiene a las potencias de matrices como un caso particular.

Definición 4.9

Dados un polinomio $p(t) = a_k t^k + a_{k-1} t^{k-1} + \ldots + a_1 t + a_0$ con $a_i \in \mathbb{R}$ $(i = 0, 1, \ldots, k)$, $a_k \neq 0$ y una matriz $A \in \mathbb{K}^{n\times n}$, el *polinomio matricial* $p(A)$ es

$$p(A) = a_k A^k + a_{k-1} A^{k-1} + \ldots + a_1 A + a_0 I$$

Ejemplo 4.9

Consideramos en $\mathbb{R}^3$ la reflexión con respecto a un plano S, como en la figura 3.13. Su complemento ortogonal es $S^\perp = \text{gen}\{u\}$, donde u es un vector que por sencillez elegimos de norma 1. De acuerdo con la definición 3.18, $H = I - 2\left(uu^t\right)$ es la matriz (de Householder) asociada a la reflexión con respecto a S. A su vez $\left(uu^t\right)$ es la matriz asociada a la proyección sobre $S^\perp$ (¿por qué?). Si definimos el polinomio

$$p(t) = -2t + 1$$

de acuerdo con la definición de polinomio matricial es $H = p\left(uu^t\right)$.
Por ser $\left(uu^t\right)$ la matriz asociada a una proyección, sus autovalores son 0 y 1. Notablemente, al evaluar $p(t)$ en ellos

$$p(0) = -2 \cdot 0 + 1 = 1 \quad \wedge \quad p(1) = -2 \cdot 1 + 1 = -1$$

obtenemos los autovalores de la reflexión. Más aún, para $\left(uu^t\right)$ el autoespacio asociado al 1 es $S^\perp$ y pasa a ser el autoespacio asociado a $p(1) = -1$ en la matriz H. Lo mismo ocurre para el autoespacio asociado al 0.

Demostraremos que la relación señalada en el ejemplo no es casual.

Teorema 4.7

Dados un polinomio $p(t) = a_k t^k + a_{k-1} t^{k-1} + ... + a_1 t + a_0$ con $a_i \in \mathbb{R}$ $(i = 0, 1, ..., k)$, $a_k \neq 0$ y una matriz $A \in \mathbb{K}^{n \times n}$

1. Si λ es autovalor de A, entonces $p(\lambda)$ es autovalor de $p(A)$.
   Además, si v es autovector de A asociado a λ, también será autovector de $p(A)$ asociado a $p(\lambda)$.
2. Para cada autovalor ω de $p(A)$ existe un autovalor λ de A tal que $p(\lambda) = \omega$.

Demostración:
El caso en que $k = 0$, es decir con $p(t) = a_0$ es inmediato y se deja como ejercicio. En lo que sigue suponemos $k \geq 1$.
Para probar 1, consideramos que v es un autovector de A asociado al autovalor λ. Vale decir que $Av = \lambda v$. Razonando inductivamente, llegamos a que para cualquier $k \in \mathbb{N}$: $A^k v = \lambda^k v$. Demostraremos entonces que $p(\lambda)$ es autovalor de $p(A)$ y que el mismo v es un autovector asociado:

$$\begin{aligned} p(A) \cdot v &= \left(a_k A^k + a_{k-1} A^{k-1} + ... + a_1 A + a_0 I\right) v \\ &= a_k A^k v + a_{k-1} A^{k-1} v + ... + a_1 Av + a_0 Iv \\ &= a_k \lambda^k v + a_{k-1} \lambda^{k-1} v + ... + a_1 \lambda v + a_0 v \\ &= \left(a_k \lambda^k + a_{k-1} \lambda^{k-1} + ... + a_1 \lambda + a_0\right) v = p(\lambda) \cdot v \end{aligned}$$

En cuanto a 2, definimos un polinomio $q(t) = p(t) - \omega$. Probaremos en primer lugar que la matriz $q(A) = p(A) - \omega I$ es singular. En efecto, sea v un autovector asociado al autovalor ω, entonces se tiene

$$q(A) \cdot v = \left[p(A) - \omega I\right] v = p(A) v - \omega Iv = \omega Iv - \omega Iv = 0$$

Por otra parte, $q(t)$ admite una factorización de la forma

$$q(t) = a_k (t - t_1)(t - t_2) \cdots (t - t_k)$$

con $t_i \in \mathbb{C}$ $(i = 1, 2, ..., k)$. Una consecuencia inmediata de la definición de polinomio matricial es que si un polinomio se factoriza como $r(t) = r_1(t) \cdot r_2(t)$ entonces para cualquier matriz cuadrada A es $r(A) = r_1(A) \cdot r_2(A)$. Con esto llegamos a

$$q(A) = a_k (A - t_1 I)(A - t_2 I) \cdots (A - t_k I)$$

y aplicando determinante a esta matriz que ya sabemos singular resulta

$$0 = \det q(A) = (a_k)^n \det(A - t_1 I) \det(A - t_2 I) \cdots \det(A - t_k I)$$

Entonces, al menos uno de estos factores debe anularse. Llamemos t^* al valor para el cual

$$\det\left(A - t^* I\right) = 0$$

Esto significa que t^* es un autovalor de A y que

$$q\left(t^*\right) = 0 = p(t^*) - \omega$$

con lo cual $p(t^*) = \omega$ y t^* es entonces un autovalor de A que se transforma en ω. □

N Al usar las propiedades 1 y 2 juntas, podemos concluir que calculando $\omega = p(\lambda)$ para cada autovalor λ de A, se agotan todos los autovalores ω de $p(A)$. Sin embargo, esto no implica una preservación de las multiplicidades geométricas ni algebraicas.

Ejemplo 4.10

La matriz

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

tiene dos autovalores $\lambda_1 = 1$ y $\lambda_2 = -1$, ambos de multiplicidad algebraica dos:

$$\det(A - tI) = (1-t)^2(-1-t)^2$$

Sin embargo, las multiplicidades geométricas son ambas uno:

$$S_{\lambda_1} = \text{gen}\{[1\,0\,0\,0]^t\} \qquad S_{\lambda_2} = \text{gen}\{[0\,0\,1\,0]^t\}$$

Al considerar el polinomio $p(t) = t^4 - 2t^2$ resulta

$$p(A) = A^4 - 2A^2 = \begin{bmatrix} -1 & 0 & 0 & 0 \\ 0 & -1 & 0 & 0 \\ 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

que tiene un autovalor $\omega = p(\lambda_1) = p(\lambda_2) = -1$ de multiplicidad algebraica y geométrica cuatro:

$$\det(A - tI) = (-1-t)^4 \qquad S_\omega = \mathbb{R}^4$$

Por lo tanto, en este caso la matriz no diagonalizable A se transformó en la matriz diagonalizable $p(A)$.

Ejercicio. Demuestre que la matriz $A = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}$ no es diagonalizable y que sin embargo A^2 lo es.

Ejercicio.

4.8 Sea $A \in \mathbb{R}^{3\times 3}$ que verifica

- A es inversible.
- $A^2 + 2A$ es singular.
- $\dim[\text{Nul}(A - 3I)] = 2$

Encuentre los autovalores de A y decida si es diagonalizable.

Ejercicio.

4.9 ¿Existen una matriz $A \in \mathbb{K}^{n\times n}$ y un polinomio $p \in \mathscr{P}$ tales que A es diagonalizable pero $p(A)$ no lo es?

Esta página fue dejada intencionalmente en blanco

5 Matrices simétricas y hermíticas

5.1 Matrices ortogonales y unitarias

Comenzamos este capítulo señalando algunos aspectos de la reflexión con respecto a un subespacio. En el ejemplo 3.20 habíamos obtenido, para la reflexión con respecto a un plano de $\mathbb{R}^3$, la siguiente matriz asociada con respecto a la base canónica:

$$A = \frac{1}{11}\begin{bmatrix} 9 & 6 & 2 \\ 6 & -7 & -6 \\ 2 & -6 & 9 \end{bmatrix} \tag{5.1}$$

Queremos señalar dos propiedades de esta matriz:

1. Es simétrica: $A = A^t$.
2. Es involutiva: $A^2 = I$, lo cual puede anticiparse por las características de la reflexión y comprobarse por un cálculo directo.

Usando estas dos propiedades podemos afirmar que

$$A^t A = I$$

Ya habíamos mencionado (al estudiar la descomposición QR en el teorema 3.14) que si esto ocurre es porque las columnas de A son una base ortonormal de Col A. Dado que las columnas de A son los transformados de los vectores de la base canónica, destacamos que en esta reflexión *la base canónica se transforma en otra base ortonormal.*

Convención

En todo este capítulo, siempre que se mencionen $\mathbb{R}^n$ y $\mathbb{C}^n$ como espacios con producto interno, se entenderá que es el *producto interno canónico* como se lo definió en 3.1 y 3.5 respectivamente.

La definición que sigue amplía la idea anterior a los $\mathbb{C}$ espacios.

Definición 5.1

Una matriz $P \in \mathbb{C}^{n\times n}$ se dice *unitaria* si verifica

$$P^H P = P P^H = I$$

o, de manera equivalente, si $P^H = P^{-1}$.
En el caso particular de una matriz unitaria $P \in \mathbb{R}^{n\times n}$ resulta

$$P^t P = P P^t = I$$

y se dice que P es una matriz *ortogonal*.

Ejercicio.

- Demuestre que si una matriz $P \in \mathbb{K}^{n\times n}$ es unitaria entonces P^H también lo es.
- Demuestre que el producto de matrices unitarias es una matriz unitaria.

El teorema que sigue establece formalmente lo que ya habíamos señalado acerca de las columnas ortonormales. Se enuncia para matrices en $\mathbb{C}^{n\times n}$ y entre paréntesis puede leerse la versión para matrices en $\mathbb{R}^{n\times n}$.

Teorema 5.1

Sea $P \in \mathbb{C}^{n\times n}$ ($P \in \mathbb{R}^{n\times n}$). Son equivalentes:

1. P es unitaria (ortogonal).
2. Las columnas de P forman una base ortonormal de $\mathbb{C}^n$ ($\mathbb{R}^n$).
3. Las filas de P forman una base ortonormal de $\mathbb{C}^n$ ($\mathbb{R}^n$).

Demostración:
Sólo demostraremos el caso en que P es unitaria, ya que el caso en que la matriz es ortogonal se deduce empleando los mismos argumentos.

En primer lugar, para probar que 1 ⇒ 2, el argumento es el mismo que se usó en el teorema 3.14 al estudiar la descomposición QR.
Recíprocamente, si las columnas de P forman una base ortonormal de $\mathbb{C}^n$, entonces $P^H P = I$. Ahora bien, esto significa que la matriz P^H es la inversa de P. [1] Luego $P^H = P^{-1}$ y por lo tanto 2 ⇒ 1.

Para probar las restantes equivalencias, alcanza con demostrar que 1 ⇔ 3. Suponiendo en primer lugar que vale 1, es decir que P es unitaria, entonces P^t también es unitaria (¡pruébelo!) con lo cual, apelando a 2 resulta que sus columnas son una base ortonormal. Pero las columnas de P^t son las filas de P y por ende 1 ⇒ 3.
Recíprocamente, si las filas de P forman una base ortonormal de $\mathbb{C}^n$, entonces las columnas de P^t forman una base ortonormal. Apelando a 2 deducimos que P^t es unitaria y luego P es también unitaria. Esto prueba que 3 ⇒ 1. □

[1] Una propiedad de las matrices cuadradas es que $BA = I \Rightarrow B = A^{-1}$, es decir que alcanza con ser inversa por izquierda para ser *la* inversa (ver [4]).

Al comienzo de esta sección estudiamos un ejemplo de la reflexión y su representación matricial. Por tratarse de un movimiento debe preservar la norma de los vectores y los ángulos: veremos a continuación que cualquier transformación determinada por una matriz unitaria verifica estas propiedades.

Teorema 5.2

Sea $P \in \mathbb{C}^{n\times n}$ una matriz unitaria. Entonces vale lo siguiente:

1. $(Pu, Pv) = (u, v)$ para todos $u, v \in \mathbb{C}^n$.
2. $\|Pu\| = \|u\|$ para todo $u \in \mathbb{C}^n$.
3. Si $\lambda \in \mathbb{C}$ es un autovalor de P, entonces $|\lambda| = 1$.
4. $|\det P| = 1$

Demostración:
Para demostrar 1 usamos la definición del producto interno canónico:

$$(Pu, Pv) = (Pu)^H Pv = u^H P^H Pv = u^H v = (u, v)$$

La propiedad 2 es una consecuencia de 1 y la dejamos como ejercicio.
Sea $\lambda \in \mathbb{C}$ un autovalor de P. Vale decir que existe un vector no nulo $v \in \mathbb{C}^n$ tal que $Pv = \lambda v$. Luego, usando 2 y teniendo en cuenta que $\|v\| \neq 0$ tenemos que

$$\|v\| = \|Pv\| = |\lambda|\,\|v\| \Rightarrow |\lambda| = 1$$

De acuerdo con la proposición 4.4, el determinante de una matriz es el producto de sus autovalores contando multiplicidades. Como en este caso se trata de autovalores de módulo uno, resulta la propiedad 4. □

Ejercicio. Demuestre la propiedad 4 del teorema 5.2 teniendo en cuenta que $\det\left(PP^H\right) = \det I$.

5.2 Diagonalización de matrices simétricas y hermíticas

En esta sección estudiaremos una clase de matrices diagonalizables de especial interés: las matrices simétricas y su generalización a $\mathbb{C}$ espacios.

Definición 5.2

Una matriz $A \in \mathbb{C}^{n\times n}$ se dice *hermítica* si verifica

$$A^H = A$$

En el caso particular de una matriz hermítica $A \in \mathbb{R}^{n\times n}$ resulta

$$A^t = A$$

y se dice que A es una matriz *simétrica*.

Ejercicio. Demuestre que los elementos de la diagonal principal de una matriz hermítica son siempre números reales.

Ejemplo 5.1

Consideremos la siguiente matriz simétrica

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

Su polinomio característico es

$$p_A(t) = t^2 - 2t - 8 = (t-4)(t+2)$$

con lo cual sus autovalores son $\lambda_1 = 4$ y $\lambda_2 = -2$. En cuanto a los autoespacios, estos son

$$S_{\lambda_1} = \text{gen}\{[1\ -1]^t\} \qquad S_{\lambda_2} = \text{gen}\{[1\ 1]^t\}$$

Entonces definiendo $v_1 = [1\ -1]^t$ y $v_2 = [1\ 1]^t$ resulta $\tilde{B} = \{v_1, v_2\}$ una *base ortogonal* de $\mathbb{R}^2$. Si además normalizamos estos vectores obtenemos una *base ortonormal* de $\mathbb{R}^2$ compuesta por autovectores de A:

$$B = \left\{ \frac{v_1}{\|v_1\|}, \frac{v_2}{\|v_2\|} \right\} = \left\{ \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{bmatrix}^t, \begin{bmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \end{bmatrix}^t \right\}$$

Si ahora consideramos las matrices

$$P = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \quad \text{y} \quad D = \begin{bmatrix} 4 & 0 \\ 0 & -2 \end{bmatrix}$$

resulta que P es ortogonal y

$$A = PDP^{-1} = PDP^t \tag{5.2}$$

Observamos que A no sólo es diagonalizable, sino que además hemos podido elegir la matriz de autovectores que aparece en su diagonalización de forma que sea *ortogonal*.

Ejercicio.

5.1 Considere la matriz asociada a la reflexión (5.1) presentada al comienzo de este capítulo y encuentre una diagonalización como en (5.2), es decir, con la matriz P ortogonal.
Sugerencia: en lugar de hacer los cálculos del polinomio característico y autoespacios, se pueden conocer sus autovalores y autovectores asociados considerando la interpretación geométrica del ejemplo 3.20.

Ejemplo 5.2

Consideremos ahora la matriz hermítica

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

cuyo polinomio característico es

$$p_A(t) = t^2 - 2t = t(t-2)$$

y sus autovalores son $\lambda_1 = 2$ y $\lambda_2 = 0$, *ambos reales*. En cuanto a los autoespacios, estos son

$$S_{\lambda_1} = \text{gen}\{[i\ 1]^t\} \qquad S_{\lambda_2} = \text{gen}\{[-i\ 1]^t\}$$

con lo cual, llamando $v_1 = [i\ 1]^t$ y $v_2 = [-i\ 1]^t$ resulta $\tilde{B} = \{v_1, v_2\}$ una *base ortogonal* de $\mathbb{C}^2$. Si además normalizamos a estos vectores obtenemos una *base ortonormal* de $\mathbb{C}^2$ compuesta por autovectores de A:

$$B = \left\{ \frac{v_1}{\|v_1\|}, \frac{v_2}{\|v_2\|} \right\} = \left\{ \left[\frac{i}{\sqrt{2}} \ \frac{1}{\sqrt{2}} \right]^t, \left[-\frac{i}{\sqrt{2}} \ \frac{1}{\sqrt{2}} \right]^t \right\}$$

Considerando entonces las matrices

$$P = \begin{bmatrix} i/\sqrt{2} & -i/\sqrt{2} \\ 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix} \quad \text{y} \quad D = \begin{bmatrix} 2 & 0 \\ 0 & 0 \end{bmatrix}$$

resulta que P es unitaria y

$$A = PDP^{-1} = PDP^H \tag{5.3}$$

En este caso hemos podido elegir la matriz de autovectores de forma que sea *unitaria*.

Las diagonalizaciones (5.2) y (5.3) serán de especial interés y llevan nombre propio.

Definición 5.3

- Una matriz $A \in \mathbb{R}^{n\times n}$ se dice *diagonalizable ortogonalmente* si existen $P \in \mathbb{R}^{n\times n}$ ortogonal y $D \in \mathbb{R}^{n\times n}$ diagonal tales que
$$A = PDP^t \tag{5.4}$$
- Una matriz $A \in \mathbb{C}^{n\times n}$ se dice *diagonalizable unitariamente* si existen $P \in \mathbb{C}^{n\times n}$ unitaria y $D \in \mathbb{C}^{n\times n}$ diagonal tales que
$$A = PDP^H \tag{5.5}$$

Es obvio que toda matriz diagonalizable unitariamente es, en particular, diagonalizable. Además, por lo que ya sabemos sobre diagonalización de matrices, las columnas de la matriz P son autovectores de A. Queda claro que si A es diagonalizable

unitariamente, entonces existe una base ortonormal de $\mathbb{C}^n$ compuesta por autovectores de A.
Por otra parte, si es posible hallar una base ortonormal de $\mathbb{C}^n$ compuesta por autovectores de A, es inmediato que podemos factorizar A en la forma (5.5): formamos P poniendo como columnas los autovectores de A que componen la base ortonormal y D poniendo los autovalores de A en la diagonal y cero en el resto.
Estos mismos argumentos pueden adaptarse para el caso de matrices reales. En conclusión

Teorema 5.3

Una matriz $A \in \mathbb{C}^{n\times n}$ ($A \in \mathbb{R}^{n\times n}$) es diagonalizable unitariamente (ortogonalmente) si y sólo si existe una base ortonormal de $\mathbb{C}^n$ ($\mathbb{R}^n$) compuesta por autovectores de A.

El teorema siguiente ratifica la generalidad de los comentarios hechos en los ejemplos 5.1 y 5.2.

Teorema 5.4 (Espectral)

Sea $A \in \mathbb{C}^{n\times n}$ una matriz hermítica. Entonces vale lo siguiente:

1. Todos sus autovalores son números reales.
2. Autoespacios asociados a diferentes autovalores son mutuamente ortogonales.
3. Es diagonalizable unitariamente.
4. Si además $A \in \mathbb{R}^{n\times n}$, y es por lo tanto simétrica, A es diagonalizable ortogonalmente.

Demostración:
En lo que sigue, usaremos la siguiente propiedad de las matrices hermíticas, válida para $u, v \in \mathbb{C}^n$:

$$(Au, v) = (u, Av)$$

que se explica como sigue:

$$(Au, v) = (Au)^H v = u^H A^H v = u^H Av = (u, Av)$$

Para demostrar 1, consideremos un autovalor λ de A. Sea v un autovector asociado a λ. Entonces, por una parte es

$$(Av, v) = (\lambda v, v) = \overline{\lambda}(v, v)$$

mientras que también tenemos

$$(Av, v) = (v, Av) = (v, \lambda v) = \lambda(v, v)$$

Luego $\overline{\lambda}(v, v) = \lambda(v, v)$ y por ser $(v, v) \neq 0$ resulta $\overline{\lambda} = \lambda$. Entonces λ es necesariamente real.

Para demostrar 2, consideremos λ y ω dos autovalores distintos de A. Para probar que $S_\lambda \perp S_\omega$ bastará con ver que si $u \in S_\lambda$ y $v \in S_\omega$ entonces $(u, v) = 0$. Teniendo en cuenta que $\lambda, \omega \in \mathbb{R}$:

$$\lambda(u, v) = (\lambda u, v) = (Au, v) = (u, Av) = (u, \omega v) = \omega(u, v)$$

Con lo cual $(\lambda - \omega)(u, v) = 0$ y al ser $(\lambda - \omega) \neq 0$ sigue el resultado.

Omitimos la demostración de 3 y 4. El lector interesado podrá encontrar los detalles en [4]. □

5.3 Formas cuadráticas

Seguramente nuestros lectores han estudiado la función homográfica

$$f(x) = \frac{1}{x}$$

Su gráfico, que se aprecia en la figura 5.1, es usualmente conocido como *hipérbola equilátera*[2] y queda caracterizado por la ecuación

$$x \cdot y = 1$$

que, a primera vista, no parece tener mucha relación con la ecuación canónica de las hipérbolas

$$\frac{x^2}{a^2} - \frac{y^2}{b^2} = 1$$

¿Se trata realmente de una hipérbola? Veremos que sí, que la curva gráfico de la función f es efectivamente una hipérbola: lo que sucede es que sus ejes principales están rotados con respecto a los ejes x e y.

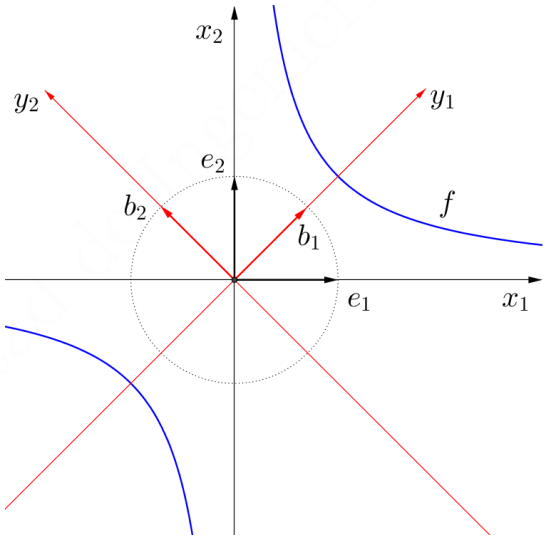


Figura 5.1: Hipérbola equilátera

Comenzamos por adoptar la notación a la usual de $\mathbb{R}^n$: es decir que estudiaremos la ecuación

$$x_1 \cdot x_2 = 1$$

a la que podemos escribir como un producto de matrices adecuadas (¡comprobarlo!):

$$[x_1 \; x_2] \underbrace{\begin{bmatrix} 0 & 1/2 \\ 1/2 & 0 \end{bmatrix}}_{A} \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = 1$$

[2] Se suele denominar "equiláteras" a las hipérbolas cuyas asíntotas son perpendiculares.

y destacamos que la matriz A es simétrica. Por lo tanto, y de acuerdo con el teorema espectral 5.4, es diagonalizable ortogonalmente. Es decir, existe una base ortonormal $B = \{b_1, b_2\}$ cuyos vectores podemos expresar como columnas de una matriz P de modo tal que $A = PDP^t$, siendo D una matriz diagonal:

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

Con esto, la ecuación anterior adopta la forma

$$x^t Ax = 1 \Leftrightarrow x^t PDP^t x = 1 \Leftrightarrow \left(P^t x\right)^t D\left(P^t x\right) = 1$$

La matriz P expresa el cambio de coordenadas de base B a base canónica, es decir $P = C_{BE}$. Por ende $C_{EB} = P^{-1} = P^t$. De esta forma, la expresión $P^t x$ indica las coordenadas del vector x con respecto a la nueva base de autovectores. Adoptamos la notación

$$P^t x = y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$$

con lo que la ecuación original en coordenadas canónicas equivale, *en coordenadas con respecto a la base B*, a la ecuación

$$y^t Dy = 1 \Leftrightarrow y^t \begin{bmatrix} 1/2 & 0 \\ 0 & -1/2 \end{bmatrix} y = 1 \Leftrightarrow \frac{1}{2}y_1^2 - \frac{1}{2}y_2^2 = 1$$

que reconocemos como la ecuación canónica de una hipérbola.

La exposición anterior pretende poner de manifiesto la importancia que tienen las expresiones de la forma $x^t Ax$ y los cambios de coordenadas que admiten. La definición que sigue permitirá generalizar estas ideas.

Definición 5.4

Una *forma cuadrática* en $\mathbb{R}^n$ es una función $Q : \mathbb{R}^n \to \mathbb{R}$ que se puede expresar en la forma

$$Q(x) = x^t Ax$$

siendo $A \in \mathbb{R}^{n\times n}$ una matriz *simétrica*.

Veamos algunos ejemplos de formas cuadráticas.

Ejemplo 5.3

La función que a cada vector de $\mathbb{R}^n$ le asigna el cuadrado de su norma, esto es,

$$Q : \mathbb{R}^n \to \mathbb{R},\ Q(x) = \|x\|^2$$

es una forma cuadrática pues $Q(x) = x^t Ix$.

Ejemplo 5.4

Sea $Q : \mathbb{R}^2 \to \mathbb{R},\ Q(x) = x^t Ax$ con

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

Un sencillo cálculo muestra que $Q(x) = 4x_1^2 + 3x_2^2$. En general, si $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ es una matriz diagonal se tiene que

$$Q(x) = a_{11}x_1^2 + \ldots + a_{nn}x_n^2 = \sum_{1\le i\le n} a_{ii}x_i^2$$

A este tipo de forma cuadrática se la denomina *forma diagonal* o *sin productos cruzados.*

Ejercicio.

5.2 Decida si una forma cuadrática $Q(x) = x^t Ax$ define una transformación lineal de $\mathbb{R}^n \to \mathbb{R}$.

Ejemplo 5.5

Sea $Q : \mathbb{R}^3 \to \mathbb{R}$, $Q(x) = x^t Ax$ con

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

Efectuando el producto $x^t Ax$ obtenemos (¡verificarlo!)

$$Q(x) = 5x_1^2 + 1x_1x_2 - 2x_1x_3 + 1x_2x_1 - 2x_2^2 + 3x_2x_3 - 2x_3x_1 + 3x_3x_2 - 1x_3^2$$

que es la sumatoria de todos los posibles productos x_ix_j multiplicados por el correspondiente coeficiente a_{ij} de la matriz A, es decir

$$Q(x) = \sum_{1\le i,j\le 3} a_{ij}x_ix_j$$

Como la matriz A es simétrica y $x_ix_j = x_jx_i$, podemos reducir la expresión de $Q(x)$ a

$$Q(x) = 5x_1^2 + 2\cdot 1x_1x_2 - 2\cdot 2x_1x_3 - 2x_2^2 + 2\cdot 3x_2x_3 - 1x_3^2$$

con lo cual sólo aparecen términos de la forma x_ix_j con $i \le j$ multiplicados por a_{ii} si $i = j$ y por $2\cdot a_{ij}$ si $i < j$. En otras palabras

$$Q(x) = \sum_{1\le i\le 3} a_{ii}x_i^2 + 2\sum_{1\le i<j\le 3} a_{ij}x_ix_j$$

El argumento del ejemplo anterior se generaliza sin dificultad para dar la expresión general de una forma cuadrática.

Proposición 5.1

Dada una forma cuadrática $Q : \mathbb{R}^n \to \mathbb{R}$, $Q(x) = x^t Ax$ con $A = [a_{ij}] \in \mathbb{R}^{n\times n}$ simétrica, su expresión es

$$Q(x) = \sum_{1\le i\le n} a_{ii}x_i^2 + 2\sum_{1\le i<j\le n} a_{ij}x_ix_j$$

Ejercicio.

5.3

- Obtenga la expresión de la forma cuadrática $Q:\mathbb{R}^3 \to \mathbb{R},\ Q(x) = x^t Ax$ siendo

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

- Dada la forma cuadrática $Q:\mathbb{R}^3 \to \mathbb{R},\ Q(x) = 2x_1^2 + 3x_1x_2 - x_1x_3 + 4x_2^2 - 2x_2x_3 - x_3^2$ encuentre una matriz simétrica $A \in \mathbb{R}^{3\times 3}$ de forma tal que sea $Q(x) = x^t Ax$.

Ejemplo 5.6

Sea una función $Q:\mathbb{R}^2 \to \mathbb{R}$ definida mediante $Q(x) = x^t Bx$ con

$$B = \begin{bmatrix} 5 & 1 \\ -4 & -2 \end{bmatrix}$$

Aparentemente $Q(x)$ no es una forma cuadrática, pues B no es simétrica. Sin embargo sí lo es, ya que

$$Q(x) = 5x_1^2 + x_1x_2 - 4x_2x_1 - 2x_2^2 = 5x_1^2 - 3x_1x_2 - 2x_2^2 = x^t Ax$$

donde

$$A = \begin{bmatrix} 5 & -3/2 \\ -3/2 & -2 \end{bmatrix}$$

El último ejemplo sugiere que cualquier expresión de la forma $x^t Bx$ define una forma cuadrática. Veamos que es efectivamente así: dada una matriz arbitraria $B \in \mathbb{R}^{n\times n}$ definimos

$$Q:\mathbb{R}^n \to \mathbb{R},\ Q(x) = x^t Bx$$

Entonces

$$x^t Bx = \left(x^t Bx\right)^t = x^t B^t x \Rightarrow 2Q(x) = x^t Bx + x^t B^t x = x^t \left(B + B^t\right) x$$

y por lo tanto

$$Q(x) = x^t \left(\frac{B + B^t}{2}\right) x$$

de modo que $Q(x)$ es una forma cuadrática, porque $\left(\frac{B+B^t}{2}\right)$ es una matriz simétrica.

Ahora que ya hemos establecido las definiciones básicas, mostraremos con un nuevo ejemplo que mediante un cambio de variable adecuado es posible transformar una forma cuadrática con productos cruzados en una forma diagonal.

Ejemplo 5.7

Consideramos la forma cuadrática $Q:\mathbb{R}^2 \to \mathbb{R},\ Q(x) = 2x_1^2 - 2x_1x_2 + 2x_2^2$. Es decir,

$$Q(x) = x^t \underbrace{\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}}_{A} x$$

Podemos realizar un cambio de variable adecuado para que, en coordenadas con respecto al nuevo sistema de referencia, no tenga términos de producto cruzado. Para ello, realizamos una diagonalización ortogonal de la matriz A:

$$A = \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_{P} \cdot \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}}_{D} \cdot \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{P^t} \tag{5.6}$$

Luego la expresión de $Q(x)$ puede escribirse como

$$Q(x) = x^t PDP^t x = \left(P^t x\right)^t D\left(P^t x\right)$$

y con el cambio de variables $y = P^t x$ equivale a

$$\tilde{Q}(y) = y^t Dy = y_1^2 + 3y_2^2$$

Como la diagonalización ortogonal puede aplicarse a cualquier matriz simétrica, el método aplicado en el ejemplo anterior tiene generalidad.

Teorema 5.5 (de los ejes principales)

Sea $A \in \mathbb{R}^{n\times n}$ una matriz simétrica. Entonces existe un cambio ortogonal de variable $x = Py$ (con $P \in \mathbb{R}^{n\times n}$ ortogonal) que cambia la forma cuadrática $Q(x) = x^t Ax$ en una sin productos cruzados, es decir,

$$Q(x) = \tilde{Q}(y) = y^t Dy$$

con $D \in \mathbb{R}^{n\times n}$ diagonal. Además, los elementos de la diagonal de D son los autovalores de A y las columnas de la matriz P son autovectores de A y forman una base ortonormal de $\mathbb{R}^n$.

Demostración:
Con el objeto de simplificar la expresión de $Q(x)$, es decir, de eliminar los términos de la forma $x_i x_j$ con $i \neq j$, consideramos el cambio de variable $x = Py$ con P inversible:

$$Q(x) = x^t Ax = (Py)^t A(Py) = y^t (P^t AP)y = \tilde{Q}(y)$$

Como queremos que $\tilde{Q}(y)$ esté libre de productos cruzados, debemos elegir P de modo tal que la matriz $D = P^t AP$ sea diagonal. Esto siempre es posible dado que A, por ser simétrica y de acuerdo con el teorema espectral 5.4, es diagonalizable ortogonalmente. Es decir, existen P ortogonal y D diagonal tales que $A = PDP^t$. Entonces, tomando tal P (cuyas columnas son autovectores de A y forman una base ortonormal de $\mathbb{R}^n$) y efectuando el cambio de variable $x = Py$, tenemos que

$$Q(x) = x^t\left(PDP^t\right)x = (P^t x)^t D(P^t x) = y^t Dy = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 = \tilde{Q}(y)$$

siendo $\lambda_1, \ldots, \lambda_n$ los autovalores (todos reales) de A. □

Al realizar el cambio de variables $x = Py$ para eliminar los productos cruzados, hemos adoptado una base ortonormal $B = \{b_1, b_2, ..., b_n\}$ (y por ende un sistema de ejes) donde la variable y representa las coordenadas de x ($y = [x]_B$) de modo tal que la expresión de $Q(x)$ en términos de esas coordenadas carece de productos cruzados. Por esta razón a las rectas que generan los vectores de la base B se las denomina *ejes principales* de la forma cuadrática.

N Para construir la matriz de cambio de base P, lo usual es ordenar la base B de modo que resulte $\det P = 1$, pues de esa manera el sistema de ejes ortogonales definidos por las columnas de P se obtiene rotando el sistema de ejes cartesiano original.

5.4 Conjuntos de nivel

Comenzamos esta sección retomando el último ejemplo, para ilustrar una aplicación del teorema de los ejes principales 5.5.

Ejemplo 5.8

Consideramos nuevamente la forma cuadrática $Q : \mathbb{R}^2 \to \mathbb{R},\ Q(x) = 2x_1^2 - 2x_1x_2 + 2x_2^2$ del ejemplo 5.7. Por ser una función de $\mathbb{R}^2 \to \mathbb{R}$, su gráfico es una superficie en $\mathbb{R}^3$, como se aprecia en la figura 5.2a. En este contexto, la ecuación

$$2x_1^2 - 2x_1x_2 + 2x_2^2 = 4 \Leftrightarrow Q(x) = 4 \tag{5.7}$$

puede leerse como la pregunta "¿Cuáles son los elementos $[x_1\ x_2]^t$ del dominio de Q cuya imagen es 4?". De acuerdo con el teorema de los ejes principales 5.5, y como ya vimos en el ejemplo 5.7, podemos realizar un cambio de variable adecuado para que la ecuación (5.7), en coordenadas con respecto al nuevo sistema de referencia, no tenga términos de producto cruzado. Elegimos entonces una base $B = \{b_1, b_2\}$ formada por las columnas de la matriz P con que diagonalizamos A en (5.6):

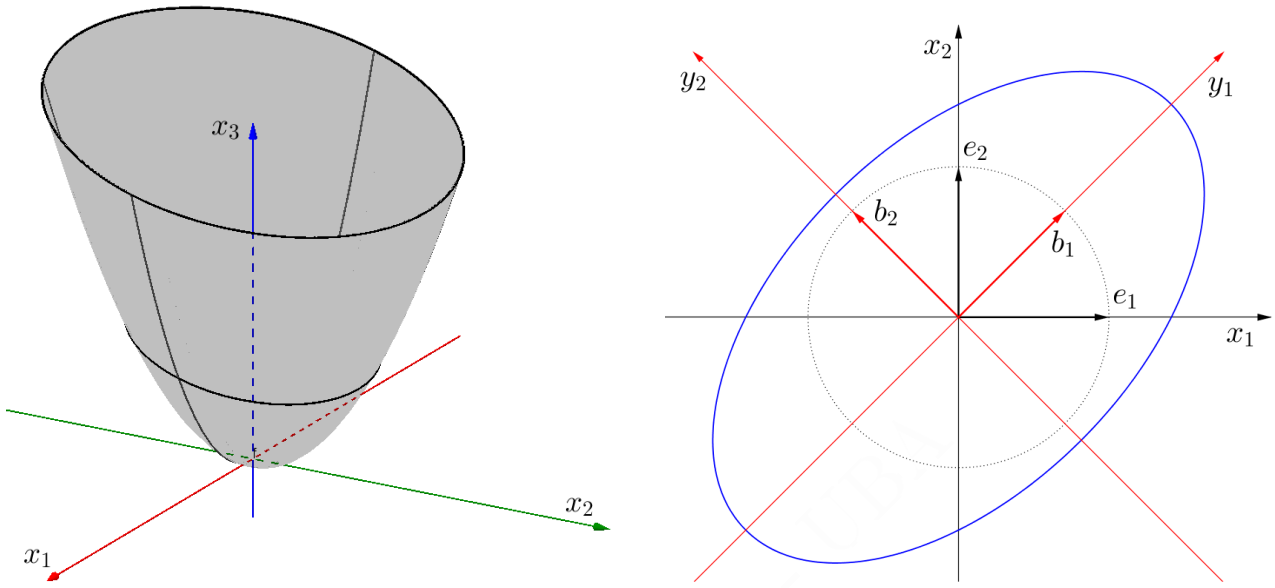
$$B = \left\{ \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix}, \begin{bmatrix} -1/\sqrt{2} \\ 1/\sqrt{2} \end{bmatrix} \right\}$$

Es decir que, considerando coordenadas $[y_1\ y_2]^t$ con respecto a la base B, los puntos que buscamos deben verificar la ecuación

$$Q(x) = 4 \Leftrightarrow \tilde{Q}(y) = 4 \Leftrightarrow y_1^2 + 3y_2^2 = 4 \Leftrightarrow \frac{y_1^2}{2^2} + \frac{y_2^2}{\left(\frac{2}{\sqrt{3}}\right)^2} = 1$$

En el nuevo sistema de referencia, resulta claro que los puntos de $\mathbb{R}^2$ que verifican la restricción constituyen una elipse, como puede verse en la figura 5.2b.

Destacamos que la nueva base se obtiene mediante una rotación de la base canónica, porque $\det P = 1$.



(a) Gráfico de la función Q

(b) Conjunto de nivel $Q(x) = 4$

Figura 5.2: Forma cuadrática $Q : \mathbb{R}^2 \to \mathbb{R}$

Ejercicio.

5.4 Considere la ecuación $2x_1^2 - 2x_1x_2 + 2x_2^2 = k$ y describa los conjuntos solución para distintos valores de $k \in \mathbb{R}$.

En el ejemplo anterior hemos indagado el conjunto de puntos del dominio para los que una forma cuadrática alcanza cierto valor dado: este conjunto tiene nombre propio.

Definición 5.5

Dada una forma cuadrática $Q : \mathbb{R}^n \to \mathbb{R}$ y dado un número $k \in \mathbb{R}$, el *conjunto de nivel k* es el conjunto de puntos de $\mathbb{R}^n$ para los cuales Q toma el valor k. Adoptando la notación $\mathcal{N}_k(Q)$, resulta

$$\mathcal{N}_k(Q) = \{x \in \mathbb{R}^n : Q(x) = k\}$$

N Aunque nuestra atención se centra en las formas cuadráticas, el concepto de conjunto de nivel puede extenderse sin modificaciones a cualquier función de $\mathbb{R}^n \to \mathbb{R}$.

Ejercicio.

5.5 Dada la forma cuadrática $Q : \mathbb{R}^2 \to \mathbb{R}$, $Q(x) = x_1^2 - 8x_1x_2 - 5x_2^2$, estudie y describa $\mathcal{N}_1(Q)$, $\mathcal{N}_{-1}(Q)$ y $\mathcal{N}_0(Q)$.

5.5 Signo de una forma cuadrática

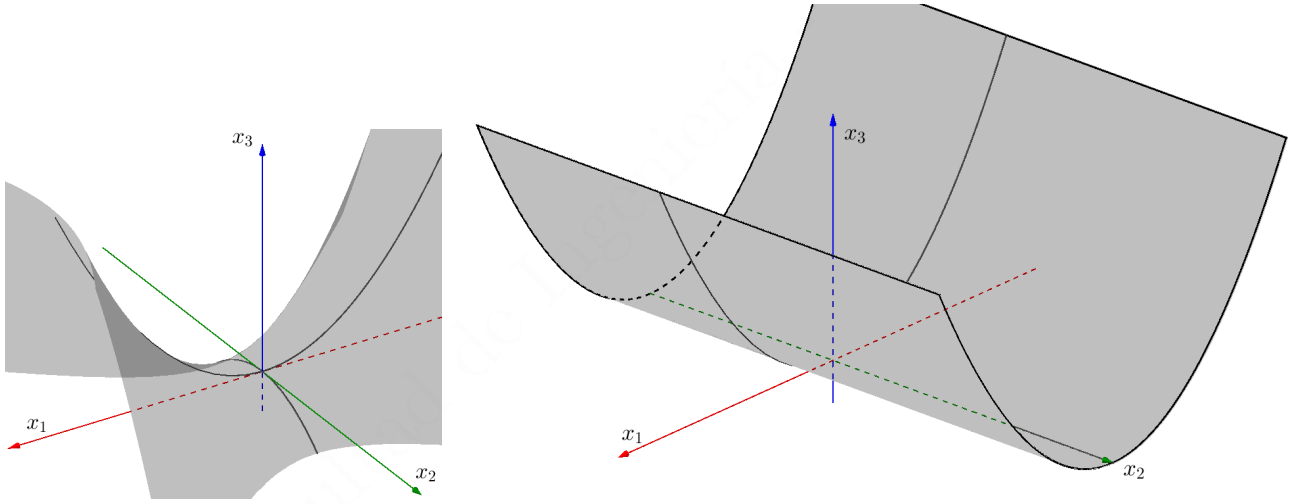
La forma cuadrática definida en el ejemplo 5.8 tiene un gráfico que se aprecia en la figura 5.2a: queremos destacar que para cualquier elemento del dominio $\mathbb{R}^2$, excepto el vector nulo, el valor de la función es positivo. Si en cambio consideramos la forma cuadrática

$$Q_1(x) = x_1^2 - x_2^2 \tag{5.8}$$

cuyo gráfico se ilustra en la figura 5.3a, en cualquier entorno del $(0,0)$ existen puntos para los que es positiva (por ejemplo los $(x_1,0)$ con $x_1 \neq 0$); pero también existen puntos para los que es negativa (por ejemplo los $(0,x_2)$ con $x_2 \neq 0$). Si ahora definimos la forma cuadrática

$$Q_2(x) = x_1^2 \tag{5.9}$$

podemos apreciar en su gráfico de la figura 5.3b que para cualquier punto es $Q_2(x) \geq 0$. Además, la función vale 0 en todos los puntos del eje x_2.



(a) Gráfico de Q_1 **(b)** Gráfico de Q_2

Figura 5.3: Formas cuadráticas y su signo

A continuación, clasificamos las formas cuadráticas según el signo que adoptan.

Definición 5.6

Dada una forma cuadrática $Q : \mathbb{R}^n \to \mathbb{R}$, decimos que es

1. *definida positiva* si para todo $x \in \mathbb{R}^n - \{0\} : Q(x) > 0$
2. *semidefinida positiva* si para todo $x \in \mathbb{R}^n : Q(x) \geq 0$
3. *definida negativa* si para todo $x \in \mathbb{R}^n - \{0\} : Q(x) < 0$
4. *semidefinida negativa* si para todo $x \in \mathbb{R}^n : Q(x) \leq 0$
5. *indefinida* si no es ninguna de las anteriores. Es decir, si existen vectores $u, v \in \mathbb{R}^n$ tales que $Q(u) > 0$ y $Q(v) < 0$

De acuerdo con esta clasificación, la forma cuadrática graficada en 5.2a es definida positiva, mientras que la graficada en 5.3a es indefinida. En cuanto a la graficada en 5.3b, resulta semidefinida positiva.

Ejercicio.

5.6 Clasifique las siguientes formas cuadráticas.

- $Q:\mathbb{R}^3 \to \mathbb{R},\ Q(x) = x_1^2 + 2x_2^2 + 3x_3^2$
- $Q:\mathbb{R}^3 \to \mathbb{R},\ Q(x) = x_1^2 + 2x_2^2$
- $Q:\mathbb{R}^4 \to \mathbb{R},\ Q(x) = x_1^2 + 2x_2^2 - 3x_3^2$
- $Q:\mathbb{R}^3 \to \mathbb{R},\ Q(x) = -x_1^2 - 2x_2^2 - 3x_3^2$
- $Q:\mathbb{R}^3 \to \mathbb{R},\ Q(x) = -x_1^2 - 2x_2^2$

Definición 5.7

Dada una matriz simétrica $A \in \mathbb{R}^{n\times n}$, decimos que es *definida positiva* si la forma cuadrática asociada $Q(x) = x^t Ax$ es definida positiva. Análogamente se define matriz *definida negativa, indefinida*, etc.

Cuando una forma cuadrática $Q:\mathbb{R}^n \to \mathbb{R}$ carece de productos cruzados, es decir, cuando $Q(x) = x^t Dx$ con

$$D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

ya vimos en el ejemplo 5.4 que

$$Q(x) = \lambda_1 x_1^2 + \ldots + \lambda_n x_n^2 = \sum_{1\le i\le n} \lambda_i x_i^2$$

En estos casos, la clasificación es sencilla:

1. definida positiva si $\lambda_i > 0$ para todo $i = 1, \ldots, n$.
2. semidefinida positiva si $\lambda_i \ge 0$ para todo $i = 1, \ldots, n$.
3. definida negativa si $\lambda_i < 0$ para todo $i = 1, \ldots, n$.
4. semidefinida negativa si $\lambda_i \le 0$ para todo $i = 1, \ldots, n$.
5. indefinida si existen índices i, j tales que $\lambda_i > 0$ mientras que $\lambda_j < 0$.

Si $Q(x) = x^t Ax$ tiene productos cruzados, es decir, si A no es una matriz diagonal, empleando un cambio de variables ortogonal $x = Py$ que elimine productos cruzados, tenemos que

$$Q(x) = y^t Dy = \tilde{Q}(y)$$

con D diagonal. Como Q y $\tilde{Q}$ tienen el mismo signo, y el signo de esta última está determinado por los elementos en la diagonal de D, que son los autovalores de A, tenemos el siguiente resultado.

Teorema 5.6

Sea $A \in \mathbb{R}^{n\times n}$ una matriz simétrica y sean $\lambda_1, \lambda_2, ..., \lambda_n$ sus autovalores (son todos números reales). Entonces la forma cuadrática $Q(x) = x^t Ax$ es

1. *definida positiva* si y sólo si $\lambda_i > 0$ para todo $i = 1, ..., n$.
2. *semidefinida positiva* si y sólo si $\lambda_i \geq 0$ para todo $i = 1, ..., n$.
3. *definida negativa* si y sólo si $\lambda_i < 0$ para todo $i = 1, ..., n$.
4. *semidefinida negativa* si y sólo si $\lambda_i \leq 0$ para todo $i = 1, ..., n$.
5. *indefinida* si existe un par de índices i, j tales que $\lambda_i > 0$ y $\lambda_j < 0$.

Ejercicio.

5.7 Clasifique la forma cuadrática $Q(x) = x^t Ax$ con $A = \begin{bmatrix} 2 & 3 & 0 \\ 3 & 2 & 4 \\ 0 & 4 & 2 \end{bmatrix}$

Ejercicio.

5.8 Determine para qué valores de c la forma cuadrática $Q : \mathbb{R}^2 \to \mathbb{R}, Q(x) = x_1^2 + 2cx_1x_2 + x_2^2$ es definida positiva.

5.6 Optimización de formas cuadráticas con restricciones

En la sección anterior presentamos a las formas cuadráticas como funciones definidas en todo $\mathbb{R}^n$. Dedicaremos esta sección a estudiar las formas cuadráticas y sus extremos cuando están restringidas a tomar valores en un subconjunto de $\mathbb{R}^n$.

Ejemplo 5.9

Retomamos la forma cuadrática $Q_1(x)$ definida en (5.8) cuyo gráfico, como función con dominio $\mathbb{R}^2$, está en la figura 5.3a. Evidentemente, la función no alcanza extremos absolutos, puesto que

- Para puntos de la forma $(x_1, 0)$ la imagen es x_1^2 y puede hacerse arbitrariamente positivo. Luego no hay máximo absoluto.
- Para puntos de la forma $(0, x_2)$ la imagen es $-x_2^2$ y puede hacerse arbitrariamente negativo. Luego no hay mínimo absoluto.

Si en cambio consideramos que la forma cuadrática sólo se evalúa en los puntos de la circunferencia de centro $(0, 0)$ y radio uno, es decir si restringimos el dominio al conjunto

$$\tilde{D} = \{x \in \mathbb{R}^2 : \|x\|^2 = 1\}$$

el gráfico correspondiente es el de la figura 5.4. Podemos ver que la función alcanzará extremos sobre puntos de la

circunferencia.[a] En efecto, para cualquier $x = (x_1, x_2)$ se verifica que

$$-\|x\|^2 \leq -x_1^2 - x_2^2 \leq \underbrace{x_1^2 - x_2^2}_{Q_1(x)} \leq x_1^2 + x_2^2 \leq \|x\|^2$$

y teniendo en cuenta la restricción $x \in \tilde{D}$, resulta

$$-1 \leq Q_1(x) \leq 1$$

Esto nos permite afirmar que la forma cuadrática, restringida al dominio $\tilde{D}$, está acotada. Veremos que además existen vectores de $\tilde{D}$ sobre los cuales se realizan efectivamente los valores extremos -1 y 1. Para probarlo, planteamos

$$\begin{aligned} Q_1(x) = 1 \ \wedge \ \|x\|^2 = 1 &\Leftrightarrow x_1^2 - x_2^2 = 1 \ \wedge \ x_1^2 + x_2^2 = 1 \\ &\Leftrightarrow x_1^2 - x_2^2 = x_1^2 + x_2^2 \ \wedge \ x_1^2 + x_2^2 = 1 \\ &\Leftrightarrow x_2^2 = 0 \ \wedge \ x_1^2 + x_2^2 = 1 \\ &\Leftrightarrow x_2 = 0 \ \wedge \ |x_1| = 1 \end{aligned}$$

En este caso, los vectores de la restricción $\tilde{D}$ sobre los que la forma realiza un valor máximo son $x = \pm(1, 0)$. Dejamos como ejercicio probar que el mínimo sujeto a la restricción se realiza sobre los vectores $x = \pm(0, 1)$.

Destacamos que la forma cuadrática es $Q_1(x) = x^t Ax$ con

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

y que los valores máximo y mínimo que alcanza sujeta a la restricción $\|x\| = 1$ coinciden con los autovalores de A. Más aun, dichos extremos se alcanzan sobre los autovectores de norma uno.

[a] Un argumento propio del análisis matemático es que una función continua realiza extremos absolutos cuando su dominio es un subconjunto cerrado y acotado de $\mathbb{R}^n$. De todos modos, para el caso de las formas cuadráticas lo demostraremos por medios algebraicos.

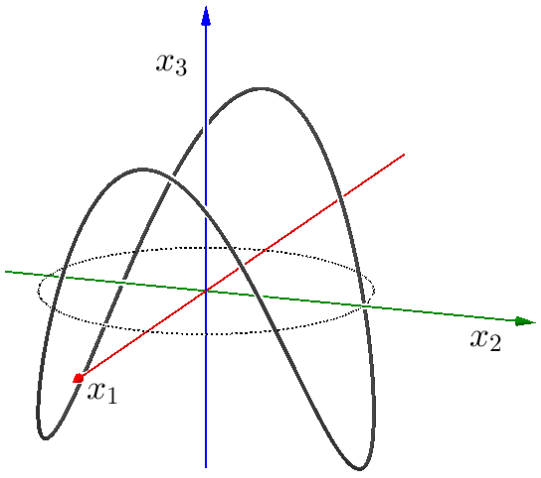


Figura 5.4: Forma cuadrática con dominio restringido a una circunferencia

Los argumentos del último ejemplo pueden generalizarse a formas cuadráticas sujetas a una restricción de la clase $\|x\|^2 = k$ (donde $k > 0$).

Teorema 5.7 (desigualdad de Rayleigh)

Sean $A \in \mathbb{R}^{n\times n}$ una matriz simétrica y $Q(x) = x^t Ax$ la forma cuadrática que define. Sean λ_M y λ_m los autovalores máximo y mínimo de A respectivamente, con S_{λ_M} y S_{λ_m} los autoespacios asociados. Entonces para todo $x \in \mathbb{R}^n$ se verifica que

$$\lambda_m \|x\|^2 \leq Q(x) \leq \lambda_M \|x\|^2$$

Además, $Q(x) = \lambda_m \|x\|^2$ si y sólo si $x \in S_{\lambda_m}$, mientras que $Q(x) = \lambda_M \|x\|^2$ si y sólo si $x \in S_{\lambda_M}$.

Demostración:
Probaremos la segunda desigualdad y que la forma cuadrática realiza su máximo en los autovectores asociados al autovalor máximo. La otra desigualdad es completamente análoga y se deja como ejercicio.

Por ser A una matriz simétrica, sabemos que existen una matriz ortogonal P y una matriz diagonal D tales que $A = PDP^t$. Además, las columnas de P forman una base ortonormal de autovectores de A y los elementos de la diagonal de D son los autovalores respectivos. Elegimos presentar los n autovalores de A en forma decreciente, denotando por μ a la multiplicidad (algebraica y geométrica) de λ_M :

$$\lambda_M = \lambda_1 = \ldots = \lambda_\mu > \lambda_{\mu+1} \geq \ldots \geq \lambda_n$$

De esta forma, denotando $P = \begin{bmatrix} v_1 & \cdots & v_n \end{bmatrix}$ resulta el autoespacio

$$S_{\lambda_M} = \text{gen}\{v_1, \ldots, v_\mu\}$$

Con el cambio de variable $x = Py$, la forma cuadrática adquiere la expresión

$$Q(x) = \tilde{Q}(y) = y^t Dy = \lambda_1 y_1^2 + \ldots + \lambda_n y_n^2$$

y siendo $\lambda_i \leq \lambda_M$ para todo $i = 1, \ldots, n$, se deduce la desigualdad

$$Q(x) \leq \lambda_M y_1^2 + \ldots + \lambda_M y_n^2 = \lambda_M \left(y_1^2 + \ldots + y_n^2\right) = \lambda_M \left\|y\right\|^2$$

válida para todo $x \in \mathbb{R}^n$. Como P es una matriz ortogonal, de acuerdo con el teorema 5.2, es $\|Px\| = \|x\|$, luego $\left\|y\right\| = \|x\|$ y obtenemos la desigualdad del enunciado:

$$Q(x) \leq \lambda_M \|x\|^2$$

Ahora buscamos los $x \in \mathbb{R}^n$ que realizan la igualdad:

$$Q(x) = \lambda_M \|x\|^2$$

Teniendo en cuenta el cambio de variable $x = Py$, esto equivale a encontrar los $y \in \mathbb{R}^n$ que verifican

$$\lambda_1 y_1^2 + \ldots + \lambda_n y_n^2 = \lambda_M \left\|y\right\|^2 = \lambda_M \left(y_1^2 + \ldots + y_n^2\right)$$

de donde

$$\underbrace{\lambda_M y_1^2 + \ldots + \lambda_M y_\mu^2}_{\mu \text{ términos}} + \lambda_{\mu+1} y_{\mu+1}^2 + \ldots + \lambda_n y_n^2 = \lambda_M y_1^2 + \ldots + \lambda_M y_n^2$$

y cancelando se obtiene que debe ser

$$\left(\lambda_M - \lambda_{\mu+1}\right) y_{\mu+1}^2 + \ldots + \left(\lambda_M - \lambda_n\right) y_n^2 = 0$$

Los factores $(\lambda_M - \lambda_i)$ de la última expresión son todos positivos (¿por qué?) luego se deduce que debe ser

$$y_{\mu+1} = \ldots = y_n = 0$$

y los vectores que realizan el máximo tienen coordenadas de la forma

$$y = \begin{bmatrix} y_1 & \cdots & y_\mu & 0 & \cdots & 0 \end{bmatrix}^t$$

con lo cual los x que verifican $Q(x) = \lambda_M \|x\|^2$ surgen de

$$x = Py = P \begin{bmatrix} y_1 & \cdots & y_\mu & 0 & \cdots & 0 \end{bmatrix}^t = y_1 v_1 + \ldots + y_\mu v_\mu$$

Como $\{v_1, \ldots, v_\mu\}$ es una base de S_{λ_M}, hemos demostrado que $Q(x) = \lambda_M \|x\|^2$ si y sólo si $x \in S_{\lambda_M}$. □

Ejemplo 5.10

Consideremos la forma cuadrática $Q(x) = x^t Ax$ con

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

Su polinomio característico es $p(t) = -(t+2)(t-4)^2$ con lo cual los autovalores máximo y mínimo son respectivamente $\lambda_M = 4$ y $\lambda_m = -2$. Por lo tanto y de acuerdo con el teorema anterior, los valores extremos que alcanza Q sujeta a la restricción $\|x\| = 1$ son

$$\max_{\|x\|=1} Q(x) = 4 \qquad \text{y} \qquad \min_{\|x\|=1} Q(x) = -2$$

Para hallar los maximizantes y minimizantes, es decir los vectores sobre los que se realizan los extremos, debemos considerar los autoespacios asociados a los autovalores

$$S_{\lambda_M} = \text{gen}\{[1\ 1\ 1]^t, [1\ -1\ 0]^t\} \qquad \text{y} \qquad S_{\lambda_m} = \text{gen}\{[1\ 1\ -2]^t\}$$

Entonces $Q(x)$ sujeto a la restricción $\|x\| = 1$ alcanza el mínimo en los vectores unitarios de S_{λ_m}, que en este caso son sólo dos

$$v_1 = \begin{bmatrix} \frac{1}{\sqrt{6}} & \frac{1}{\sqrt{6}} & -\frac{2}{\sqrt{6}} \end{bmatrix}^t \qquad \text{y} \qquad v_2 = -v_1 = \begin{bmatrix} -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{6}} & \frac{2}{\sqrt{6}} \end{bmatrix}^t$$

Con respecto a los maximizantes, son los vectores unitarios de S_{λ_M}, es decir los de la forma

$$x = \alpha [1\ 1\ 1]^t + \beta [1\ -1\ 0]^t \text{ con } \|x\| = 1$$

Para obtener una expresión algo más explícita de los maximizantes, es conveniente trabajar con una base ortonormal $\{v_1, v_2\}$ del autoespacio, ya que entonces $\|\alpha_1 v_1 + \alpha_2 v_2\|^2 = \alpha_1^2 + \alpha_2^2$. Con esto resulta que los x que maximizan Q sujeta a la restricción son aquellos de la forma

$$x = \alpha_1 v_1 + \alpha_2 v_2 \text{ con } \alpha_1^2 + \alpha_2^2 = 1$$

Como en nuestro caso una base ortonormal de S_{λ_M} es

$$\left\{ \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}^t, \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}^t \right\}$$

los maximizantes son los $x \in \mathbb{R}^3$ de la forma:

$$x = \alpha_1 \begin{bmatrix} \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} & \frac{1}{\sqrt{3}} \end{bmatrix}^t + \alpha_2 \begin{bmatrix} \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}^t \text{ con } \alpha_1^2 + \alpha_2^2 = 1$$

Notamos que en este caso hay un número infinito de maximizantes y que constituyen la circunferencia de radio uno centrada en el origen y contenida en el plano S_{λ_M}.

Presentamos ahora un ejemplo en que la restricción no es de la forma $\|x\|^2 = k$.

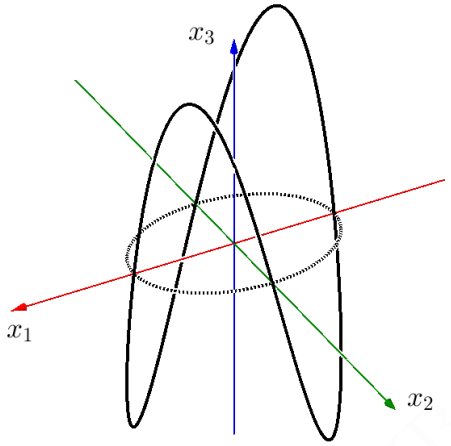


Figura 5.5: Forma cuadrática con dominio restringido a una elipse

Ejemplo 5.11

Buscamos los extremos de la forma cuadrática $Q : \mathbb{R}^2 \to \mathbb{R},\ Q(x) = 4x_1x_2$ sujeta a la restricción

$$x_1^2 + 2x_2^2 = 2$$

Vale decir que debemos encontrar los extremos que realiza la función Q cuando su dominio se restringe a la elipse de ecuación

$$\left(\frac{x_1}{\sqrt{2}}\right)^2 + x_2^2 = 1 \tag{5.10}$$

como se aprecia en la figura 5.5.

Para poder aplicar el teorema de Rayleigh 5.7, señalamos que el cambio de variable

$$y_1 = \frac{x_1}{\sqrt{2}} \qquad y_2 = x_2$$

permite reescribir la restricción (5.10) como

$$y_1^2 + y_2^2 = 1 \tag{5.11}$$

y la forma cuadrática como

$$Q(x) = \tilde{Q}(y) = 4\sqrt{2}y_1y_2 \tag{5.12}$$

Entonces $\tilde{Q}(y)$ es una forma cuadrática sujeta a una restricción (5.11) que sí verifica las condiciones del teorema de Rayleigh 5.7. Para aplicarlo, reescribimos la forma cuadrática (5.12)

$$\tilde{Q}(y) = y^t \underbrace{\begin{bmatrix} 0 & 2\sqrt{2} \\ 2\sqrt{2} & 0 \end{bmatrix}}_{A} y$$

Es un cálculo de rutina ver que los autovalores de A son $\lambda_M = 2\sqrt{2}$ y $\lambda_m = -2\sqrt{2}$ con autoespacios asociados

$$S_{\lambda_M} = \text{gen}\{[1\ 1]^t\} \qquad S_{\lambda_m} = \text{gen}\{[-1\ 1]^t\}$$

Luego el valor máximo de $\tilde{Q}(y)$ sujeto a la restricción $\|y\|^2 = 1$ es $2\sqrt{2}$ y se realiza en los vectores de coordenadas $y = \pm\left[1/\sqrt{2}\ \ 1/\sqrt{2}\right]^t$. Para conocer los vectores originales, debemos deshacer el cambio de coordenadas: entonces el valor máximo de $Q(x)$ sujeta a la restricción (5.10) es $2\sqrt{2}$ y se realiza en los vectores $x = \pm\left[1\ \ 1/\sqrt{2}\right]^t$ (véase el gráfico). Se puede comprobar que estos vectores verifican la ecuación de la elipse y al evaluar $Q(x)$ en ellos se obtiene el valor predicho.

Dejamos como ejercicio probar que el valor mínimo es $-2\sqrt{2}$ y se realiza en los vectores $x = \pm\left[-1\ 1/\sqrt{2}\right]^t$.

Ejercicio.

5.9 Encuentre el valor mínimo que realiza la forma cuadrática $Q(x) = \frac{4}{9}x_1^2 - \frac{80}{9}x_1x_2 - \frac{125}{9}x_2^2$ sujeta a la restricción $4x_1^2 + 25x_2^2 = 9$. Indique sobre qué vectores se alcanza dicho mínimo.

Otra dificultad con que podríamos encontrarnos es el caso en que la restricción tiene términos de producto cruzado. En el ejemplo que sigue presentamos dos estrategias de resolución, la segunda de gran generalidad.

Ejemplo 5.12

Buscamos los extremos de la forma cuadrática $Q : \mathbb{R}^2 \to \mathbb{R},\ Q(x) = x_1^2 + x_2^2$ sujeta a la restricción

$$2x_1^2 - 2x_1x_2 + 2x_2^2 = 4$$

La forma cuadrática es el cuadrado de la norma, mientras que la restricción ya fue estudiada en el ejemplo 5.8 y graficada en la figura 5.2b. Por lo tanto, el problema tiene una interpretación geométrica inmediata: se trata de encontrar los puntos de la elipse más próximos y más alejados del origen.

En primer lugar, y para no confundir con la forma $Q(x) = \|x\|^2$, caracterizamos a la restricción mediante

$$R(x) = 2x_1^2 - 2x_1x_2 + 2x_2^2 = x^t \underbrace{\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}}_{B} x$$

Como en el ejemplo 5.8, realizamos una diagonalización ortogonal de la matriz B:

$$B = \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}}_{P} \cdot \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}}_{D} \cdot \underbrace{\frac{1}{\sqrt{2}}\begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{P^t}$$

Luego la restricción con el cambio de variables $y = P^t x$ equivale a

$$y_1^2 + 3y_2^2 = 4 \Leftrightarrow \frac{y_1^2}{2^2} + \frac{y_2^2}{\left(\frac{2}{\sqrt{3}}\right)^2} = 1 \tag{5.13}$$

Los puntos de la elipse cuya distancia al origen es máxima tienen, en el nuevo sistema de referencia, coordenadas $\pm[2\ 0]^t$. De acuerdo con el cambio de coordenadas $x = Py$, la máxima distancia al origen se realiza en los puntos

$$x = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}\begin{bmatrix} \pm 2 \\ 0 \end{bmatrix} = \pm\frac{1}{\sqrt{2}}\begin{bmatrix} 2 \\ 2 \end{bmatrix} = \pm\begin{bmatrix} \sqrt{2} \\ \sqrt{2} \end{bmatrix}$$

El valor que alcanza la forma cuadrática sobre estos puntos es $Q(x) = 4$ (verifíquelo) porque Q es *el cuadrado* de la norma. De manera análoga se puede comprobar que en los puntos $x = \pm\left[-\sqrt{\frac{2}{3}}\ \sqrt{\frac{2}{3}}\right]^t$ se realiza el valor mínimo sobre la elipse y es $Q(x) = \frac{4}{3}$.

Presentamos a continuación una alternativa que tiene mayor generalidad, puesto que se puede aplicar a problemas sin una interpretación geométrica tan inmediata (por ejemplo, aquellos en que $Q(x)$ no es el cuadrado de la norma). En este problema la dificultad radica en que la restricción no es de la forma $x_1^2 + x_2^2 = 1$ que ya sabemos manejar. Lo que buscamos entonces es un cambio de variable que transforme nuestra restricción en una de la forma estándar. Se propone entonces un segundo cambio de variable a partir de (5.13) que permita escribir la restricción en la forma $\|u\|^2 = 1$. Dado que

$$\left(\frac{y_1}{2}\right)^2 + \left(\frac{y_2}{\left(\frac{2}{\sqrt{3}}\right)}\right)^2 = 1$$

la propuesta es $u_1 = \frac{y_1}{2}$, $u_2 = \frac{y_2}{\left(\frac{2}{\sqrt{3}}\right)}$, con la cual la restricción se reduce a

$$u_1^2 + u_2^2 = 1$$

Despejando $y_1 = 2u_1$, $y_2 = \frac{2}{\sqrt{3}}u_2$, el cambio de variable es, matricialmente,

$$y = \underbrace{\begin{bmatrix} 2 & 0 \\ 0 & \frac{2}{\sqrt{3}} \end{bmatrix}}_{M} u$$

de modo que $x = Py = PMu$ y la forma cuadrática se puede reescribir como

$$Q(x) = x^t \underbrace{\begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}}_{A} x = (PMu)^t A (PMu) = u^t \underbrace{M^t P^t A P M}_{S} u$$

Entonces se deben hallar los extremos de $u^t S u$ sujeta a la restricción $\|u\|^2 = 1$. Dado que

$$S = \begin{pmatrix} 4 & 0 \\ 0 & 4/3 \end{pmatrix}$$

En este caso, y por haber resultado una matriz diagonal, podemos leer inmediatamente sus autovalores y autovectores. La desigualdad de Rayleigh 5.7 nos permite afirmar, una vez más, que

- El máximo es 4 y se alcanza para $u = \pm[1\ 0]^t$, es decir para $x = PMu = \pm\left[\sqrt{2}\ \sqrt{2}\right]^t$
- El mínimo es $\frac{4}{3}$ y se alcanza para $u = \pm[0\ 1]^t$, es decir para $x = PMu = \pm\left[-\sqrt{\frac{2}{3}}\ \sqrt{\frac{2}{3}}\right]^t$

Ejercicio.

5.10 Hallar los puntos de la curva de ecuación $6x_1^2 + 4x_1x_2 + 6x_2^2 = 1$ cuya distancia al origen es mínima y aquellos cuya distancia es máxima.

N Los recursos utilizados en el ejemplo anterior pueden generalizarse para demostrar que dada una forma cuadrática $Q(x) = x^t Ax$, sujeta a la restricción $x^t Bx = k$, siempre es posible hallar los extremos *cuando B es definida positiva.*

El ejemplo que sigue ilustra las dificultades que pueden ocurrir en otros casos.

Ejemplo 5.13

Buscamos el máximo de la forma cuadrática $Q : \mathbb{R}^2 \to \mathbb{R}$, $Q(x) = x_1^2 + x_2^2$ sujeta a la restricción $x_1 x_2 = 1$. La forma cuadrática es el cuadrado de la norma, mientras que la restricción ya fue estudiada al comienzo de la sección 5.3 y graficada en la figura 5.1. Se trata de encontrar los puntos de una hipérbola más alejados del origen. Sin embargo, ya se vio cómo un cambio ortogonal de variables $x = Py$ transforma a la restricción en

$$\frac{1}{2}y_1^2 - \frac{1}{2}y_2^2 = 1$$

y es claro que pueden elegirse vectores de coordenadas $y = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^t$ de modo tal que $\|y\| = \|x\|$ sea arbitrariamente grande. De esta forma, no existe un máximo para la norma restringida a la hipérbola, como el gráfico permitía anticipar.

Ejercicio.

5.11 Encuentre los extremos que realiza la forma cuadrática $Q(x) = x_1 x_2$ sujeta a la restricción $2x_1^2 - 2x_1 x_2 + 2x_2^2 = 4$. Indique sobre qué vectores se alcanzan dichos extremos.

Ejercicio.

5.12 Considere la forma cuadrática $Q : \mathbb{R}^2 \to \mathbb{R}, Q(x) = 9x_1^2 - 6x_1 x_2 + x_2^2$.

- Describa los conjuntos de nivel.
- Clasifique el signo (definida positiva, semidefinida,...).
- Encuentre los valores máximo y mínimo que alcanza sujeta a la restricción $\|x\|^2 = 1$. Indique en qué vectores alcanza dichos extremos.

Esta página fue dejada intencionalmente en blanco

6 Descomposición en valores singulares

6.1 Introducción

Como en la última sección del capítulo anterior, consideramos el problema de conocer los valores extremos que alcanza una forma cuadrática sujeta a una restricción. En este caso, comenzamos por definir una trasformación lineal mediante

$$T : \mathbb{R}^2 \to \mathbb{R}^3, T(x) = Ax$$

donde

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & -1 \end{bmatrix} \tag{6.1}$$

Consideraremos la forma cuadrática que calcula, para cada vector $x \in \mathbb{R}^2$, el cuadrado de la norma de su transformado:

$$Q : \mathbb{R}^2 \to \mathbb{R},\ Q(x) = \|Ax\|^2 \tag{6.2}$$

donde el producto interno es el canónico. Veamos que la fórmula de Q se puede escribir de la manera usual (definida en 5.4) con una matriz simétrica:

$$Q(x) = \|Ax\|^2 = (Ax)^t (Ax) = x^t \left(A^t A\right) x$$

siendo en este caso

$$A^t A = \begin{bmatrix} 2 & 1 \\ 1 & 2 \end{bmatrix} \tag{6.3}$$

En la figura 6.1 se ilustra cómo T transforma el espacio $\mathbb{R}^2$ en un plano de $\mathbb{R}^3$ (recordemos que para una transformación definida como $T(x) = Ax$ es $\operatorname{Im} T = \operatorname{Col} A$). En particular, los vectores de $\mathbb{R}^2$ que verifican la restricción $\|x\| = 1$ se transforman en una elipse. [1]

[1] Aunque es intuitivamente aceptable, esta afirmación no es obvia. Más adelante en este mismo capítulo tendremos elementos para demostrarla.

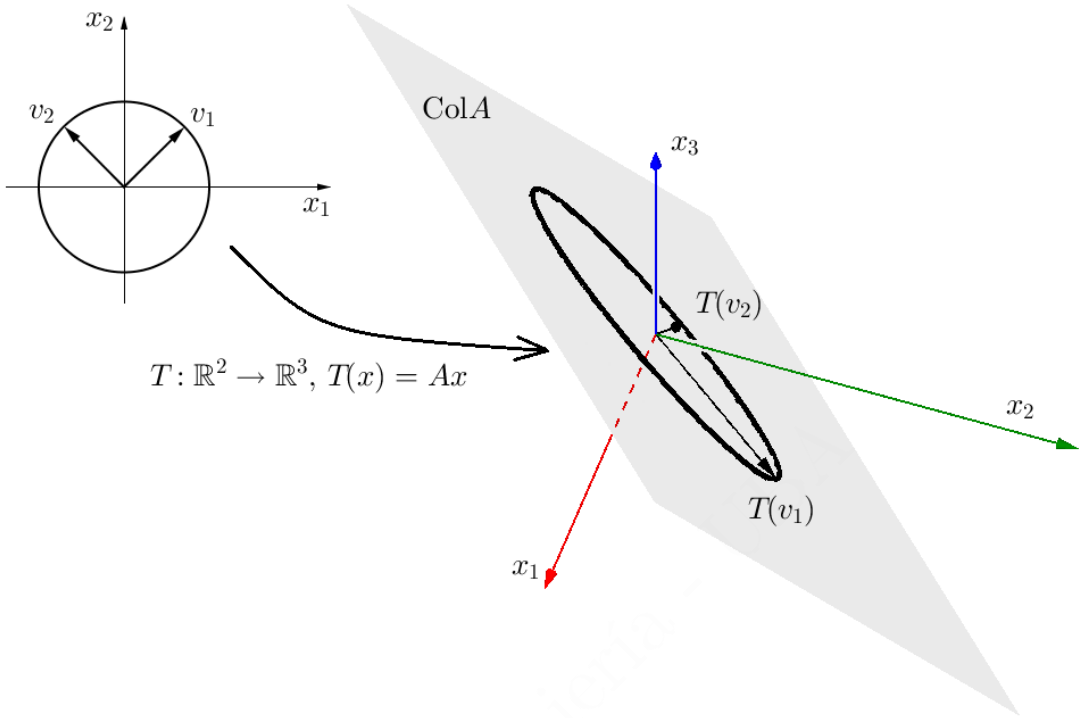


Figura 6.1: Transformación de $\mathbb{R}^2$ y de la circunferencia

Buscamos ahora los vectores de $\mathbb{R}^2$ sujetos a la restricción $\|x\| = 1$ en los que Q alcanza sus extremos. De acuerdo con el teorema de Rayleigh 5.7, esto ocurrirá en los autovectores de norma uno de la matriz A^tA que define a la forma cuadrática. En el caso de la matriz en (6.3) los autovalores son 3 y 1 y sus correspondientes autoespacios son

$$S_3 = \text{gen}\left\{\frac{1}{\sqrt{2}}[1\ 1]^t\right\} \qquad S_1 = \text{gen}\left\{\frac{1}{\sqrt{2}}[-1\ 1]^t\right\}$$

Por lo tanto, el máximo de $\|Ax\|^2$ sujeta a la restricción $\|x\| = 1$ es 3 y se alcanza en los vectores $\pm v_1 = \pm\frac{1}{\sqrt{2}}[1\ 1]^t$. El mínimo es 1 y se alcanza en los vectores $\pm v_2 = \pm\frac{1}{\sqrt{2}}[-1\ 1]^t$. De acuerdo con nuestra interpretación geométrica, los semiejes de la elipse en la figura 6.1 miden $\sqrt{3}$ y 1: estos valores surgen de $\|Av_1\|$ y $\|Av_2\|$. Por cierto, Av_1 y Av_2 generan los ejes de simetría de la elipse.

Las definiciones que siguen apuntan a generalizar, para cualquier matriz $A \in \mathbb{R}^{m\times n}$, las observaciones e interpretaciones hechas en este ejemplo.

Señalamos, en primer lugar, algunas propiedades que serán de gran utilidad a lo largo del capítulo.

Proposición 6.1

Dada una matriz $A \in \mathbb{R}^{m\times n}$, la matriz $A^tA \in \mathbb{R}^{n\times n}$ resulta simétrica y semidefinida positiva.

Demostración:
Verificamos que A^tA es simétrica trasponiéndola:

$$\left(A^tA\right)^t = A^t\left(A^t\right)^t = A^tA$$

Por lo tanto, de acuerdo con el teorema espectral 5.4, sus autovalores son números reales. Además, para cualquier vector $x \in \mathbb{R}^n$ resulta

$$x^t A^t A x = (Ax)^t (Ax) = \|Ax\|^2 \geq 0$$

con lo cual es semidefinida positiva y sus autovalores son, como ya vimos en el teorema 5.6, números mayores o iguales que cero. □

Ejercicio.

6.1 Probar que para toda $A \in \mathbb{R}^{m\times n}$, la matriz $A^t A \in \mathbb{R}^{n\times n}$ resulta definida positiva si y sólo si Nul $A = \{0\}$.

Estas propiedades garantizan que tiene sentido la siguiente definición.

Definición 6.1

Sea $A \in \mathbb{R}^{m\times n}$. Sean $\lambda_1, \lambda_2, ..., \lambda_n$ los autovalores de $A^t A \in \mathbb{R}^{n\times n}$ ordenados en forma decreciente, es decir

$$\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n \geq 0$$

Entonces $\sigma_i = \sqrt{\lambda_i}$ con $i = 1, 2, ..., n$ es el i-ésimo *valor singular* de A. O sea, los valores singulares de una matriz A son las raíces cuadradas de los autovalores de $A^t A$.

En particular denotaremos como $\sigma_{\text{máx}}$ al mayor valor singular y $\sigma_{\text{mín}}$ al menor:

$$\sigma_{\text{máx}} = \sqrt{\lambda_1} \qquad \sigma_{\text{mín}} = \sqrt{\lambda_n}$$

Ya habíamos observado que los autovalores de $A^t A$ dan cuenta del máximo y mínimo alcanzados por $\|Ax\|^2$ sujeta a la restricción $\|x\| = 1$. A continuación, generalizamos los argumentos usados y los explicamos en términos de valores singulares.

Teorema 6.1

Para toda $A \in \mathbb{R}^{m\times n}$ resulta

$$\max_{\|x\|=1} \|Ax\| = \sigma_{\text{máx}} \qquad \text{y} \qquad \min_{\|x\|=1} \|Ax\| = \sigma_{\text{mín}}$$

Demostración:
Sea $\lambda_{\text{máx}} = \lambda_1$ el máximo autovalor de $A^t A$. Entonces, usando el teorema de Rayleigh 5.7

$$\max_{\|x\|=1} \|Ax\|^2 = \max_{\|x\|=1} x^t \left(A^t A\right) x = \lambda_{\text{máx}}$$

y entonces

$$\sigma_{\text{máx}} = \sqrt{\lambda_{\text{máx}}} = \sqrt{\max_{\|x\|=1} x^t \left(A^t A\right) x} = \sqrt{\max_{\|x\|=1} \|Ax\|^2} = \max_{\|x\|=1} \sqrt{\|Ax\|^2} = \max_{\|x\|=1} \|Ax\|$$

de donde sigue el resultado.[2] De manera análoga se demuestra el caso del mínimo. □

Como corolario del teorema obtenemos una desigualdad que expresa, en términos de los valores singulares, cómo afecta a la norma la multiplicación por una matriz A.

[2] Para intercambiar la raíz con el máximo hemos tenido en cuenta que la función raíz cuadrada es creciente, por lo tanto alcanzará su máximo cuando el argumento sea máximo.

Proposición 6.2

Sea $A \in \mathbb{R}^{m\times n}$. Entonces se tiene para todo $x \in \mathbb{R}^n$

$$\sigma_{\text{mín}} \|x\| \leq \|Ax\| \leq \sigma_{\text{máx}} \|x\|$$

Demostración:
Si $x = 0$ las desigualdades son inmediatamente ciertas. En el caso de $x \neq 0$ podemos definir el vector unitario $v = \frac{x}{\|x\|}$ y aplicarle el teorema anterior, con lo cual $\sigma_{\text{mín}} \leq \|Av\| \leq \sigma_{\text{máx}}$. Como $\|Av\| = \frac{\|Ax\|}{\|x\|}$ obtenemos

$$\sigma_{\text{mín}} \leq \frac{\|Ax\|}{\|x\|} \leq \sigma_{\text{máx}}$$

de donde sigue el resultado. □

Al ser A^tA una matriz simétrica, siempre podemos hallar una base ortonormal de autovectores. El teorema que sigue establece importantes propiedades acerca de estas bases.

Teorema 6.2

Sea $A \in \mathbb{R}^{m\times n}$. Supongamos que $\lambda_1, ..., \lambda_n$ son los autovalores de A^tA, están ordenados en forma decreciente y la cantidad de autovalores no nulos es r:

$$\lambda_1 \geq ... \geq \lambda_r > 0 \qquad \text{y} \qquad \lambda_{r+1} = ... = \lambda_n = 0$$

Sea $\{v_1, v_2, ..., v_n\}$ una base ortonormal de $\mathbb{R}^n$ formada por autovectores de A^tA asociados a $\lambda_1, \lambda_2, ..., \lambda_n$ respectivamente. Entonces se cumple que

1. $\{Av_1, Av_2, ..., Av_n\}$ es un conjunto ortogonal y para todo $i = 1, ..., n$ es $\|Av_i\| = \sqrt{\lambda_i} = \sigma_i$.
2. $\left\{\frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}, ..., \frac{Av_r}{\sigma_r}\right\}$ es una base ortonormal de Col A.
3. rango $A = r$ es la cantidad de valores singulares no nulos de A.
4. $\{v_{r+1}, v_{r+2}, ..., v_n\}$ es una base ortonormal de Nul A.

Demostración:
Para demostrar 1 tenemos en cuenta que $\{v_1, v_2, ..., v_n\}$ es una base ortonormal:

$$\left(Av_i, Av_j\right) = (Av_i)^t \left(Av_j\right) = v_i^t \left(A^tAv_j\right) = v_i^t \left(\lambda_j v_j\right) = \lambda_j \left(v_i^t v_j\right) = \lambda_j \left(v_i, v_j\right)$$

Entonces si $i \neq j$ resulta $\left(Av_i, Av_j\right) = 0$. Para el caso $i = j$ tenemos $(Av_i, Av_i) = \|Av_i\|^2 = \lambda_i$, de donde sigue la tesis.
En cuanto al punto 2, ya sabemos que el conjunto $\left\{\frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}, ..., \frac{Av_r}{\sigma_r}\right\}$ es ortonormal y por lo tanto linealmente independiente: falta probar que genera Col A. Para ello, consideramos la transformación lineal $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(x) = Ax$. Entonces, teniendo en cuenta que Im T está generada por los transformados de los vectores de cualquier base de $\mathbb{R}^n$:

$$\text{Col } A = \text{Im } T = \text{gen}\{Av_1, Av_2, ..., Av_n\} = \text{gen}\{Av_1, Av_2, ..., Av_r\}$$

donde la última igualdad se debe a que, como ya demostramos en 1, $Av_i = 0$ si $i > r$. De esto se deduce inmediatamente 3. Finalmente, para probar la propiedad 4 comenzamos por señalar que dim (Nul A) $= n - r$. Dado que $\{v_{r+1}, v_{r+2}, ..., v_n\}$ es un

conjunto ortonormal de $n-r$ vectores que pertenecen a Nul A (¿por qué?) tenemos la base buscada. □

En la figura 6.2 se ilustran las relaciones establecidas por el teorema 6.2. Los vectores $\{v_{r+1}, v_{r+2}, ..., v_n\}$ constituyen una base ortonormal de Nul A. Por lo tanto $\{v_1, v_2, ..., v_r\}$ es una base del complemento ortogonal de Nul A, es decir, Fil A. Si consideramos la transformación lineal $T : \mathbb{R}^n \to \mathbb{R}^m$, $T(x) = Ax$, el espacio de partida $\mathbb{R}^n$ puede descomponerse como la suma directa de Nul A y Nul $A^\perp$. Evidentemente, los vectores de Nul A se transforman en cero. Por su parte, la base $\{v_1, v_2, ..., v_r\}$ de Nul $A^\perp$ se transforma en una base ortogonal de la imagen Col A, que puede normalizarse como $\left\{\frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}, ..., \frac{Av_r}{\sigma_r}\right\}$.

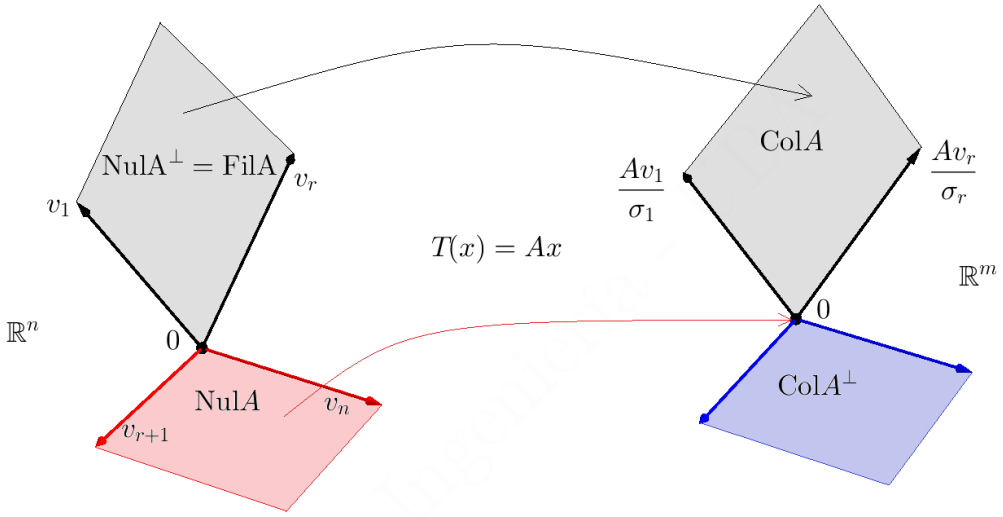


Figura 6.2: Los espacios fundamentales de A

Ejemplo 6.1

En el caso de la matriz

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

estudiada al comienzo del capítulo, la base de autovectores de A^tA descrita en el teorema anterior es

$$\{v_1, v_2\} = \left\{\frac{1}{\sqrt{2}}[1\ 1]^t, \frac{1}{\sqrt{2}}[-1\ 1]^t\right\} \tag{6.4}$$

y al multiplicarlos por A obtenemos un conjunto ortogonal en $\mathbb{R}^3$

$$\{Av_1, Av_2\} = \left\{\frac{1}{\sqrt{2}}[1\ 2\ -1]^t, \frac{1}{\sqrt{2}}[-1\ 0\ -1]^t\right\}$$

que al normalizarlo resulta una base ortonormal de Col A

$$B = \left\{\frac{Av_1}{\|Av_1\|}, \frac{Av_2}{\|Av_2\|}\right\} = \left\{\frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}\right\} = \left\{\frac{1}{\sqrt{6}}[1\ 2\ -1]^t, \frac{1}{\sqrt{2}}[-1\ 0\ -1]^t\right\} \tag{6.5}$$

Ejercicio.

6.2 Dada la matriz

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix}$$

- Calcule los valores singulares de A.
- Encuentre una base ortonormal $B = \{v_1, v_2, v_3\}$ de $\mathbb{R}^3$ de modo tal que $\{v_3\}$ sea una base ortonormal de Nul A y $\left\{\frac{Av_1}{\|Av_1\|}, \frac{Av_2}{\|Av_2\|}\right\}$ de Col A.

En el próximo ejemplo demostramos formalmente la afirmación que habíamos hecho al comienzo del capítulo acerca de la circunferencia unitaria y su transformación en elipse. Aunque su lectura no es indispensable para lo que sigue, pensamos que puede resultar de interés.

Ejemplo 6.2

Consideremos una transformación lineal $T : \mathbb{R}^2 \to \mathbb{R}^3, T(x) = Ax$ con rango $A = 2$. Probaremos que, como ya se mencionó al comienzo del capítulo e ilustró en la figura 6.1, la imagen de la circunferencia unitaria es una elipse (eventualmente una circunferencia).

Para demostrarlo, usaremos el teorema 6.2. Consideramos entonces $\{v_1, v_2\}$ una base ortonormal de $\mathbb{R}^2$ formada por autovectores de A^tA. Los vectores de $\mathbb{R}^2$ son de la forma $\alpha v_1 + \beta v_2$, y los que pertenecen a la circunferencia quedan caracterizados por tener norma uno. Usando el hecho de que $\{v_1, v_2\}$ es una base ortonormal de $\mathbb{R}^2$ resulta

$$\|\alpha v_1 + \beta v_2\|^2 = 1 \Leftrightarrow \alpha^2 + \beta^2 = 1$$

por lo tanto los elementos de la circunferencia son de la forma

$$x = \alpha v_1 + \beta v_2 \qquad \text{con} \qquad \alpha^2 + \beta^2 = 1$$

y se transforman en vectores de $\mathbb{R}^3$ de la forma

$$Ax = \alpha Av_1 + \beta Av_2 \qquad \text{con} \qquad \alpha^2 + \beta^2 = 1$$

A continuación, expresaremos Ax a partir de la base ortonormal de Col A descrita en el ítem 2 del teorema anterior:

$$Ax = \sigma_1 \alpha \cdot \left(\frac{Av_1}{\sigma_1}\right) + \sigma_2 \beta \cdot \left(\frac{Av_2}{\sigma_2}\right) \qquad \text{con} \qquad \alpha^2 + \beta^2 = 1$$

De esta forma, las coordenadas de Ax con respecto a la base ortonormal $B = \left\{\frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}\right\}$ son

$$[Ax]_B = \begin{bmatrix} \sigma_1 \alpha & \sigma_2 \beta \end{bmatrix}^t = \begin{bmatrix} y_1 & y_2 \end{bmatrix}^t$$

y verifican la ecuación de una elipse cuyos semiejes miden σ_1 y σ_2:

$$\left(\frac{y_1}{\sigma_1}\right)^2 + \left(\frac{y_2}{\sigma_2}\right)^2 = \alpha^2 + \beta^2 = 1$$

Ejercicio.

6.3 Considere la transformación lineal $T:\mathbb{R}^2 \to \mathbb{R}^3, T(x) = Ax$ definida al comienzo del capítulo mediante la matriz (6.1). Pruebe que la circunferencia unitaria se transforma, eligiendo un sistema de coordenadas adecuado en Col A, en la elipse de ecuación $\left(\frac{y_1}{\sqrt{3}}\right)^2 + (y_2)^2 = 1$.

Ejercicio.

6.4 Considere la transformación lineal $T:\mathbb{R}^2 \to \mathbb{R}^3, T(x) = Ax$ definida mediante la matriz

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

y describa el conjunto de $\mathbb{R}^3$ en que transforma a la circunferencia unitaria.

Ejemplo 6.3

Consideramos ahora una transformación lineal $T:\mathbb{R}^3 \to \mathbb{R}^2, T(x) = Ax$ con rango $A = 2$ definida por la matriz

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

Probaremos que, como se ilustra en la figura 6.3, la imagen de la esfera unitaria es una región de borde elíptico. Los recursos que emplearemos son totalmente análogos a los del ejemplo 6.2.

Usaremos una vez más el teorema 6.2, para lo cual calculamos

$$A^tA = \begin{bmatrix} 5 & 10 & 0 \\ 10 & 20 & 0 \\ 0 & 0 & 5 \end{bmatrix}$$

Sea entonces $\{v_1, v_2, v_3\}$ una base ortonormal de $\mathbb{R}^3$ formada por autovectores de A^tA. Dado que los autoespacios son

$$S_{25} = \text{gen}\{[1\ 2\ 0]^t\} \quad S_5 = \text{gen}\{[0\ 0\ 1]^t\} \quad \text{gen}\, S_0 = \text{gen}\{[-2\ 1\ 0]^t\}$$

definimos

$$v_1 = \frac{1}{\sqrt{5}}\begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix} \quad v_2 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \quad v_3 = \frac{1}{\sqrt{5}}\begin{bmatrix} -2 \\ 1 \\ 0 \end{bmatrix}$$

Luego los vectores de $\mathbb{R}^3$ que pertenecen a la esfera unitaria son de la forma

$$x = \alpha v_1 + \beta v_2 + \gamma v_3 \qquad \text{con} \qquad \alpha^2 + \beta^2 + \gamma^2 = 1$$

y se transforman en vectores de $\mathbb{R}^2$ de la forma

$$Ax = \alpha Av_1 + \beta Av_2 + \gamma Av_3 \qquad \text{con} \qquad \alpha^2 + \beta^2 + \gamma^2 = 1$$

Teniendo en cuenta que $Av_3 = 0$, resulta que los vectores de la esfera unitaria se transforman en vectores de la forma

$$Ax = \alpha Av_1 + \beta Av_2 \qquad \text{con} \qquad \alpha^2 + \beta^2 \leq 1$$

Falta expresar Ax en términos de una base ortonormal de Col A. Para ello apelamos al ítem ❷ del teorema 6.2:

$$Ax = \sigma_1\alpha \cdot \left(\frac{Av_1}{\sigma_1}\right) + \sigma_2\beta \cdot \left(\frac{Av_2}{\sigma_2}\right) \qquad \text{con} \qquad \alpha^2 + \beta^2 \leq 1$$

En este caso es $\sigma_1 = \sqrt{25} = 5$ y $\sigma_2 = \sqrt{5}$, luego

$$\frac{Av_1}{\sigma_1} = \frac{1}{5}\begin{bmatrix} \sqrt{5} \\ 2\sqrt{5} \end{bmatrix} \qquad \frac{Av_2}{\sigma_2} = \frac{1}{\sqrt{5}}\begin{bmatrix} 2 \\ -1 \end{bmatrix}$$

Con lo cual las coordenadas de Ax con respecto a la base ortonormal $B = \left\{\frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}\right\}$ son

$$[Ax]_B = \begin{bmatrix} 5\alpha & \sqrt{5}\beta \end{bmatrix}^t = [y_1\ y_2]^t$$

y determinan una *región de borde elíptico* cuyos semiejes miden 5 y $\sqrt{5}$:

$$\left(\frac{y_1}{5}\right)^2 + \left(\frac{y_2}{\sqrt{5}}\right)^2 = \alpha^2 + \beta^2 \leq 1$$

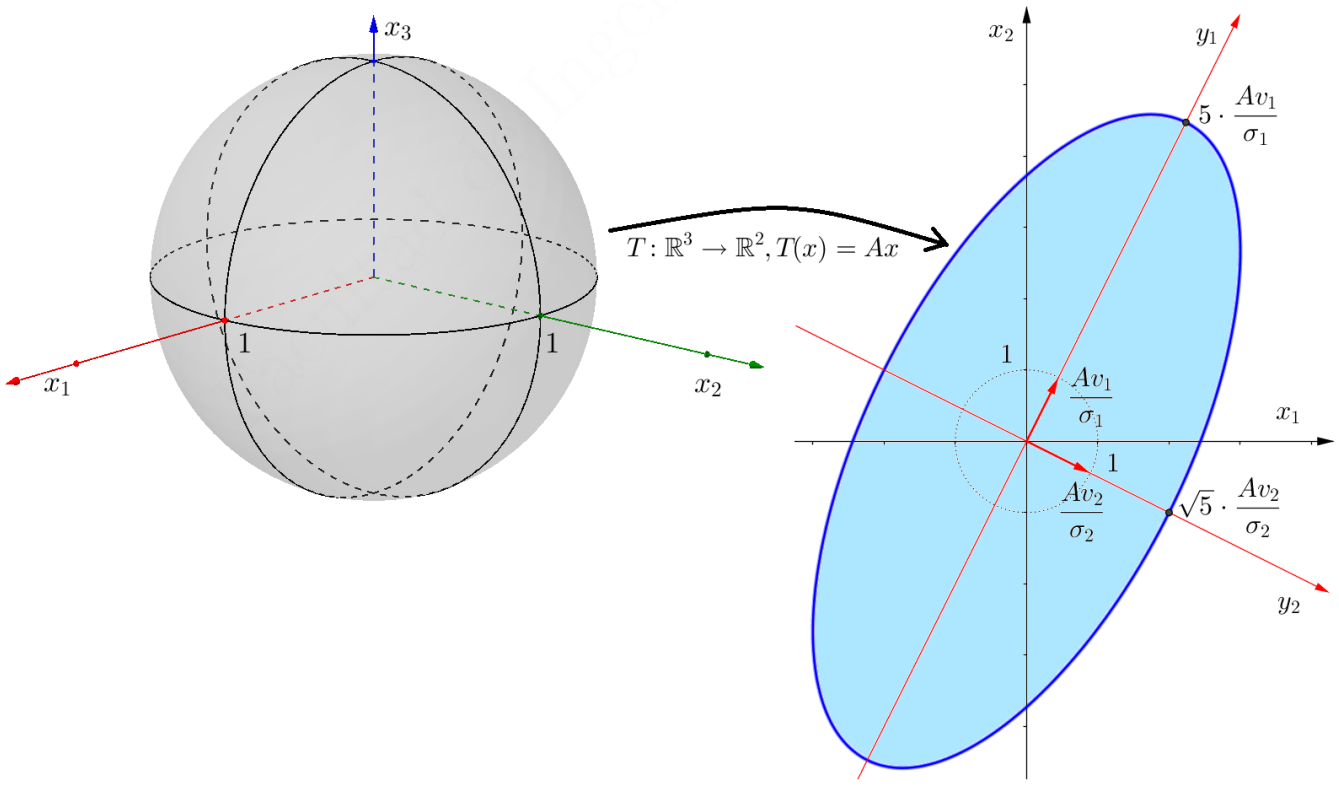


Figura 6.3: Imagen de la esfera unitaria

Ejercicio.

6.5 Considere una transformación lineal $T:\mathbb{R}^3 \to \mathbb{R}^3$, $T(x) = Ax$ y estudie la imagen de la esfera unitaria para los dos casos siguientes

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

6.2 Descomposición en valores singulares

En el capítulo 4 presentamos la diagonalización de matrices cuadradas buscando una factorización de la forma $A = PDP^{-1}$, siendo D una matriz diagonal. Una de las ideas que motivan esta factorización es que al elegir bases de autovectores podemos comprender mejor la transformación lineal asociada con la matriz A. Evidentemente, tal factorización sólo puede lograrse (y no en cualquier caso) para *matrices cuadradas*, puesto que la ecuación $Av = \lambda v$ no tiene sentido si $A \in \mathbb{R}^{m\times n}$ con $m \neq n$. En esta sección presentaremos una factorización que puede obtenerse para cualquier matriz rectangular y brinda información acerca de los subespacios fundamentales.

Cada matriz $A \in \mathbb{R}^{m\times n}$ determina una transformación lineal $T:\mathbb{R}^n \to \mathbb{R}^m$, $T(x) = Ax$. Como ya mencionamos en la sección previa y de acuerdo con el teorema 6.2, podemos elegir una base ortonormal del espacio $\mathbb{R}^n$ tal que sus transformados permiten obtener una base ortonormal de Col A, conociendo además cuáles son los vectores de norma uno que realizan la máxima y mínima dilatación al transformarse. La factorización que sigue pone en evidencia toda esta información.

Definición 6.2

Sea $A \in \mathbb{R}^{m\times n}$. Una *descomposición en valores singulares (DVS)* de A es una factorización de la forma

$$A = U\Sigma V^t \tag{6.6}$$

con $U \in \mathbb{R}^{m\times m}$ y $V \in \mathbb{R}^{n\times n}$ matrices ortogonales y $\Sigma \in \mathbb{R}^{m\times n}$ de la forma

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \quad \text{donde} \quad D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix} \tag{6.7}$$

y $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_r > 0$. Vale decir que Σ es una matriz con un primer bloque diagonal y el resto de los bloques son matrices de ceros con el tamaño adecuado.

N La descomposición en valores singulares es de gran utilidad y tiene muchas aplicaciones en la ingeniería. Aunque no las expondremos en este curso, los lectores interesados pueden consultar la bibliografía, por ejemplo [5].

Definición 6.3

Dada una matriz $A \in \mathbb{R}^{m\times n}$ con descomposición en valores singulares $A = U\Sigma V^t$ como la definida en 6.2, a los vectores que aparecen como columnas de la matriz V se los denomina *vectores singulares derechos* de A mientras que a los que aparecen como columnas de U se los denomina *vectores singulares izquierdos* de A.

Hasta ahora hemos definido la descomposición en valores singulares, pero no hemos garantizado su existencia ni exhibido un método para construirla. El teorema que sigue resuelve ambas cuestiones: garantiza la existencia de la descomposición para cualquier matriz y en la demostración se explica cómo construirla.

Teorema 6.3

Sea $A \in \mathbb{R}^{m\times n}$. Entonces existe una descomposición en valores singulares de A.

Demostración:
Siguiendo la exposición del teorema 6.2, llamamos $\lambda_1, ..., \lambda_n$ a los autovalores de A^tA, de manera que estén ordenados en forma decreciente y la cantidad de elementos no nulos sea r:

$$\lambda_1 \geq ... \geq \lambda_r > 0 \qquad \text{y} \qquad \lambda_{r+1} = ... = \lambda_n = 0$$

Sea $\{v_1, v_2, ..., v_n\}$ una base ortonormal de $\mathbb{R}^n$ formada por autovectores de A^tA de modo tal que para todo $i = 1, ..., n$ resulta $A^tAv_i = \lambda_i v_i$. La matriz V se define con estos vectores como columnas:

$$V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

La matriz $\Sigma \in \mathbb{R}^{m\times n}$ se define como

$$\Sigma = \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \qquad \text{donde} \qquad D = \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r \end{bmatrix}$$

y $\sigma_i = \sqrt{\lambda_i}$ con $i = 1, ..., r$. Para definir U, tenemos en cuenta que el conjunto $\left\{ \frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}, ..., \frac{Av_r}{\sigma_r} \right\}$ es ortonormal, luego definimos

$$u_1 = \frac{Av_1}{\sigma_1} \qquad u_2 = \frac{Av_2}{\sigma_2} \qquad ... \qquad u_r = \frac{Av_r}{\sigma_r}$$

Si $r < m$, es decir si los u_i no completan una base de $\mathbb{R}^m$, buscamos vectores $u_{r+1}, ..., u_m$ de modo tal que $\{u_1, u_2, ..., u_m\}$ sea una base ortonormal de $\mathbb{R}^m$. Con esto definimos la matriz ortogonal

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$$

A continuación, comprobamos que efectivamente es $A = U\Sigma V^t$ para las matrices definidas. En primer lugar, teniendo en cuenta la definición de los u_i y que $Av_i = 0$ si $i > r$:

$$AV = \begin{bmatrix} Av_1 & Av_2 & \cdots & Av_n \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_r v_r & 0 & \cdots & 0 \end{bmatrix}$$

Por otra parte,

$$U\Sigma = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} \begin{bmatrix} \sigma_1 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix} = \begin{bmatrix} \sigma_1 u_1 & \sigma_2 u_2 & \cdots & \sigma_r u_r & 0 & \cdots & 0 \end{bmatrix}$$

Por lo tanto $AV = U\Sigma$. Teniendo en cuenta que $V^{-1} = V^t$, multiplicamos a derecha por V^t y obtenemos $AVV^t = A = U\Sigma V^t$ □

Ejemplo 6.4

Una vez más, consideramos la matriz (6.1) definida al comienzo del capítulo y estudiada en el ejemplo 6.1. Siguiendo los pasos de la demostración del teorema 6.3, señalamos en primer lugar los autovalores de A^tA:

$$\lambda_1 = 3 \qquad \lambda_2 = 1$$

de donde obtenemos los valores singulares

$$\sigma_1 = \sqrt{3} \qquad \sigma_2 = 1$$

que nos permiten definir

$$\Sigma = \begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}$$

La base de $\mathbb{R}^2$ formada por autovectores de A^tA es la que ya se presentó en (6.4):

$$\{v_1, v_2\} = \left\{ \frac{1}{\sqrt{2}} [1\ 1]^t, \frac{1}{\sqrt{2}} [-1\ 1]^t \right\}$$

y ya podemos definir

$$V = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

A continuación definimos la base ortonormal de Col A como en (6.5):

$$\{u_1, u_2\} = \left\{ \frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2} \right\} = \left\{ \frac{1}{\sqrt{6}} [1\ 2\ -1]^t, \frac{1}{\sqrt{2}} [-1\ 0\ -1]^t \right\}$$

Falta un vector u_3 que complete una base ortonormal de R^3: de las dos posibilidades que existen elegimos $u_3 = \frac{1}{\sqrt{3}} [1\ -1\ -1]^t$ y definimos

$$U = \begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{bmatrix}$$

Con lo cual la descomposición en valores singulares resulta

$$A = \begin{bmatrix} 1 & 0 \\ 1 & 1 \\ 0 & -1 \end{bmatrix} = \underbrace{\begin{bmatrix} \frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & \frac{1}{\sqrt{3}} \\ \frac{2}{\sqrt{6}} & 0 & -\frac{1}{\sqrt{3}} \\ -\frac{1}{\sqrt{6}} & -\frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{3}} \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} \sqrt{3} & 0 \\ 0 & 1 \\ 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ -1 & 1 \end{bmatrix}}_{V^t}$$

N Debe notarse que para una matriz dada $A \in \mathbb{R}^{m\times n}$ hay más de una descomposición en valores singulares. Al releer la demostración del teorema 6.3 se aprecia que puede haber diversas maneras de completar la base $\{u_1, u_2, ..., u_m\}$. También es posible elegir otros autovectores v_i de A^tA. El ejemplo más obvio son las descomposiciones $A = U\Sigma V^t = (-U)\Sigma(-V)^t$.

Ejercicio.

6.6 Encuentre una descomposición en valores singulares de la matriz estudiada en el ejercicio 6.2.

Ya hemos mencionado que para una misma matriz pueden encontrarse distintas descomposiciones en valores singulares. De todos modos, existen algunas características que toda DVS necesariamente tiene, independientemente de la forma en que haya sido obtenida. Aunque no serán indispensables para los temas que siguen, ofrecemos al lector interesado ciertas cualidades que debe tener cualquier DVS.

Teorema 6.4

Sea $A = U\Sigma V^t$ una descomposición en valores singulares de $A \in \mathbb{R}^{m\times n}$, con U, Σ y V como en la definición 6.2. Entonces

1. $\sigma_1, \sigma_2, ..., \sigma_r$ son los valores singulares no nulos de A.
2. Si $V \in \mathbb{R}^{n\times n}$ es una matriz de la forma
$$V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$
entonces los vectores columna constituyen una base ortonormal $\{v_1, v_2, ..., v_n\}$ de $\mathbb{R}^n$ y verifican que $A^tAv_i = \sigma_i^2 v_i$ para todo $i = 1, ..., r$ y $A^tAv_i = 0$ para $i = r+1, ..., n$.
3. Si $U \in \mathbb{R}^{m\times m}$ es una matriz de la forma
$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix}$$
entonces $Av_i = \sigma_i u_i$ para todo $i = 1, ..., r$.

Demostración:
Como $A = U\Sigma V^t$, tenemos que $A^tA = V\Sigma^t U^t U\Sigma V^t = V\left(\Sigma^t\Sigma\right)V^t$ donde

$$\Sigma^t\Sigma = \begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_r^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & 0 \end{bmatrix}$$

de modo tal que $V\left(\Sigma^t\Sigma\right)V^t$ es una diagonalización ortogonal de la matriz simétrica A^tA. En consecuencia, $\sigma_1^2, \sigma_2^2, ..., \sigma_r^2$, son los autovalores no nulos de A^tA ordenados en forma decreciente. [3] Por lo tanto sus raíces cuadradas $\sigma_1, \sigma_2, ..., \sigma_r$ son los

[3] Estamos usando el teorema 4.3 que garantiza que en cualquier diagonalización $A = PDP^{-1}$ de una matriz los elementos de la diagonal son necesariamente autovalores de A y las columnas de P sus autovectores.

valores singulares no nulos de A y queda probada ❶.
Las columnas de V, a su vez, conforman una base ortogonal de $\mathbb{R}^n$ y son autovectores de A^tA. Más precisamente: $A^tAv_i = \sigma_i^2 v_i$ para $i = 1,...,r$ y $A^tAv_i = 0$ para $i = r+1,...,n$, ya que $v_{r+1},...,v_n$ son autovectores asociados al autovalor nulo. Con esto hemos demostrado ❷.
Finalmente el punto ❸ resulta inmediato a partir de la igualdad $AV = U\Sigma$. □

Ejercicio. Sea $A = U\Sigma V^t$ una descomposición en valores singulares de $A \in \mathbb{R}^{m\times n}$.

- Demuestre que $A^t = V\Sigma^t U^t$ es una DVS de A^t.
- Demuestre que A y A^t tienen los mismos valores singulares no nulos.

6.3 La DVS y los subespacios fundamentales de una matriz

En esta sección veremos que en una descomposición en valores singulares $A = U\Sigma V^t$, las matrices U y V están compuestas por bases ortonormales de los cuatro subespacios fundamentales de A.

Teorema 6.5

Sea $A \in \mathbb{R}^{m\times n}$ una matriz de rango r con una DVS $A = U\Sigma V^t$ donde

$$U = \begin{bmatrix} u_1 & u_2 & \cdots & u_m \end{bmatrix} \quad \text{y} \quad V = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix}$$

Definiendo las matrices

$$U_r = \begin{bmatrix} u_1 & \cdots & u_r \end{bmatrix} \qquad U_{m-r} = \begin{bmatrix} u_{r+1} & \cdots & u_m \end{bmatrix} \tag{6.8}$$

junto con

$$V_r = \begin{bmatrix} v_1 & \cdots & v_r \end{bmatrix} \qquad V_{n-r} = \begin{bmatrix} v_{r+1} & \cdots & v_n \end{bmatrix} \tag{6.9}$$

resulta

❶ $\{v_1,...,v_r\}$ es una base ortonormal de $(\text{Nul}\,A)^\perp = \text{Fil}\,A$.

❷ $V_r^tV_r = I$ y $V_rV_r^t = [P_{\text{Fil}\,A}]_E$

❸ $\{v_{r+1},...,v_n\}$ es una base ortonormal de $\text{Nul}\,A$.

❹ $\{u_1,...,u_r\}$ es una base ortonormal de $\text{Col}\,A$.

❺ $U_r^tU_r = I$ y $U_rU_r^t = [P_{\text{Col}\,A}]_E$

❻ $\{u_{r+1},...,u_m\}$ es una base ortonormal de $(\text{Col}\,A)^\perp = \text{Nul}\,A^t$.

Demostración:
Como el rango de A es r, de acuerdo con la propiedad ❸ del teorema 6.2, A es una matriz con r valores singulares no nulos. Por lo tanto A^tA tiene r autovalores no nulos. Por el teorema 6.4, los vectores $\{v_1, v_2,..., v_n\}$ son autovectores de A^tA, estando los primeros r de ellos asociados a los autovalores no nulos de A^tA y los restantes al autovalor nulo de A^tA. Por lo tanto, por la propiedad ❹ del teorema 6.2, $\{v_{r+1},...,v_n\}$ es una base ortonormal de $\text{Nul}\,A$. Como además $\{v_1,...,v_n\}$ es base ortonormal

de $\mathbb{R}^n$, necesariamente $\{v_1,\dots,v_r\}$ es una base ortonormal de $(\text{Nul}\,A)^\perp = \text{Fil}\,A$. Con esto, y la forma en que se construyen las matrices asociadas a una proyección ortogonal (explicada en el teorema 3.19), quedan demostrados ❶, ❷ y ❸.
La demostración de ❹, ❺ y ❻ es análoga y se deja como ejercicio. □

Ejercicio. Sea $A \in \mathbb{R}^{m\times n}$ una matriz de rango r. Manteniendo la notación del teorema 6.5, demuestre que

- $V_{n-r}^t V_{n-r} = I$ y $V_{n-r}V_{n-r}^t = [P_{\text{Nul}\,A}]_E$
- $U_{m-r}^t U_{m-r} = I$ y $U_{m-r}U_{m-r}^t = \left[P_{\text{Nul}\,A^t}\right]_E$

Ejemplo 6.5

Dada

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 1 & -1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2} & 0 & 0 \\ 0 & \sqrt{18} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1 & 3\sqrt{2} & 1 \\ 1 & -3\sqrt{2} & 1 \\ 0 & 0 & 0 \end{bmatrix}$$

Buscamos los valores singulares de A, bases de sus cuatro subespacios fundamentales y calculamos las matrices asociadas a la proyección sobre dichos subespacios.

La descomposición que tenemos de A es similar a una DVS, salvo por el hecho de que la matriz que aparece a la izquierda, aunque tiene columnas mutuamente ortogonales, no son de norma uno. Procedemos entonces a normalizar sus columnas, dividiendo cada una de ellas por su norma. Para que el producto siga siendo A, multiplicamos adecuadamente las filas de la matriz central. Queda entonces la factorización

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \sqrt{2}\sqrt{2} & 0 & 0 \\ 0 & \sqrt{2}\sqrt{18} & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 0 & 1 & 0 \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}$$

Esta factorización no es aún una DVS, porque los elementos de la diagonal de la matriz central no están ordenados de mayor a menor. Para ello lo que hacemos es permutar las dos primeras columnas de la matriz de la izquierda y las dos primeras filas de la matriz de la derecha. Entonces obtenemos

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \tag{6.10}$$

que ahora sí es una DVS de A, ya que $A = U\Sigma V^t$ con

$$U = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad \Sigma = \begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \quad V = \begin{bmatrix} 0 & 1/\sqrt{2} & -1/\sqrt{2} \\ 1 & 0 & 0 \\ 0 & 1/\sqrt{2} & 1/\sqrt{2} \end{bmatrix}$$

Los valores singulares de A son $\sigma_1 = 6$, $\sigma_2 = 2$ y $\sigma_3 = 0$. Como A tiene dos valores singulares distintos de cero, deducimos que su rango es dos. Las primeras dos columnas de V son una base ortonormal de $\text{Fil}\,A$, mientras que la última columna de V es una base ortonormal de $\text{Nul}\,A$. Con respecto a $\text{Col}\,A$, las dos primeras columnas de U son una base ortonormal de ese subespacio, mientras que la última columna de U es base ortonormal de $\text{Nul}\,A^t$. Las matrices asociadas a las proyecciones son

$$[P_{\text{Fil}\,A}]_E = \begin{bmatrix} 0 & 1/\sqrt{2} \\ 1 & 0 \\ 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} = \begin{bmatrix} 1/2 & 0 & 1/2 \\ 0 & 1 & 0 \\ 1/2 & 0 & 1/2 \end{bmatrix}$$

$$[P_{\operatorname{Nul} A}]_E = I - [P_{\operatorname{Fil} A}]_E = \begin{bmatrix} 1/2 & 0 & -1/2 \\ 0 & 0 & 0 \\ -1/2 & 0 & 1/2 \end{bmatrix}$$

$$\left[P_{\operatorname{Nul} A^t}\right]_E = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

$$[P_{\operatorname{Col} A}]_E = I - \left[P_{\operatorname{Nul} A^t}\right]_E = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

Ejercicio.

6.7 Considere la matriz definida en el ejercicio 6.2, cuya DVS se obtuvo en el ejercicio 6.6.

- Encuentre bases ortonormales para los subespacios fundamentales.
- Calcule $P_{\operatorname{Col} A}(v)$ siendo $v = [1 \ \ -1]^t$.
- Calcule $P_{\operatorname{Fil} A}(w)$ siendo $w = [1 \ \ -1 \ \ 1]^t$.

Si releemos la ecuación (6.10) del ejemplo anterior, se aprecia que la última fila de V^t no aportará nada al producto, en el sentido de que todos sus elementos estarán multiplicados por los ceros de la última columna de Σ. Algo análogo ocurrirá con la última columna de U. Por lo tanto, en la DVS de A podemos prescindir de estos elementos y reducir el producto a la expresión

$$A = \begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} \\ -1/\sqrt{2} & 1/\sqrt{2} \\ 0 & 0 \end{bmatrix} \begin{bmatrix} 6 & 0 \\ 0 & 2 \end{bmatrix} \begin{bmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix} \tag{6.11}$$

Esta factorización de A permite leer sus valores singulares no nulos, una base ortonormal de $\operatorname{Col} A$ y una base ortonormal de $\operatorname{Fil} A$.

Veamos que para cualquier matriz $A \in \mathbb{R}^{m\times n}$ puede obtenerse una factorización como en (6.11). En efecto, si mantenemos la notación usada en el teorema 6.5 y la matriz Σ como se definió en (6.7), entonces una DVS de A puede escribirse como

$$A = U\Sigma V^t = \begin{bmatrix} U_r & U_{m-r} \end{bmatrix} \begin{bmatrix} D & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_r^t \\ V_{n-r}^t \end{bmatrix} = U_r D V_r^t$$

Esto justifica la siguiente definición.

Definición 6.4

Sea $A \in \mathbb{R}^{m\times n}$. Una *descomposición en valores singulares (DVS) reducida* de A es una factorización de la forma

$$A = U_r D V_r^t \tag{6.12}$$

obtenida a partir de una DVS como se definió en 6.2, siendo $U_r \in \mathbb{R}^{m\times r}$ y $V_r \in \mathbb{R}^{n\times r}$ las matrices definidas en (6.8) y (6.9) respectivamente.

Ejercicio.

6.8 Encuentre una DVS reducida de la matriz estudiada en el ejercicio 6.2.

6.4 Matriz seudoinversa. Cuadrados mínimos.

Consideramos, una vez más, la matriz estudiada en el ejemplo 6.5. Allí habíamos obtenido la DVS

$$A = \begin{bmatrix} 1 & 3\sqrt{2} & 1 \\ 1 & -3\sqrt{2} & 1 \\ 0 & 0 & 0 \end{bmatrix} = \underbrace{\begin{bmatrix} 1/\sqrt{2} & 1/\sqrt{2} & 0 \\ -1/\sqrt{2} & 1/\sqrt{2} & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{U} \underbrace{\begin{bmatrix} 6 & 0 & 0 \\ 0 & 2 & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\Sigma} \underbrace{\begin{bmatrix} 0 & 1 & 0 \\ 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ -1/\sqrt{2} & 0 & 1/\sqrt{2} \end{bmatrix}}_{V^t}$$

Observamos que

1. Las filas de V^t, esto es las columnas de V, constituyen una base ortonormal de $B = \{v_1, v_2, v_3\}$ de $\mathbb{R}^3$, donde

$$v_1 = \begin{bmatrix} 0 \\ 1 \\ 0 \end{bmatrix} \qquad v_2 = \begin{bmatrix} 1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix} \qquad v_3 = \begin{bmatrix} -1/\sqrt{2} \\ 0 \\ 1/\sqrt{2} \end{bmatrix}$$

   de modo tal que $\{v_1, v_2\}$ es una base ortonormal de Fil A y $\{v_3\}$ es una base ortonormal de Nul A. Además, la matriz V realiza el cambio de base: $V = C_{BE}$.

2. Las columnas de U constituyen una base ortonormal $C = \{u_1, u_2, u_3\}$ de $\mathbb{R}^3$ donde

$$u_1 = \begin{bmatrix} 1/\sqrt{2} \\ -1/\sqrt{2} \\ 0 \end{bmatrix} \qquad u_2 = \begin{bmatrix} 1/\sqrt{2} \\ 1/\sqrt{2} \\ 0 \end{bmatrix} \qquad u_3 = \begin{bmatrix} 0 \\ 0 \\ 1 \end{bmatrix}$$

   de modo tal que $\{u_1, u_2\}$ es una base ortonormal de Col A y $\{u_3\}$ es una base ortonormal de $(\text{Col } A)^{\perp}$. En este caso, $U = C_{CE}$.

De esta forma, A resulta la matriz asociada a la transformación

$$T : \mathbb{R}^3 \to \mathbb{R}^3,\ T(x) = Ax$$

con respecto a la base canónica, mientras que Σ es la matriz con respecto a las bases B y C:

$$[T]_E = A = \underbrace{C_{CE}}_{U} \cdot \underbrace{[T]_{BC}}_{\Sigma} \cdot \underbrace{C_{EB}}_{V^t}$$

Evidentemente, la transformación T no es inversible. Sin embargo, y como sugiere la figura 6.2, si restringimos el dominio a Fil A obtenemos una transformación biyectiva de Fil $A \to$ Col A. Nos proponemos entonces invertir T en esta restricción en que es posible. La transformación está caracterizada por cómo transforma a la base B:

$$Av_1 = 6u_1 \qquad Av_2 = 2u_2 \qquad Av_3 = 0$$

De esta manera, si pretendemos definir una transformación *seudoinversa* y representarla con una matriz $A^\dagger$, deberá verificar

$$A^\dagger(6u_1) = v_1 \Rightarrow A^\dagger u_1 = \frac{1}{6}v_1$$
$$A^\dagger(2u_2) = v_2 \Rightarrow A^\dagger u_2 = \frac{1}{2}v_2$$

Para que la pseudoinversa quede bien definida sobre la base $C = \{u_1, u_2, u_3\}$, imponemos

$$A^\dagger u_3 = 0$$

Todo este proceso se resume matricialmente:

$$A^\dagger = \underbrace{V}_{C_{BE}} \underbrace{\begin{bmatrix} \frac{1}{6} & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 \end{bmatrix}}_{\Sigma^\dagger} \underbrace{U^t}_{C_{EC}} = \begin{bmatrix} 1/4 & 1/4 & 0 \\ \sqrt{2}/12 & -\sqrt{2}/12 & 0 \\ 1/4 & 1/4 & 0 \end{bmatrix}$$

Donde la matriz $\Sigma^\dagger$ representa la transformación seudoinversa con respecto a las bases C y B, mientras que U^t y V realizan los cambios de coordenadas a la base canónica.

Esta misma matriz puede obtenerse a partir de la DVS reducida de A, calculada en 6.11 como $A = U_r D V_r^t$. Allí se encuentra efectivamente toda la información necesaria para invertir la transformación de Fil $A \to$ Col A:

$$A^\dagger = V_r D^{-1} U_r^t = \begin{bmatrix} 0 & 1/\sqrt{2} \\ 1 & 0 \\ 0 & 1/\sqrt{2} \end{bmatrix} \begin{bmatrix} 1/6 & 0 \\ 0 & 1/2 \end{bmatrix} \begin{bmatrix} 1/\sqrt{2} & -1/\sqrt{2} & 0 \\ 1/\sqrt{2} & 1/\sqrt{2} & 0 \end{bmatrix} = \begin{bmatrix} 1/4 & 1/4 & 0 \\ \sqrt{2}/12 & -\sqrt{2}/12 & 0 \\ 1/4 & 1/4 & 0 \end{bmatrix}$$

A continuación mostraremos cómo es posible aplicar la matriz $A^\dagger$ a la resolución de un problema de cuadrados mínimos. Consideremos el sistema incompatible

$$Ax = \begin{bmatrix} 1 & 3\sqrt{2} & 1 \\ 1 & -3\sqrt{2} & 1 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = b$$

Dado que la matriz A tiene sus columnas linealmente dependientes, podemos anticipar que habrá infinitas soluciones por cuadrados mínimos $\hat{x} = [\hat{x}_1\ \hat{x}_2\ \hat{x}_3]^t$. Ahora bien, dado que $b \notin$ Col A, resolver el problema de cuadrados mínimos equivale (como ya se vio en la sección 3.5 del capítulo 3) a encontrar los $\hat{x}$ tales que

$$A\hat{x} = P_{\text{Col}\,A}(b) \tag{6.13}$$

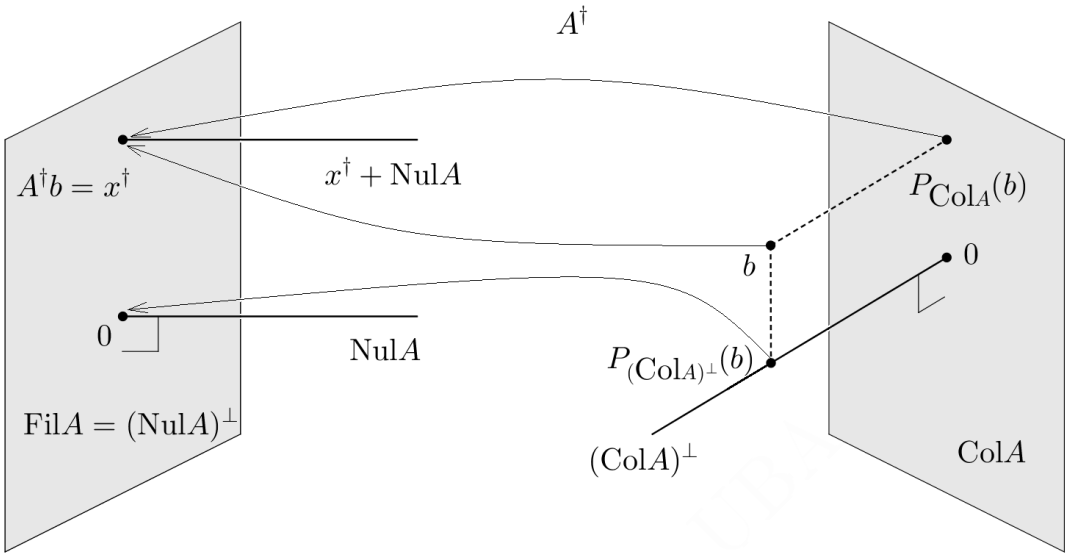


Figura 6.4: $A^\dagger$ y su relación con los subespacios fundamentales

Como sugiere la figura 6.4, aquí es donde podemos usar la seudoinversa $A^\dagger$: con ella obtenemos el único elemento $x^\dagger$ en Fil A que verifica (6.13). En efecto, si descomponemos a b como la suma de sus proyecciones sobre Col A y $(\mathrm{Col}\,A)^\perp$, y teniendo en cuenta que por su propia definición $A^\dagger$ se anula sobre los elementos de $(\mathrm{Col}\,A)^\perp$, resulta

$$A^\dagger b = A^\dagger \left(P_{\mathrm{Col}\,A}(b) + P_{(\mathrm{Col}\,A)^\perp}(b) \right) = A^\dagger P_{\mathrm{Col}\,A}(b) = x^\dagger$$

En nuestro ejemplo tenemos

$$x^\dagger = A^\dagger b = \begin{bmatrix} 1/4 & 1/4 & 0 \\ \sqrt{2}/12 & -\sqrt{2}/12 & 0 \\ 1/4 & 1/4 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 1/2 \\ 0 \\ 1/2 \end{bmatrix}$$

La solución $x^\dagger$ tendrá una característica notable dentro del conjunto de soluciones $\hat{x}$ por cuadrados mínimos: dado que los $\hat{x}$ surgen de resolver la ecuación normal

$$A^t A\hat{x} = A^t b$$

cualquier solución se obtiene como una solución particular sumada con una solución del sistema homogéneo asociado, es decir un elemento de $\mathrm{Nul}\,A^t A = \mathrm{Nul}\,A$. Por ende, podemos escribir la solución general por cuadrados mínimos en la forma

$$x^\dagger + \mathrm{Nul}\,A = \left\{ x^\dagger + x_0 : x_0 \in \mathrm{Nul}\,A \right\}$$

En nuestro ejemplo, este conjunto solución es una recta paralela a Nul A cuyo punto más cercano al origen es justamente $x^\dagger$. En otras palabras: $x^\dagger$ es la solución por cuadrados mínimos *de norma mínima*.

Dedicaremos el resto de esta sección a formalizar y generalizar las observaciones hechas hasta aquí.

Definición 6.5

Sea $A \in \mathbb{R}^{m\times n}$ con una DVS reducida $U_r D V_r^t$. La matriz

$$A^\dagger = V_r D^{-1} U_r^t \tag{6.14}$$

es la *matriz seudoinversa* de A. También se la conoce como *seudoinversa de Moore-Penrose*.

N Ya sabemos que la DVS reducida de una matriz A no es única. Sin embargo, para cualquier DVS reducida que se elija se obtiene la misma matriz seudoinversa. Esto se puede justificar pensando que $A^\dagger$ representa una transformación lineal que está totalmente determinada por cómo transforma a las bases de $\operatorname{Col} A$ y de $(\operatorname{Col} A)^\perp$. El lector interesado podrá encontrar detalles en la bibliografía.

Ejercicio.

6.9 Encuentre la matriz seudoinversa de la matriz estudiada en el ejercicio 6.2.

Teorema 6.6

Sean $A \in \mathbb{R}^{m\times n}$ y $A^\dagger \in \mathbb{R}^{n\times m}$ su seudoinversa. Entonces

1. $A^\dagger A = [P_{\operatorname{Fil} A}]_E$
2. $AA^\dagger = [P_{\operatorname{Col} A}]_E$

Demostración:
Consideramos una DVS reducida de A:

$$A = U_r D V_r^t$$

a partir de la cual obtenemos la seudoinversa

$$A^\dagger = V_r D^{-1} U_r^t$$

Las propiedades 1 y 2 surgen como consecuencia del teorema 6.5 al realizar los productos indicados

$$A^\dagger A = V_r D^{-1} U_r^t U_r D V_r^t = V_r V_r^t$$
$$AA^\dagger = U_r D V_r^t V_r D^{-1} U_r^t = U_r U_r^t$$

□

Ejercicio.

- Demuestre que si $A \in \mathbb{R}^{n\times n}$ es inversible entonces $A^\dagger = A^{-1}$.
- Demuestre que para cualquier $A \in \mathbb{R}^{m\times n}$ es

$$A^\dagger A A^\dagger = A^\dagger \qquad AA^\dagger A = A$$

- Demuestre que para cualquier $b \in \mathbb{R}^m$ el vector $A^\dagger b$ es un elemento de $\operatorname{Fil} A$.
  Sugerencia: probar que $P_{\operatorname{Fil} A}\left(A^\dagger b\right) = A^\dagger b$ usando el teorema 6.6.

Teorema 6.7

Sea $A \in \mathbb{R}^{m\times n}$ y sea $A^\dagger \in \mathbb{R}^{n\times m}$ su seudoinversa. Sea $b \in \mathbb{R}^m$. Entonces en el sistema de ecuaciones

$$Ax = b$$

el vector

$$x^\dagger = A^\dagger b$$

es la solución por cuadrados mínimos de norma mínima.

Demostración:
Para ver que $x^\dagger$ es solución por cuadrados mínimos, debemos probar que $Ax^\dagger = P_{\mathrm{Col}\,A}(b)$. Ahora bien, usando la propiedad 2 del teorema 6.6 tenemos

$$Ax^\dagger = AA^\dagger b = [P_{\mathrm{Col}\,A}]_E\, b$$

Para probar que $x^\dagger$ es la solución de norma mínima, tendremos en cuenta que las soluciones por cuadrados mínimos son de la forma $x^\dagger + x_0$ con $x_0 \in \mathrm{Nul}\,A$, como ya justificamos al comienzo de esta sección. Para calcular el cuadrado de la norma de una solución genérica $x^\dagger + x_0$ aprovechamos que $x^\dagger \in \mathrm{Fil}\,A = (\mathrm{Nul}\,A)^\perp$ y por ende podemos aplicar el teorema de Pitágoras

$$\left\|x^\dagger + x_0\right\|^2 = \left\|x^\dagger\right\|^2 + \|x_0\|^2 \geq \left\|x^\dagger\right\|^2$$

con lo cual obtenemos el resultado buscado. □

Ejercicio.

6.10 Considerando la matriz definida en el ejercicio 6.2:

- Resuelva el sistema de ecuaciones $Ax = b$ con $b = [1\ 1]^t$. Utilice la matriz seudoinversa calculada en el ejercicio 6.9.
- Compruebe que las matrices A y $A^\dagger$ verifican las propiedades del teorema 6.6.

7 Ecuaciones diferenciales y sistemas de ecuaciones diferenciales

7.1 Introducción

Este capítulo tiene por objetivo mostrar cómo pueden aplicarse los recursos del álgebra lineal para estudiar algunas ecuaciones diferenciales. Por eso es que la exposición quedará limitada a ciertas ecuaciones específicas, pudiendo los lectores interesados en profundizar estos temas consultar la abundante bibliografía que existe sobre el tema. De ningún modo pretende ser un curso de ecuaciones diferenciales, como sí lo son los textos [10] y [2], que hemos consultado para elaborar estas notas y recomendamos.

Comencemos considerando un tema clásico y familiar: el movimiento rectilíneo de una partícula. En la figura 7.1 se aprecia un sistema de referencia vertical en el que hemos elegido el sentido positivo "hacia abajo", puesto que pretendemos representar el movimiento de caída de los cuerpos.

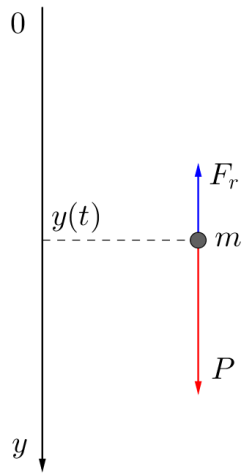


Figura 7.1: Sistema de referencia para la caída de un cuerpo

Dada una partícula de masa m, su altura será una función $y(t)$. Es decir, una cantidad y que depende del tiempo t. Para poder aplicar la segunda ley de Newton

$$F = ma \tag{7.1}$$

a las fuerzas que concurren en la partícula, recordemos que la primera derivada de la posición es la *velocidad instantánea*

$$y'(t) = v(t)$$

mientras que la segunda derivada es la *aceleración instantánea*

$$y''(t) = v'(t) = a(t)$$

Suponemos, en una primera aproximación, que no existe fuerza de rozamiento. Entonces la única fuerza que actúa sobre la partícula es el peso $P = mg$, donde g es la aceleración gravitatoria debida a la atracción terrestre (razonablemente podemos suponerla constante para alturas bajas). De esta forma, la segunda ley de Newton (7.1) resulta

$$mg = my''(t) \Leftrightarrow y''(t) = g \tag{7.2}$$

Vale decir que la posición $y(t)$ es una función con derivada segunda constante. O sea que se trata de una función polinómica de grado dos:

$$y(t) = \frac{1}{2}gt^2 + At + B \qquad \text{con } A, B \in \mathbb{R} \tag{7.3}$$

Es usual en esta clase de problemas conocer datos iniciales como la posición y_0 y la velocidad v_0 al comienzo del movimiento. En tal caso, al evaluar la fórmula (7.3) en $t = 0$ deducimos que $y(0) = B = y_0$. De manera similar, derivando la fórmula (7.3) y evaluando en $t = 0$ deducimos que $y'(0) = A = v_0$. Con esto llegamos a la clásica expresión

$$y(t) = \frac{1}{2}gt^2 + v_0 t + y_0 \tag{7.4}$$

Debemos tener en cuenta que la ecuación (7.4) es tan sólo una aproximación a la posición verdadera de la partícula. Esto se debe a las simplificaciones que hemos hecho, como suponer que no existe el rozamiento. Así, la fórmula (7.4) puede ser de utilidad para intervalos relativamente breves de tiempo, aunque no puede tomarse seriamente cuando $t \to \infty$.

Para dar cuenta de estas dificultades, incorporamos ahora la fuerza de rozamiento: como se ve en la figura 7.1, se trata de una fuerza de sentido opuesto al peso. El rozamiento es proporcional a la velocidad: [1]

$$F_r = -kv$$

donde k es una constante positiva que depende del aire y de las características aerodinámicas del cuerpo que cae. El signo menos da cuenta de que el sentido de F_r es siempre opuesto al del movimiento. En este caso, al aplicar la ley de Newton (7.1) queda

$$mg - ky'(t) = my''(t) \tag{7.5}$$

Encontrar las funciones $y(t)$ que verifican la relación (7.5) no es tan sencillo como en el caso anterior. Sin embargo, podemos anticipar algo aun cuando no conozcamos una fórmula explícita de $y(t)$. Dado que la velocidad $y'(t)$ tiende a crecer durante la caída, la fuerza peso y la fuerza de rozamiento tienden a compensarse. Por eso esperamos que a largo plazo la aceleración $y''(t)$ tienda a cero: este movimiento tiende a un movimiento de velocidad constante (es fácil de imaginar pensando, por ejemplo, en un paracaídas).

[1] En rigor, el fenómeno del rozamiento es un amplio tema de la física. Existen, por ejemplo, modelos alternativos en los que el rozamiento es proporcional a otra potencia v^n de la velocidad.

La relación (7.5) involucra derivadas de primer y segundo orden. Podemos simplificarla si la escribimos en términos de la velocidad $v(t) = y'(t)$, con lo cual obtenemos

$$mg - kv(t) = mv'(t) \Leftrightarrow v'(t) + \frac{k}{m}v(t) = g \tag{7.6}$$

Un recurso habitual para encontrar las funciones $v(t)$ que verifican esta relación es usar lo que se denomina un *factor integrante* . En este caso, el factor integrante es la función $e^{\frac{k}{m}t}$: como nunca se anula, si multiplicamos ambos miembros por este factor obtenemos un problema equivalente al original

$$\underbrace{v'(t)e^{\frac{k}{m}t} + \frac{k}{m}v(t)e^{\frac{k}{m}t}}_{\frac{d}{dt}\left(v(t)e^{\frac{k}{m}t}\right)} = ge^{\frac{k}{m}t}$$

con la ventaja de que el primer miembro es la derivada de un producto [2], por lo que integramos para obtener

$$v(t)e^{\frac{k}{m}t} = \frac{m}{k}ge^{\frac{k}{m}t} + C \qquad \text{con } C \in \mathbb{R}$$

y por lo tanto

$$v(t) = \frac{m}{k}g + Ce^{-\frac{k}{m}t} \qquad \text{con } C \in \mathbb{R}$$

Si además conocemos la velocidad inicial v_0, evaluando la solución en $t = 0$ llegamos a

$$v(t) = \frac{m}{k}g + \left(v_0 - \frac{m}{k}g\right)e^{-\frac{k}{m}t} \tag{7.7}$$

Como habíamos anticipado, la velocidad tiende a estabilizarse:

$$\lim_{t\to\infty} v(t) = \frac{m}{k}g$$

Las ecuaciones (7.2), (7.5) y (7.6) son ejemplos de lo que denominamos *ecuaciones diferenciales* , es decir, problemas en los que se trata de encontrar aquellas funciones que verifican cierta ecuación que involucra a sus derivadas. En particular (7.2) y (7.5) son *ecuaciones diferenciales de segundo orden* porque contienen derivadas segundas de la función incógnita, mientras que (7.6) es una *ecuación diferencial de primer orden*. En las secciones que siguen estudiaremos algunas de las ecuaciones llamadas *ecuaciones diferenciales lineales*, que abarcan a estos ejemplos como casos particulares.

7.2 Ecuación diferencial lineal de primer orden

En esta sección generalizamos el estudio de ecuaciones como (7.6). Ya vimos en la ecuación (7.6) cómo una ecuación de la forma

$$a_1(t)y'(t) + a_0(t)y(t) = b(t)$$

puede reescribirse como

$$y'(t) + \frac{a_0(t)}{a_1(t)}y(t) = \frac{b(t)}{a_1(t)}$$

en la medida en que la función $a_1(t)$ no se anule. Eso eso lo que asumiremos de aquí en adelante, así como la continuidad de la funciones $a_1(t)$, $a_0(t)$ y $b(t)$.

[2] Más adelante explicaremos un método general para obtener el factor integrante.

Definición 7.1

Una ecuación diferencial de la forma

$$y'(t) + p(t)y(t) = f(t) \tag{7.8}$$

(donde $p(t)$ y $f(t)$ son funciones continuas en cierto intervalo abierto I) se denomina ***ecuación diferencial lineal de primer orden***.

En el caso de que sea $f(t) = 0$ para todo $t \in I$ se dice que es una ecuación ***homogénea***.

Si además se buscan soluciones que verifiquen una condición de la forma $y(t_0) = y_0$ para cierto $t_0 \in I$ se tiene un ***problema de valor inicial***.

Ejercicio. Considere la ecuación diferencial homogénea

$$y' + \frac{1}{t}y = 0$$

- Verifique que la función $\phi : (0, +\infty) \to \mathbb{R}$, $\phi(t) = \frac{1}{t}$ es una solución.
- Encuentre dos soluciones más en el mismo intervalo $(0, +\infty)$.
- Encuentre una solución que verifique la condición inicial $y(2) = 4$.

Convención

A menudo abreviaremos la notación omitiendo la variable independiente de las funciones. Así, por ejemplo, en lugar de $y(t)$ escribiremos meramente y. El contexto debería permitir en cada caso discernir qué clase de funciones se está manejando.

Ya hemos visto en la sección anterior cómo una ecuación diferencial puede tener infinitas soluciones. También vimos que al imponer ciertas condiciones iniciales quedaba determinada una única solución. Esto permite suponer que, bajo hipótesis adecuadas, un problema de valor inicial tiene solución y además es única. El teorema que sigue establece esta afirmación para el caso de la ecuación diferencial lineal de primer orden. La demostración es especialmente instructiva porque además brinda un método de resolución de la ecuación.

Teorema 7.1

Dado un problema de valor inicial

$$\begin{cases} y' + py = f \\ y(t_0) = y_0 \end{cases}$$

como se definió en 7.1, existe una única solución $\phi : I \to \mathbb{R}$ que verifica $\phi(t_0) = y_0$.

Demostración:
Como en la sección anterior, multiplicamos por un factor integrante para transformar la ecuación. Consideramos entonces una primitiva de $p(t)$: esto es, una función $P(t)$ tal que $P' = p$. Si multiplicamos ambos miembros de la ecuación por el factor integrante $e^{P(t)}$ (que no se anula para ningún valor de t) obtenemos la ecuación equivalente

$$e^{P(t)}y' + e^{P(t)}P'(t)y = e^{P(t)}f$$

de manera tal que una función y es solución si y sólo si verifica que

$$\frac{d}{dt}\left[e^{P(t)}y\right] = e^{P(t)}f(t)$$

Por lo tanto $e^{P(t)}y(t)$ es una primitiva de $e^{P(t)}f(t)$:

$$e^{P(t)}y = \int e^{P(t)}f(t)dt \Leftrightarrow y = e^{-P(t)}\int e^{P(t)}f(t)dt$$

Si $F(t)$ es una primitiva de $e^{P(t)}f(t)$, cualquier otra primitiva difiere en una constante. Luego

$$y = e^{-P(t)}\left[F(t)+C\right] = F(t)e^{-P(t)} + Ce^{-P(t)} \qquad \text{con } C \in \mathbb{R} \tag{7.9}$$

Esta última expresión es la *solución general* de la ecuación diferencial definida en 7.1, pues contiene todas las posibles soluciones.

Finalmente, si en la solución general (7.9) imponemos la condición inicial resulta

$$y(t_0) = F(t_0)e^{-P(t_0)} + Ce^{-P(t_0)} = y_0$$

de donde se deduce que hay una sola elección posible de la constante de integración. □

Ejemplo 7.1

Hallaremos la solución del problema de valor inicial

$$\begin{cases} y' - ty = 0 \\ y(1) = -1 \end{cases}$$

Siguiendo los pasos de la demostración del teorema 7.1, calculamos un factor integrante hallando una primitiva de la función $p(t) = -t$. En este caso proponemos $P(t) = -\frac{t^2}{2}$ (aunque podríamos haber elegido cualquiera de la forma $P(t) = -\frac{t^2}{2} + k$). Luego el factor integrante es $e^{-\frac{t^2}{2}}$ y la ecuación diferencial original es equivalente a

$$e^{-\frac{t^2}{2}}y' - te^{-\frac{t^2}{2}}y = 0 \Leftrightarrow \frac{d}{dt}\left[e^{-\frac{t^2}{2}}y\right] = 0$$

con lo cual deducimos que

$$e^{-\frac{t^2}{2}}y = C \Leftrightarrow y = Ce^{\frac{t^2}{2}} \qquad \text{con } C \in \mathbb{R}$$

Al imponer la condición inicial deducimos el único valor adecuado de $C \in \mathbb{R}$:

$$y(1) = Ce^{\frac{1}{2}} = -1 \Rightarrow C = -e^{-\frac{1}{2}}$$

luego la solución al problema de valor incial es

$$\phi(t) = -e^{-\frac{1}{2}}e^{\frac{t^2}{2}} = -e^{\frac{t^2-1}{2}}$$

En la figura 7.2 se aprecian varios elementos de la solución general, incluyendo la solución del problema de valor inicial. Este gráfico permite una interpretación sencilla del teorema 7.1 de existencia y unicidad de soluciones: de las infinitas curvas solución de la ecuación diferencial, estamos eligiendo la única que pasa por el punto $(1,-1)$.

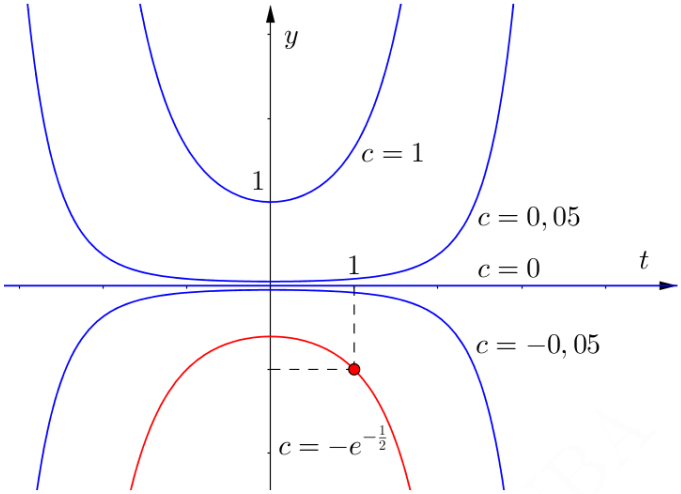


Figura 7.2: Diversas curvas solución

Ejemplo 7.2

Consideramos el problema de valor inicial

$$\begin{cases} y' - \tan t \cdot y = \operatorname{sen} t \\ y(0) = 0 \end{cases}$$

Como $\tan t$ está definida y es continua en los intervalos abiertos de la forma $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right), \left(\frac{\pi}{2}, \frac{3\pi}{2}\right)$... y el punto en el cual tenemos dada la condición inicial es $t = 0$, trabajamos en el intervalo $\left(-\frac{\pi}{2}, \frac{\pi}{2}\right)$. Para calcular el factor integrante

$$\int -\tan t \, dt = -\int \frac{\operatorname{sen} t}{\cos t} dt = \ln|\cos t| + C$$

En el intervalo considerado es $|\cos t| = \cos t$, por lo que elegimos factor integrante $e^{\ln \cos t} = \cos t$. La ecuación original es entonces equivalente a

$$\cos t\, y' - \cos t \frac{\operatorname{sen} t}{\cos t} y = \operatorname{sen} t \cos t \Leftrightarrow \frac{d}{dt}\left[\cos t\, y\right] = \operatorname{sen} t \cos t$$

con lo cual

$$\cos t\, y = -\frac{\cos^2 t}{2} + C \qquad \text{con } C \in \mathbb{R}$$

y la solución general resulta

$$y(t) = -\frac{\cos t}{2} + \frac{C}{\cos t} \qquad \text{con } C \in \mathbb{R}$$

Imponiendo la condición inicial resulta

$$y(0) = -\frac{1}{2} + C = 0 \Rightarrow C = \frac{1}{2}$$

con lo cual la solución buscada es

$$\phi(t) = -\frac{\cos t}{2} + \frac{1}{2\cos t}$$

En la figura 7.3 se aprecian diversas curvas solución y los intervalos en los que están definidas las funciones $y(t)$.

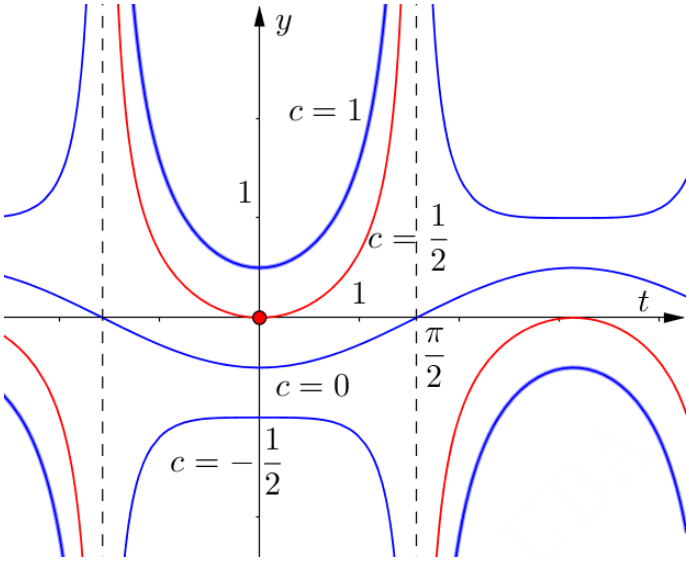


Figura 7.3: Diversas curvas solución

Ejercicio.

7.1 Encuentre la solución del problema de valor inicial

$$\begin{cases} y' + \frac{\cosh t}{\operatorname{senh} t} y = \frac{-t}{\operatorname{senh} t} \\ y(1) = \frac{1}{2\operatorname{senh} 1} \end{cases}$$

e indique en qué intervalo es válida.

7.3 Estructura de las soluciones en la ecuación de primer orden

En esta sección aplicaremos los conceptos del álgebra lineal para caracterizar las soluciones de la ecuación diferencial lineal de primer orden. Para ello, comenzamos por señalar que la ecuación diferencial lineal definida en 7.1 puede pensarse en términos de una transformación lineal. En efecto, si definimos la función

$$L : \mathscr{C}^1(I) \to \mathscr{C}(I),\ L(y) = y' + py \tag{7.10}$$

es inmediato que L es una transformación lineal (es usual en este contexto denominarlo *operador lineal*). Por lo tanto, resolver la ecuación diferencial lineal de primer orden

$$y' + py = f$$

equivale a encontrar las funciones $y \in \mathscr{C}^1(I)$ tales que $L(y) = f$. En otras palabras, la solución general es la preimagen de f por L:

$$L^{-1}(f) = \{y \in \mathscr{C}^1(I) : L(y) = f\} = \{y \in \mathscr{C}^1(I) : y' + py = f\}$$

(Insistimos con la advertencia, ya hecha en el capítulo de transformaciones lineales, de no confundir preimagen con función inversa). En el caso particular de la ecuación homógenea, la solución general es la preimagen de la función nula, o sea el núcleo del operador: $L^{-1}(0) = \operatorname{Nu} L$.

En lo que sigue veremos cómo la estructura de la solución general es totalmente análoga a la que ya hemos estudiado para los sistemas de ecuaciones lineales. Esto se debe a que resolver un sistema de ecuaciones lineales, tanto como resolver una ecuación diferencial lineal, consiste en estudiar preimágenes de una transformación lineal.

Teorema 7.2

Dada una transformación lineal $T : V \to W$ entre dos espacios vectoriales V y W, dados $w \in \operatorname{Im} T$ y $v_p \in V$ tal que $T(v_p) = w$, el conjunto de soluciones de la ecuación

$$T(v) = w$$

es

$$T^{-1}(w) = v_p + \operatorname{Nu} T = \{v_p + v_h : v_h \in \operatorname{Nu} T\}$$

Demostración:
Debemos probar una doble inclusión. Probaremos en primer lugar que todos los elementos de la forma $v_p + v_h$ pertenecen a $T^{-1}(w)$. Para ello, alcanza con ver que $T(v_p + v_h) = T(v_p) + T(v_h) = w + 0 = w$.
Consideramos ahora un elemento $v \in T^{-1}(w)$. Probaremos que es de la forma $v_p + v_h$ para algún $v_h \in \operatorname{Nu} T$. En efecto, dado que $v \in T^{-1}(w)$ y que T es lineal, resulta $T(v - v_p) = T(v) - T(v_p) = w - w = 0$. Por lo tanto $v - v_p$ es un elemento de $\operatorname{Nu} T$: $v - v_p = v_h$ para algún $v_h \in \operatorname{Nu} T$, como queríamos. □

El resultado anterior es conocido para el caso de los sistemas de ecuaciones lineales: cualquier solución es una solución particular v_p sumada con algún elemento del núcleo, esto es, con una solución del problema homogéneo. Veamos cómo se expresa en el caso de la ecuación diferencial lineal.

Teorema 7.3

Sea y_p una solución particular de la ecuación diferencial lineal (7.8). Entonces la función $y \in \mathscr{C}^1(I)$ es solución de dicha ecuación si y sólo si es de la forma $y = y_p + y_h$ siendo y_h una solución de la ecuación homogénea asociada.

Este resultado nos dice que para hallar todas las soluciones de la ecuación diferencial lineal alcanza con resolver la ecuación homogénea asociada y hallar *una* solución particular de la ecuación no homogénea.

En cuanto a la la ecuación homogénea, ya hemos mencionado que su conjunto solución es el núcleo de L, por lo que se trata de un subespacio de $\mathscr{C}^1(I)$. Dejamos como ejercicio para el lector adaptar la demostración del teorema 7.1 al caso homogéneo $y' + py = 0$, para ver que la solución general es

$$y = Ce^{-P(t)} \qquad \text{con } C \in \mathbb{R}$$

donde $P(t)$ es una primitiva de $p(t)$. Por lo tanto el núcleo del operador L está generado por $e^{-P(t)}$ y es un subespacio de dimensión uno. Resumimos estas observaciones en un teorema.

Teorema 7.4

Consideramos la ecuación diferencial homogénea

$$y' + py = 0$$

con p continua en un intervalo abierto I. Entonces

1. Si $P(t)$ es una primitiva de $p(t)$, entonces $\phi_h(t) = e^{-P(t)}$ es una solución no trivial de la ecuación diferencial.
2. Dada cualquier solución no trivial ϕ_h de la ecuación diferencial, $\{\phi_h\}$ es una base de soluciones. Esto es, cualquier otra solución es de la forma
$$y_h = C\phi_h \qquad \text{con } C \in \mathbb{R}$$

Ejercicio.

7.2 Considere el operador lineal

$$L : \mathscr{C}^1(0,+\infty) \to \mathscr{C}(0,+\infty),\ L(y) = y' - \frac{1}{t}y$$

- Resuelva la ecuación diferencial $L(y) = 3t$
- Encuentre Nu L.

Ejercicio. Considere una vez más el problema de valor inicial del ejercicio 7.1.

- Defina un operador lineal $L : \mathscr{C}^1(I) \to \mathscr{C}(I)$ que permita representar a la ecuación diferencial.
- Calcule el núcleo de dicho operador.
- Calcule $L^{-1}\left(\frac{-t}{\operatorname{senh} t}\right)$
- Elija el elemento de $L^{-1}\left(\frac{-t}{\operatorname{senh} t}\right)$ que verifica la condición inicial dada.

7.4 Ecuación diferencial lineal de segundo orden

En la introducción de este capítulo hemos visto ejemplos de ecuaciones diferenciales con derivadas segundas, como (7.2) y (7.5). La forma general de la ecuación diferencial lineal de segundo orden es

$$a_2(t)y'' + a_1(t)y' + a_0(t)y = b(t)$$

con $a_2(t)$, $a_1(t)$, $a_0(t)$ y $b(t)$ funciones continuas en cierto intervalo abierto. Asumiendo que $a_2(t)$ no se anula en dicho intervalo, podemos reescribir la ecuación como

$$y'' + \frac{a_1(t)}{a_2(t)}y' + \frac{a_0(t)}{a_2(t)}y = \frac{b(t)}{a_2(t)}$$

Para la ecuación de primer orden habíamos exhibido un método general de resolución en 7.1. Como es de imaginar, el caso de segundo orden presenta mayores dificultades. Por ello es que limitaremos nuestra atención al caso en que las funciones $a_2(t)$, $a_1(t)$ y $a_0(t)$ son constantes.

Definición 7.2

Una ecuación diferencial de la forma

$$y'' + a_1 y' + a_0 y = f(t) \tag{7.11}$$

donde $a_1, a_0 \in \mathbb{R}$ y $f(t)$ es una función continua en cierto intervalo abierto I, se denomina ***ecuación diferencial lineal de segundo orden con coeficientes constantes***.

En el caso de que sea $f(t) = 0$ para todo $t \in I$ se dice que es una ecuación ***homogénea***.

Si además se buscan soluciones que verifiquen una condición de la forma

$$\begin{cases} y(t_0) = a \\ y'(t_0) = b \end{cases}$$

para cierto $t_0 \in I$ y a, b dos valores dados, se tiene un ***problema de valores iniciales***.

Muchos de los recursos y resultados acerca de la ecuación de primer orden pueden adaptarse sin dificultad para la de segundo orden. Como en (7.10), si definimos un operador lineal

$$L : \mathscr{C}^2(I) \to \mathscr{C}(I),\ L(y) = y'' + a_1 y' + a_0 y \tag{7.12}$$

resolver la ecuación de segundo orden definida en 7.2 equivale a encontrar la preimagen de f por L:

$$L^{-1}(f) = \{y \in \mathscr{C}^2(I) : L(y) = f\} = \{y \in \mathscr{C}^2(I) : y'' + a_1 y' + a_0 y = f\}$$

En particular, la solución de la ecuación homogénea es $L^{-1}(0)$, es decir el subespacio $\operatorname{Nu} L$.

Aplicando el teorema 7.2 como en el caso de primer orden, podemos concluir nuevamente que las soluciones de la ecuación se obtienen con una solución particular sumada a una solución de la ecuación homogénea.

Teorema 7.5

Sea y_p una solución particular de la ecuación diferencial lineal (7.11). Entonces la función $y \in \mathscr{C}^2(I)$ es solución de dicha ecuación si y sólo si es de la forma $y = y_p + y_h$ con y_h una solución de la ecuación homogénea asociada.

Volvamos a pensar la ecuación diferencial (7.2) con que describimos la caída libre de los cuerpos

$$y'' = g$$

aplicando ahora los conceptos que hemos estudiado. Ya sabemos que la solución general es la descrita en (7.3)

$$y(t) = \frac{1}{2} g t^2 + At + B \qquad \text{con } A, B \in \mathbb{R}$$

y ahora podemos señalar que $y_p = \frac{1}{2} g t^2$ es una solución particular, mientras que las funciones de la forma $At + B$ con $A, B \in \mathbb{R}$ son soluciones de la ecuación homogénea asociada

$$y'' = 0$$

o, si se prefiere, constituyen el núcleo del operador lineal

$$L : \mathscr{C}^2(I) \to \mathscr{C}(I),\ L(y) = y''$$

Queremos destacar que

$$\operatorname{Nu} L = \operatorname{gen}\{t, 1\}$$

es decir, que el núcleo del operador L es de dimensión dos.

Parece razonable conjeturar que el núcleo de cualquier operador de la forma (7.12) será de dimensión dos. Más adelante probaremos que esto es efectivamente cierto, lo cual resulta de gran importancia. En efecto, si el núcleo del operador es de dimensión dos, está generado por dos funciones y_1, y_2 linealmente independientes. Luego para resolver una ecuación como la definida en 7.2 alcanzará con

1. Encontrar dos soluciones y_1, y_2 de la ecuación homogénea que sean linealmente independientes.
2. Encontrar una solución particular y_p de la ecuación.

y la solución general será de la forma

$$y_p + Ay_1 + By_2 \qquad \text{con } A, B \in \mathbb{R} \tag{7.13}$$

El ejemplo que sigue ilustra este método.

Ejemplo 7.3

Consideramos el problema de valores iniciales

$$\begin{cases} y'' - 9y = t^2 + 2 \\ y(0) = 0 \\ y'(0) = 1 \end{cases}$$

En primer lugar buscamos las soluciones de la ecuación homogénea, sabiendo que alcanza con encontrar dos soluciones independientes:

$$y'' - 9y = 0 \Leftrightarrow y'' = 9y$$

Como se trata de una función cuya derivada segunda es un múltiplo de la misma función, proponemos soluciones de la forma $y(t) = e^{rt}$ para algún factor r a determinar. Reemplazando en la ecuación obtenemos

$$r^2 e^{rt} - 9e^{rt} = 0 \Leftrightarrow e^{rt}\left(r^2 - 9\right) = 0$$

y como $e^{rt} > 0$ para cualquier valor de $t \in \mathbb{R}$, debe ser

$$\left(r^2 - 9\right) = 0$$

con lo cual es $r = 3$ ó $r = -3$ y obtenemos las soluciones

$$y_1(t) = e^{3t} \qquad y_2(t) = e^{-3t}$$

Dejamos como ejercicio comprobar que son efectivamente linealmente independientes. Luego, las soluciones de la ecuación homogénea son de la forma

$$Ae^{3t} + Be^{-3t} \qquad \text{con } A, B \in \mathbb{R}$$

Buscamos ahora una solución particular. Dado que la ecuación contiene derivadas que deben dar por resultado un polinomio, podemos suponer que existe una solución polinómica. Proponemos entonces una función de la forma $y_p(t) = Ct^2 + Dt + E$. Al reemplazar en la ecuación queda

$$2C - 9(Ct^2 + Dt + E) = t^2 + 2 \Leftrightarrow -9Ct^2 - 9Dt + 2C - 9E = t^2 + 2$$

e igualando los coeficientes de los polinomios deducimos que

$$C = -\frac{1}{9} \qquad D = 0 \qquad E = -\frac{20}{81}$$

de modo tal que la solución particular es

$$y_p(t) = -\frac{1}{9}t^2 - \frac{20}{81}$$

y la solución general resulta

$$-\frac{1}{9}t^2 - \frac{20}{81} + Ae^{3t} + Be^{-3t} \qquad \text{con } A, B \in \mathbb{R}$$

De toda esta familia de funciones debemos elegir aquella función $y(t)$ que verifique las condiciones iniciales:

$$y(0) = -\frac{20}{81} + A + B = 0$$
$$y'(0) = 3A - 3B = 1$$

de donde deducimos que $A = \frac{47}{162}$ y $B = -\frac{7}{162}$. La solución al problema de valores iniciales resulta

$$y(t) = -\frac{1}{9}t^2 - \frac{20}{81} + \frac{47}{162}e^{3t} - \frac{7}{162}e^{-3t}$$

Ejercicio.

7.3 Resuelva el siguiente problema de valores iniciales

$$\begin{cases} y'' - 4y = t^2 + t \\ y(0) = 1 \\ y'(0) = 0 \end{cases}$$

7.5 La ecuación homogénea de segundo orden

Como ya habíamos anticipado en la sección anterior, probaremos que el conjunto solución de la ecuación homogénea

$$y'' + a_1 y' + a_0 = 0 \tag{7.14}$$

definida en 7.2 es un subespacio de $\mathscr{C}^2(I)$ de dimensión dos. Más aun, indicaremos un método para calcular las soluciones. La demostración es un poco extensa y los lectores que tan sólo busquen un resumen del método podrán encontrarlo en los teoremas de esta sección. Para los lectores más curiosos ofrecemos los lineamientos de la demostración, aunque omitimos algunos detalles. Consideramos que leer cuidadosamente los argumentos puede ser un buen repaso de los temas estudiados hasta ahora y abrir un interesante panorama acerca de cómo se aplican las ideas del álgebra lineal.

Comenzamos con una definición que nos permitirá caracterizar al operador definido en (7.12).

Definición 7.3

Adoptando la notación

$$Dy = y' \qquad D^2y = D(Dy) = y'' \qquad \dots \qquad D^n y = y^{(n)} \dots$$

y dado un polinomio $p(r) = a_n r^n + a_{n-1} r^{n-1} + \dots + a_1 r + a_0$ definimos el operador lineal

$$p(D) : \mathscr{C}^n(I) \to \mathscr{C}(I),\ p(D)y = a_n D^n y + a_{n-1} D^{n-1} y + \dots + a_1 Dy + a_0 y$$

N Destacamos la completa analogía que existe entre el operador $p(D)$ y el polinomio matricial $p(A)$ estudiado en el capítulo 4: en el caso de $p(A)$ la matriz A también representa un operador, aunque entre espacios de dimensión finita.

Con la definición 7.3 podemos reescribir la ecuación homogénea (7.14) como

$$p(D)y = 0 \qquad \text{con } p(r) = r^2 + a_1 r + a_0 \tag{7.15}$$

En lo que sigue, aprovecharemos la factorización de este polinomio para resolver la ecuación.

Ejemplo 7.4

Consideramos la ecuación homogénea

$$y'' + y' - 2y = 0 \Leftrightarrow D^2 y + Dy - 2y = 0 \Leftrightarrow (D^2 + D - 2)y = 0$$

En este caso, el polinomio definido en (7.15) es $p(r) = r^2 + r - 2 = (r-1)(r+2)$. Veamos que el operador $(D^2 + D - 2)$ puede factorizarse como $(D-1)(D+2)$. Para ello, probamos que $(D-1)(D+2)$ aplicado a cualquier función da exactamente lo mismo que la aplicación de $(D^2 + D - 2)$. Usamos las definiciones y la propiedad asociativa de los operadores:

$$\begin{aligned}(D-1)(D+2)y &= (D-1)\left[(D+2)y\right] \\ &= (D-1)(Dy + 2y) \\ &= D(Dy + 2y) - (Dy + 2y) \\ &= (D^2 + D - 2)y\end{aligned}$$

La ecuación original puede escribirse entonces como

$$(D-1)(D+2)y = 0 \Leftrightarrow (D-1)\left[(D+2)y\right] = 0$$

vale decir que la función $(D+2)y$ debe ser un elemento del núcleo del operador $(D-1)$. Este núcleo no es difícil de calcular:

$$(D-1)v = 0 \Leftrightarrow Dv = v \Leftrightarrow v = ke^t \qquad \text{con } k \in \mathbb{R}$$

Hemos reducido la ecuación original a la ecuación de primer orden

$$(D+2)y = ke^t \Leftrightarrow y' + 2y = ke^t \tag{7.16}$$

que podemos resolver multiplicando por el factor integrante e^{2t}:

$$\begin{aligned}e^{2t}y' + e^{2t}2y &= ke^{3t} \\ \left(e^{2t}y\right)' &= ke^{3t} \\ e^{2t}y &= \frac{k}{3}e^{3t} + C_2 \qquad \text{con } C_2 \in \mathbb{R}\end{aligned}$$

Entonces la solución general de (7.16), que coincide con la de la ecuación homogénea, es el conjunto de funciones de la forma

$$y(t) = \frac{k}{3}e^t + C_2 e^{-2t}$$

y renombrando las constantes queda

$$y(t) = C_1 e^t + C_2 e^{-2t} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

Los argumentos usados en el ejemplo anterior se pueden adaptar a cualquier caso en que las raíces del polinomio $p(r) = r^2 + a_1 r + a_0$ sean dos números reales distintos. De allí el interés en la siguiente definición.

Definición 7.4

Dada una ecuación diferencial como la definida en 7.2, la ecuación

$$r^2 + a_1 r + a_0 = 0$$

se conoce como su ***ecuación característica***.

Dejamos como ejercicio para el lector generalizar el trabajo hecho en el ejemplo para demostrar el teorema siguiente.

Teorema 7.6 (Raíces reales distintas)

Si la ecuación característica de la ecuación homogénea (7.14) tiene dos raíces reales distintas r_1 y r_2, entonces $\{e^{r_1 t}, e^{r_2 t}\}$ es una base del conjunto de soluciones y la solución general está formada por las funciones de la forma

$$y = C_1 e^{r_1 t} + C_2 e^{r_2 t} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

Ejercicio.

7.4 Resuelva el siguiente problema de valores iniciales

$$\begin{cases} y'' + y' - 12y = 0 \\ y(0) = 1 \\ y'(0) = 2 \end{cases}$$

Si la ecuación característica tiene una raíz real doble, no podremos encontrar dos soluciones independientes de la forma $e^{r_1 t}, e^{r_2 t}$. El teorema que sigue caracteriza al conjunto solución de tal caso.

Teorema 7.7 (Raíz real doble)

Si la ecuación característica de la ecuación homogénea (7.14) tiene una raíz real doble r_1, entonces $\{te^{r_1 t}, e^{r_1 t}\}$ es una base del conjunto de soluciones y la solución general está formada por las funciones de la forma

$$y = C_1 te^{r_1 t} + C_2 e^{r_1 t} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

Demostración:
Como en este caso es $(r - r_1)^2 = 0$, la ecuación (7.14) se escribe como

$$(D - r_1)(D - r_1)y = 0 \Leftrightarrow (D - r_1)\left[(D - r_1)y\right] = 0$$

Luego la función $(D - r_1)y$ debe ser un elemento del núcleo del operador $(D - r_1)$. Vale decir que $(D - r_1)y$ debe ser de la forma

$$(D - r_1)y = C_1 e^{r_1 t} \qquad \text{con } C_1 \in \mathbb{R}$$

De esta forma, la ecuación original es equivalente a la ecuación de primer orden

$$y' - r_1 y = C_1 e^{r_1 t} \tag{7.17}$$

Al multiplicar por el factor integrante $e^{-r_1 t}$ obtenemos

$$\begin{aligned} e^{-r_1 t} y' - r_1 e^{-r_1 t} y &= C_1 \\ \left(e^{-r_1 t} y\right)' &= C_1 \\ e^{-r_1 t} y &= C_1 t + C_2 \qquad \text{con } C_2 \in \mathbb{R} \end{aligned}$$

Luego la solución general de la ecuación (7.17) y por ende de la ecuación homogénea, es

$$y(t) = C_1 t e^{r_1 t} + C_2 e^{r_1 t}$$

Queda como ejercicio verificar que las funciones $te^{r_1 t}$ y $e^{r_1 t}$ son independientes. □

Ejercicio.

7.5 Resuelva el siguiente problema de valores iniciales

$$\begin{cases} y'' + 4y' + 4y = 0 \\ y(0) = 1 \\ y'(0) = 0 \end{cases}$$

Siendo el polinomio definido en (7.15) con coeficientes reales, sólo nos queda por considerar el caso en que tiene dos raíces complejas conjugadas. Para poder pensar esta situación, consideraremos funciones de $I \to \mathbb{C}$ donde I es un intervalo abierto en $\mathbb{R}$. Podemos representarlas con la forma general

$$f(t) = u(t) + i\,v(t)$$

donde $u(t)$ y $v(t)$ son funciones de $I \to \mathbb{R}$. En caso de que $u(t)$ y $v(t)$ sean derivables, la derivación se define mediante[3]

$$f'(t) = u'(t) + i\,v'(t)$$

La resolución de las ecuaciones diferenciales que hemos estudiado puede extenderse sin dificultades a esta clase de funciones. En particular, prestaremos atención a la función exponencial e^{it}. De acuerdo con la fórmula de Euler, es

$$e^{it} = \cos t + i \operatorname{sen} t$$

Luego, definiendo para $b \in \mathbb{R}$ la función

$$f : \mathbb{R} \to \mathbb{C},\ f(t) = e^{ibt} = \underbrace{\cos(bt)}_{u(t)} + i \underbrace{\operatorname{sen}(bt)}_{v(t)}$$

resulta

$$\begin{aligned} \left(e^{ibt}\right)' &= -b\operatorname{sen}(bt) + ib\cos(bt) \\ &= ib\left[\cos(bt) + i\operatorname{sen}(bt)\right] \\ &= ibe^{ibt} \end{aligned}$$

como podía esperarse.

[3] Esta escueta presentación de las funciones con valores complejos no pretende ser rigurosa. Remitimos a los lectores interesados en el tema a cualquier texto de análisis matemático.

Ejercicio. Demuestre que para cualquier número complejo $z = a + bi$ resulta

$$\left(e^{zt}\right)' = ze^{zt}$$

Usando estas funciones y sus propiedades, se puede probar un teorema totalmente análogo al 7.6 demostrado para raíces reales distintas, con lo cual la solución general está formada por las funciones de la forma

$$y = z_1 e^{r_1 t} + z_2 e^{r_2 t}$$

donde $r_1 = a + bi$ y $r_2 = a - bi$ son las raíces conjugadas y $z_1, z_2 \in \mathbb{C}$. Usualmente, de toda esta familia ampliada de soluciones interesan las funciones a valores reales, como las que habíamos obtenido en los problemas anteriores. Para hallar estas funciones a valores reales, debemos desarrollar la expresión de la solución general:

$$\begin{aligned} z_1 e^{r_1 t} + z_2 e^{r_2 t} &= z_1 e^{(a+bi)t} + z_2 e^{(a-bi)t} \\ &= z_1 e^{at} e^{bit} + z_2 e^{at} e^{-bit} \\ &= z_1 e^{at} [\cos(bt) + i \operatorname{sen}(bt)] + z_2 e^{at} [\cos(bt) - i \operatorname{sen}(bt)] \\ &= e^{at} [(z_1 + z_2)\cos(bt) + (iz_1 - iz_2)\operatorname{sen}(bt)] \end{aligned}$$

Dado que $z_1, z_2 \in \mathbb{C}$ son arbitrarias pueden elegirse de manera que verifiquen el sistema (siempre compatible)

$$\begin{aligned} z_1 + z_2 &= C_1 \\ iz_1 - iz_2 &= C_2 \end{aligned}$$

donde C_1, C_2 son arbitrarios. En resumen, las soluciones son de la forma

$$e^{at} [C_1 \cos(bt) + C_2 \operatorname{sen}(bt)]$$

y llegamos a la solución general de la ecuación homogénea con funciones reales.

Teorema 7.8 (Raíces complejas conjugadas)

Si la ecuación característica de la ecuación homogénea (7.14) tiene dos raíces complejas conjugadas $r_1 = a + bi$ y $r_2 = a - bi$ (con $b \neq 0$), entonces $\{e^{at}\cos(bt), e^{at}\operatorname{sen}(bt)\}$ es una base del conjunto de soluciones y la solución general está formada por las funciones de la forma

$$y = C_1 e^{at}\cos(bt) + C_2 e^{at}\operatorname{sen}(bt) \qquad \text{con } C_1, C_2 \in \mathbb{R}\ ^a$$

[a] Eventualmente, estas constantes podrían ser números complejos, aunque en esta exposición nos restringimos a las funciones con valores reales.

Ejemplo 7.5 (Problema de valores en la frontera)

Consideramos la ecuación diferencial

$$y'' + 16y = 0$$

cuya ecuación característica es $r^2 + 16 = 0$. Sus raíces son los números complejos conjugados $r_1 = 4i$ y $r_2 = -4i$. De acuerdo con el teorema 7.8, una base del conjunto de soluciones será $\{e^{0\cdot t}\cos(4t), e^{0\cdot t}\operatorname{sen}(4t)\} = \{\cos(4t), \operatorname{sen}(4t)\}$. La solución general es

$$y = C_1\cos(4t) + C_2\operatorname{sen}(4t) \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

En las aplicaciones, es muy común encontrar problemas en que la solución debe verificar ciertas condiciones en los extremos de un intervalo. Es lo que se conoce como un *problema con valores en la frontera*. Consideremos, por ejemplo, el problema

$$\begin{cases} y'' + 16y = 0 \\ y(0) = 0 \\ y\left(\frac{\pi}{8}\right) = 1 \end{cases}$$

Al evaluar la solución general en $t = 0$ deducimos que debe ser $C_1 = 0$, mientras que al evaluar en $t = \frac{\pi}{8}$ deducimos que $C_2 = 1$. Por lo tanto, la única solución al problema de valor en la frontera es

$$y(t) = \operatorname{sen}(4t)$$

Ejercicio.

7.6 Resuelva, cuando sea posible, los siguientes problema de valores en la frontera

$$\begin{cases} y'' + 16y = 0 \\ y(0) = 0 \\ y\left(\frac{\pi}{4}\right) = 0 \end{cases} \qquad \begin{cases} y'' + 16y = 0 \\ y(0) = 0 \\ y\left(\frac{\pi}{4}\right) = 1 \end{cases}$$

Ejercicio.

7.7 Resuelva el siguiente problema de valores iniciales

$$\begin{cases} y'' - 2y' + 5y = 0 \\ y(0) = 2 \\ y'(0) = -1 \end{cases}$$

7.6 Método de los coeficientes indeterminados

Como ya señalamos en el teorema 7.5, para resolver la ecuación de segundo orden definida en 7.2 necesitamos la solución general de la ecuación homogénea y *una* solución particular. En la sección anterior hemos indicado cómo resolver la ecuación homogénea. En cuanto a encontrar una solución particular, en el ejemplo 7.3 lo hicimos basándonos en una conjetura: dado que el término independiente de la ecuación era un polinomio, propusimos como solución un polinomio con coeficientes a determinar. Veamos algunos ejemplos más en que se usan recursos similares.

Ejemplo 7.6

Consideramos la ecuación diferencial

$$y'' - 2y' + y = 3e^{2t}$$

cuya ecuación característica es $r^2 - 2r + 1 = (r-1)^2 = 0$. De acuerdo con el teorema 7.7, la solución general del problema homogéneo tiene a todas las funciones de la forma

$$y = C_1 te^t + C_2 e^t \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

Para encontrar una solución particular, observamos que las derivadas de cualquier orden de la función e^{2t} son múl-

tiplos de la misma función. Luego, si la reemplazamos en la expresión $y'' - 2y' + y$ obtendremos un múltiplo de e^{2t}. Para que ese múltiplo sea el adecuado, consideraremos la función $y_p(t) = Ae^{2t}$:

$$y_p'' - 2y_p' + y_p = 4Ae^{2t} - 4Ae^{2t} + Ae^{2t} = Ae^{2t} = 3e^{2t}$$

de donde deducimos que $A = 3$; la solución particular buscada es $y_p(t) = 3e^{2t}$ y la solución general resulta

$$y = 3e^{2t} + C_1 te^t + C_2 e^t \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

Ejemplo 7.7

Consideramos ahora el caso

$$y'' - 3y' + 2y = 3e^t$$

En este caso, la ecuación característica es $r^2 - 3r + 2 = (r-1)(r-2) = 0$. De acuerdo con el teorema 7.6, la solución general de la ecuación homogénea tiene a todas las funciones de la forma

$$y = C_1 e^t + C_2 e^{2t} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

La novedad en este problema es que no podemos proponer soluciones de la forma $y_p(t) = Ae^t$, puesto que son soluciones del problema homogéneo y se anulan al reemplazarlas en la expresión $y'' - 3y' + 2y$. Proponemos entonces $y_p(t) = Ate^t$. Antes de reemplazar en la ecuación, calculamos las derivadas necesarias:

$$y_p' = Ae^t + Ate^t$$
$$y_p'' = Ae^t + Ae^t + Ate^t = 2Ae^t + Ate^t$$

y entonces

$$\begin{aligned} 3e^t &= y_p'' - 3y_p' + 2y_p \\ &= 2Ae^t + Ate^t - 3\left(Ae^t + Ate^t\right) + 2Ate^t \\ &= -Ae^t \end{aligned}$$

de donde deducimos que $A = -3$; la solución particular buscada es $y_p(t) = -3te^t$ y la solución general resulta

$$y = -3te^t + C_1 e^t + C_2 e^{2t} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

Ejemplo 7.8

La ecuación

$$y'' - 3y' + 2y = 2\cos t$$

tiene asociada la misma ecuación homogénea que el ejemplo anterior. Sólo nos falta entones encontrar una solución particular. En este caso, debemos tener en cuenta que las derivadas de la función coseno son, alternativamente, múltiplos del seno y del coseno. Proponemos entonces $y_p(t) = A\cos t + B\,\text{sen}\, t$ y calculamos las derivadas necesarias:

$$y_p' = -A\,\text{sen}\, t + B\cos t$$
$$y_p'' = -A\cos t - B\,\text{sen}\, t$$

y entonces

$$\begin{aligned} 2\cos t &= y_p'' - 3y_p' + 2y_p \\ &= -A\cos t - B\operatorname{sen} t - 3(-A\operatorname{sen} t + B\cos t) + 2(A\cos t + B\operatorname{sen} t) \\ &= (A-3B)\cos t + (3A+B)\operatorname{sen} t \end{aligned}$$

Dado que $\cos t$ y $\operatorname{sen} t$ son funciones linealmente independientes, existe una única combinación lineal de ellas que da la función $2\cos t$. Esto justifica que podamos igualar los coeficientes respectivos y plantear el sistema de ecuaciones

$$\begin{aligned} A - 3B &= 2 \\ 3A + B &= 0 \end{aligned}$$

de donde deducimos que $A = \frac{1}{5}$ y $B = -\frac{3}{5}$; la solución particular buscada es $y_p(t) = \frac{1}{5}\cos t - \frac{3}{5}\operatorname{sen} t$ y la solución general resulta

$$y = \frac{1}{5}\cos t - \frac{3}{5}\operatorname{sen} t + C_1 e^t + C_2 e^{2t} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

Los ejemplos anteriores ilustran los recursos que usualmente se usan cuando el término independiente de la ecuación diferencial es un polinomio, una exponencial, un seno o un coseno. Creemos que deberían ser claros e indicar cómo proceder en cada situación. De todos modos, dejamos asentada una expresión general del método, cuya comprobación puede buscarse en la bibliografía.

Proposición 7.1 (Método de los coeficientes indeterminados)

Si en la ecuación diferencial

$$y'' + a_1 y' + a_0 y = f(t)$$

definida en 7.2 la función $f(t)$ es de la forma

$$f(t) = q(t)e^{\alpha t}\cos(\beta t) \qquad \text{ó} \qquad f(t) = q(t)e^{\alpha t}\operatorname{sen}(\beta t)$$

donde q es un polinomio a coeficientes reales y $\alpha, \beta \in \mathbb{R}$, entonces existe una solución particular de la forma

$$y_p(t) = t^s e^{\alpha t}\left(q_1(t)\cos(\beta t) + q_2(t)\operatorname{sen}(\beta t)\right)$$

donde q_1 y q_2 son polinomios a coeficientes reales de grado menor o igual que el de q y s es la multiplicidad de $\alpha + i\beta$ como raíz de la ecuación característica (en particular, $s = 0$ si $\alpha + i\beta$ no es raíz).

Dejamos como ejercicio para el lector comprobar que todos los ejemplos anteriores se pueden interpretar como casos particulares de esta proposición.

N Una pregunta que naturalmente surge en este contexto es cómo encontrar una solución particular cuando no se cumplen las condiciones del método de los coeficientes indeterminados, válido tan sólo para una clase reducida de funciones. Aunque no lo expondremos en este curso, mencionamos que existe un método general conocido como *variación de parámetros*. Remitimos a los lectores interesados a la bibliografía de ecuaciones diferenciales.

Ejercicio.

7.8 Resuelva los siguientes problemas de valores iniciales

$$\begin{cases} y'' - 6y' + 9y = e^{3t} \\ y(0) = 1 \\ y'(0) = -1 \end{cases} \qquad \begin{cases} y'' - 9y = t^2 e^t \\ y(0) = 1 \\ y'(0) = 0 \end{cases}$$

7.7 Sistemas de ecuaciones diferenciales

Para presentar este tema, comenzaremos planteando un modelo de crecimiento de una población (entendida en sentido amplio: podría ser de humanos, animales, plantas...). Una primera aproximación al problema sería considerar que el crecimiento es proporcional a la cantidad de individuos que dicha población tiene. Adoptando la notación $x(t)$ para representar la cantidad de individuos en función del tiempo, la ecuación diferencial

$$x'(t) = kx(t)$$

representa nuestro modelo: x' da cuenta de la variación de individuos por unidad de tiempo, que resulta proporcional a la cantidad de individuos que la población efectivamente tiene.[4] La constante de proporcionalidad k es un dato empírico de cada población. La solución general de esta ecuación es

$$x(t) = Ce^{kt} \qquad \text{con } C \in \mathbb{R}$$

y la primera objeción que puede hacerse a este modelo es que la población tiende a infinito conforme avanza el tiempo (si $k > 0$). Evidentemente, se trata de un modelo poco realista, que no tiene en cuenta las limitaciones con que se encuentra una población creciente. Un modelo un poco más sofisticado es el que admite las poblaciones $x_1(t)$ y $x_2(t)$ de dos especies que interactúan. Las ecuaciones

$$\begin{aligned} x_1'(t) &= a_{11}x_1(t) + a_{12}x_2(t) \\ x_2'(t) &= a_{21}x_1(t) + a_{22}x_2(t) \end{aligned}$$

(donde $a_{11}, a_{12}, a_{21}, a_{22} \in \mathbb{R}$) reflejan esta interacción, puesto que el cambio de cada población depende de las dos especies. Consideremos el caso en que $x_1(t)$ es una población de ratones y $x_2(t)$ de lechuzas: estamos frente a un modelo de *predador-presa*. La cantidad de ratones tenderá a aumentar proporcionalmente a si misma, pero disminuirá proporcionalmente a la cantidad de lechuzas que los cazan. Vale decir que deberá ser $a_{11} > 0$ y $a_{12} < 0$. A su vez, la población de lechuzas deberá crecer proporcionalmente a la cantidad de ratones disponibles para alimentarse. Por otra parte, cuanto mayor sea la población de lechuzas menor disponibilidad de alimento tendrá para cada individuo. Esto permite suponer que será $a_{21} > 0$ y $a_{22} < 0$. Para ilustrar este modelo, consideramos el sistema de ecuaciones diferenciales[5]

$$\begin{aligned} x_1'(t) &= 1x_1(t) - 2x_2(t) \\ x_2'(t) &= 2x_1(t) - 4x_2(t) \end{aligned}$$

[4] Cualquier población tiene una cantidad entera de individuos. Al pensarla como una función derivable $x(t)$ que puede asumir valores fraccionarios estamos haciendo una simplificación que nos permite aplicar los recursos del cálculo.

[5] El lector debe estar al tanto de todas las simplificaciones que se han hecho en este modelo y, por ende, reconocer sus limitaciones. Los modelos biológicos más sofisticados exceden los propósitos de este ejemplo.

Para resolver este sistema lo representamos matricialmente

$$\underbrace{\begin{bmatrix} x_1'(t) \\ x_2'(t) \end{bmatrix}}_{X'(t)} = \underbrace{\begin{bmatrix} 1 & -2 \\ 2 & -4 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} x_1(t) \\ x_2(t) \end{bmatrix}}_{X(t)}$$

y realizamos una diagonalización de la matriz A:

$$A = \underbrace{\begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} -3 & 0 \\ 0 & 0 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} -1/3 & 2/3 \\ 2/3 & -1/3 \end{bmatrix}}_{P^{-1}}$$

Entonces el sistema original puede escribirse como

$$X' = AX = PDP^{-1}X$$

Multiplicando a izquierda por P^{-1} queda

$$P^{-1}X' = DP^{-1}X$$

Dado que P^{-1} es una matriz que contiene constantes, las reglas usuales de derivación nos permiten escribir

$$\left(P^{-1}X\right)' = D\left(P^{-1}X\right)$$

De esta forma, el cambio de variable $Y(t) = P^{-1}X(t)$ transforma al sistema original en un sistema *desacoplado*

$$Y' = DY \Leftrightarrow \begin{bmatrix} y_1'(t) \\ y_2'(t) \end{bmatrix} = \begin{bmatrix} -3 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} -3y_1(t) \\ 0 \end{bmatrix} \Leftrightarrow \begin{cases} y_1'(t) & = -3y_1(t) \\ y_2'(t) & = 0 \end{cases}$$

donde cada función incógnita se encuentra en una ecuación diferencial lineal sin vínculo con la otra. La resolución en este caso es

$$\begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} = \begin{bmatrix} C_1 e^{-3t} \\ C_2 \end{bmatrix} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

y para conocer la solución del problema original deshacemos el cambio de variables:

$$X(t) = PY(t) = \begin{bmatrix} 1 & 2 \\ 2 & 1 \end{bmatrix} \begin{bmatrix} C_1 e^{-3t} \\ C_2 \end{bmatrix} = \begin{bmatrix} C_1 e^{-3t} + 2C_2 \\ 2C_1 e^{-3t} + C_2 \end{bmatrix} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

o sea que

$$x_1(t) = C_1 e^{-3t} + 2C_2$$
$$x_2(t) = 2C_1 e^{-3t} + C_2$$

Las constantes $C_1, C_2 \in \mathbb{R}$ quedan determinadas por las condiciones iniciales. Si, por ejemplo, la población inicial de ratones cuenta con 500 individuos y la de lechuzas con 400, al evaluar la solución general en $t = 0$ queda el sistema de ecuaciones

$$C_1 + 2C_2 = 500$$
$$2C_1 + C_2 = 400$$

de donde se deduce que $C_1 = 100$ y $C_2 = 200$ y por ende

$$x_1(t) = 100e^{-3t} + 400$$
$$x_2(t) = 200e^{-3t} + 200$$

El modelo predice que a medida que pasa el tiempo, las poblaciones de ratones y lechuzas tenderán a estabilizarse en 400 y 200 individuos respectivamente. Más aún, en la solución general general se aprecia que, para cualquier condición inicial, la población de ratones tenderá a duplicar a la de lechuzas.

Ejercicio.

7.9 Considere el siguiente sistema de ecuaciones diferenciales

$$\begin{aligned} x_1'(t) &= 3x_1(t) - x_2(t) \\ x_2'(t) &= -2x_1(t) + 2x_2(t) \end{aligned}$$

- Encuentre la solución que verifica las condiciones iniciales $x_1(0) = 50 \wedge x_2(0) = 250$.
- ¿Qué tipo de relación podría estar representando este sistema? Dé un ejemplo.

La definición que sigue generaliza el problema que hemos estudiado en el ejemplo.

Definición 7.5

Un sistema de ecuaciones diferenciales de la forma

$$\begin{cases} x_1'(t) &= a_{11}x_1(t) + \ldots + a_{1n}x_n(t) + f_1(t) \\ &\vdots \\ x_n'(t) &= a_{n1}x_1(t) + \ldots + a_{nn}x_n(t) + f_n(t) \end{cases}$$

donde las funciones derivables $x_i(t)$ son las incógnitas, los coeficientes $a_{ij} \in \mathbb{R}$ son constantes y las $f_i(t)$ son funciones continuas en cierto intervalo abierto I, se denomina ***sistema de ecuaciones diferenciales lineales con coeficientes constantes***.

Adoptando la notación

$$X(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix} \quad X'(t) = \begin{bmatrix} x_1'(t) \\ \vdots \\ x_n'(t) \end{bmatrix} \quad A = \begin{bmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{n1} & \cdots & a_{nn} \end{bmatrix} \quad F(t) = \begin{bmatrix} f_1(t) \\ \vdots \\ f_n(t) \end{bmatrix}$$

el sistema puede escribirse como

$$X' = AX + F$$

En el caso de que sea $F(t) = 0$ para todo $t \in I$ se dice que es un sistema ***homogéneo***.

Si además se buscan soluciones que verifiquen una condición de la forma $X(t_0) = X_0$ para ciertos $t_0 \in I$ y $X_0 \in \mathbb{R}^n$, se tiene un ***problema de valores iniciales***.

Como ya hicimos en el caso de la ecuación de segundo orden, comenzaremos estudiando el caso homogéneo para luego pasar al general.

7.8 Sistemas homogéneos

En los textos de ecuaciones diferenciales se demuestra que cualquier sistema homogéneo en las condiciones que estamos estudiando tiene una solución general, que dicha solución es un subespacio del espacio de funciones $\mathscr{C}^1(I, \mathbb{R}^n)$ y que la

dimensión de ese subespacio es n. Nuestro estudio se limitará al caso en que la matriz es diagonalizable, para el que aplicaremos los mismos recursos que en el ejemplo de la sección anterior.

Teorema 7.9

Dado un sistema de ecuaciones lineales homogéneo

$$X' = AX$$

con las condiciones definidas en 7.5, si existe una base $\{v_1, ..., v_n\}$ de $\mathbb{R}^n$ compuesta por autovectores de A de modo tal que $Av_i = \lambda_i v_i$ con $i = 1, ..., n$ y $\lambda_i \in \mathbb{R}$, entonces

$$\left\{e^{\lambda_1 t} v_1, ..., e^{\lambda_n t} v_n\right\}$$

es una base del espacio de soluciones del sistema.

Demostración:
Planteamos en primer lugar la diagonalización de la matriz de coeficientes del sistema: $A = PDP^{-1}$ donde

$$P = \begin{bmatrix} v_1 & v_2 & \cdots & v_n \end{bmatrix} \qquad D = \begin{bmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_n \end{bmatrix}$$

Al reemplazar en el sistema obtenemos

$$X' = PDP^{-1}X \Leftrightarrow \left(P^{-1}X\right)' = D\left(P^{-1}X\right)$$

con lo cual el cambio de variable $Y = P^{-1}X$ conduce al sistema

$$Y' = DY \Leftrightarrow \begin{cases} y_1'(t) &= \lambda_1 y_1(t) \\ \vdots & \vdots \\ y_n'(t) &= \lambda_n y_n(t) \end{cases}$$

Dado que las ecuaciones están desacopladas, la solución se calcula inmediatamente como

$$Y(t) = \begin{bmatrix} C_1 e^{\lambda_1 t} \\ \vdots \\ C_n e^{\lambda_n t} \end{bmatrix} \qquad \text{con } C_1, ..., C_n \in \mathbb{R}$$

Para conocer la solución del sistema deshacemos el cambio de variables:

$$X(t) = PY(t) = P \begin{bmatrix} C_1 e^{\lambda_1 t} \\ \vdots \\ C_n e^{\lambda_n t} \end{bmatrix} = C_1 e^{\lambda_1 t} v_1 + ... + C_n e^{\lambda_n t} v_n \qquad \text{con } C_1, ..., C_n \in \mathbb{R}$$

Esto prueba que el conjunto $\left\{e^{\lambda_1 t} v_1, ..., e^{\lambda_n t} v_n\right\}$ genera todas las soluciones. Para probar su independencia lineal, planteamos la combinación lineal que da la función nula:

$$C_1 e^{\lambda_1 t} v_1 + ... + C_n e^{\lambda_n t} v_n = 0 \qquad \forall t \in I$$

Ahora bien, como los vectores $v_1, \ldots, v_n$ son independientes, los coeficientes deben ser nulos:

$$C_1 e^{\lambda_1 t} = 0$$
$$\vdots$$
$$C_n e^{\lambda_n t} = 0$$

y dado que las funciones exponenciales no se anulan, se deduce que $C_1 = \ldots = C_n = 0$, como queríamos. □

Ejemplo 7.9

Resolveremos el siguiente sistema de ecuaciones diferenciales

$$\begin{aligned} x_1'(t) &= -x_1(t) + x_2(t) \\ x_2'(t) &= x_1(t) + 2x_2(t) - 3x_3(t) \\ x_3'(t) &= x_1(t) + x_2(t) - 2x_3(t) \end{aligned}$$

cuya matriz de coeficientes

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

es diagonalizable y sus autoespacios son

$$S_{-2} = \operatorname{gen}\{[-1\ 1\ 1]^t\} \qquad S_1 = \operatorname{gen}\{[1\ 2\ 1]^t\} \qquad S_0 = \operatorname{gen}\{[1\ 1\ 1]^t\}$$

Por lo tanto, la solución general es

$$X(t) = C_1 e^{-2t} \begin{bmatrix} -1 \\ 1 \\ 1 \end{bmatrix} + C_2 e^t \begin{bmatrix} 1 \\ 2 \\ 1 \end{bmatrix} + C_3 \begin{bmatrix} 1 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -C_1 e^{-2t} + C_2 e^t + C_3 \\ C_1 e^{-2t} + 2C_2 e^t + C_3 \\ C_1 e^{-2t} + C_2 e^t + C_3 \end{bmatrix}$$

donde $C_1, C_2, C_3 \in \mathbb{R}$.

Ejercicio.

7.10 Considere el siguiente sistema de ecuaciones diferenciales

$$\begin{aligned} x_1'(t) &= x_2(t) - x_3(t) \\ x_2'(t) &= 2x_2(t) - 2x_3(t) \\ x_3'(t) &= x_2(t) - x_3(t) \end{aligned}$$

y encuentre la solución general.

De manera similar a lo que había ocurrido en el caso de las ecuaciones diferenciales de segundo orden, es posible que la matriz de coeficientes del sistema tenga autovalores complejos. En tal caso, vale un teorema como 7.9 admitiendo autovalores complejos y autovectores en $\mathbb{C}^n$. La demostración es idéntica a la que hemos hecho para el caso real. Como en las ecuaciones de segundo orden, será de interés encontrar una base de soluciones en $\mathbb{R}$.

Ejemplo 7.10

Resolvamos el sistema $X' = AX$ con

$$A = \begin{bmatrix} 4 & -3 \\ 3 & 4 \end{bmatrix}$$

Los autovalores son dos números complejos conjugados: $\lambda_1 = 4 - 3i$ y $\lambda_2 = 4 + 3i$. Los autoespacios asociados son

$$S_{\lambda_1} = \text{gen}\{[1 \;\; i]^t\} \qquad S_{\lambda_2} = \text{gen}\{[1 \;\; -i]^t\}$$

y por lo tanto la solución general del sistema es

$$X(t) = C_1 e^{(4-3i)t} \begin{bmatrix} 1 \\ i \end{bmatrix} + C_2 e^{(4+3i)t} \begin{bmatrix} 1 \\ -i \end{bmatrix}$$

donde $C_1, C_2 \in \mathbb{C}$. Para encontrar funciones de $I \to \mathbb{R}$ que generen la solución general, desarrollamos la expresión obtenida:

$$\begin{aligned} X(t) &= C_1 e^{4t} [\cos(3t) - i \operatorname{sen}(3t)] \begin{bmatrix} 1 \\ i \end{bmatrix} + C_2 e^{4t} [\cos(3t) + i \operatorname{sen}(3t)] \begin{bmatrix} 1 \\ -i \end{bmatrix} \\ &= e^{4t} \begin{bmatrix} C_1 \cos(3t) + C_2 \cos(3t) \\ C_1 \operatorname{sen}(3t) + C_2 \operatorname{sen}(3t) \end{bmatrix} + i e^{4t} \begin{bmatrix} -C_1 \operatorname{sen}(3t) + C_2 \operatorname{sen}(3t) \\ C_1 \cos(3t) - C_2 \cos(3t) \end{bmatrix} \\ &= (C_1 + C_2) e^{4t} \begin{bmatrix} \cos(3t) \\ \operatorname{sen}(3t) \end{bmatrix} + i(C_1 - C_2) e^{4t} \begin{bmatrix} -\operatorname{sen}(3t) \\ \cos(3t) \end{bmatrix} \end{aligned}$$

y podemos entonces decir que todas las soluciones de $I \to \mathbb{R}$ son de la forma

$$X(t) = D_1 e^{4t} \begin{bmatrix} \cos(3t) \\ \operatorname{sen}(3t) \end{bmatrix} + D_2 e^{4t} \begin{bmatrix} -\operatorname{sen}(3t) \\ \cos(3t) \end{bmatrix}$$

con $D_1, D_2 \in \mathbb{R}$.

Queremos destacar que las dos soluciones independientes

$$\Phi_1(t) = e^{4t} \begin{bmatrix} \cos(3t) \\ \operatorname{sen}(3t) \end{bmatrix} \qquad \Phi_2(t) = e^{4t} \begin{bmatrix} -\operatorname{sen}(3t) \\ \cos(3t) \end{bmatrix}$$

son las partes real e imaginaria, respectivamente, de la solución

$$e^{4t} [\cos(3t) - i \operatorname{sen}(3t)] \begin{bmatrix} 1 \\ i \end{bmatrix}$$

obtenida a partir de λ_1.

Al final del ejemplo anterior señalamos que las partes real e imaginaria de una solución eran a su vez dos soluciones reales. Este hecho no es una casualidad sino que puede generalizarse adaptando el desarrollo hecho en el ejemplo.

Teorema 7.10

Dado un sistema de ecuaciones lineales homogéneo

$$X' = AX$$

con las condiciones definidas en 7.5, si la matriz A tiene un autovalor complejo $\lambda = a+bi$ (con $b \neq 0$) con un autovector asociado v entonces

1. Las funciones a valores complejos

$$\Phi_1(t) = e^{\lambda t} v \qquad \text{y} \qquad \Phi_2(t) = e^{\overline{\lambda} t} \overline{v}$$

son dos soluciones independientes del sistema.

2. $\Phi_2(t) = \overline{\Phi_1(t)}$
3. Las partes real e imaginaria de $\Phi_1(t)$ son soluciones a valores reales del sistema y resultan linealmente independientes.

Demostración:
Veamos que $\Phi_1(t)$ es solución:

$$\Phi_1'(t) = \lambda e^{\lambda t} v = e^{\lambda t} Av = A\left(e^{\lambda t} v\right) = A\Phi_1(t)$$

El caso de $\Phi_2(t)$ se basa en el hecho (estudiado en el capítulo 4) de que para matrices reales el conjugado de un autovalor también es autovalor, con autovector conjugado:

$$A\overline{v} = \overline{Av} = \overline{\lambda v} = \overline{\lambda}\,\overline{v}$$

y se demuestra de manera análoga. La independencia lineal surge de plantear la combinación que da la función nula

$$C_1\Phi_1(t) + C_2\Phi_2(t) = 0 \Leftrightarrow C_1 e^{\lambda t} v + C_2 e^{\overline{\lambda} t}\overline{v} = 0$$

y recordar que a autovalores distintos corresponden autovectores independientes, con lo cual deben ser

$$C_1 e^{\lambda t} = C_2 e^{\overline{\lambda} t} = 0$$

y necesariamente $C_1 = C_2 = 0$.

La propiedad 2 es una consecuencia inmediata de las propiedades de la conjugación y se deja como ejercicio.

Para demostrar 3, señalamos que por ser el conjunto solución del sistema homogéneo un subespacio, cualquier combinación lineal

$$C_1\Phi_1(t) + C_2\Phi_2(t) = C_1\Phi_1(t) + C_2\overline{\Phi_1(t)} \tag{7.18}$$

será solución. Recordando entonces que las parte real e imaginaria de un número complejo z surgen de

$$\operatorname{Re} z = \frac{z+\overline{z}}{2} \qquad \operatorname{Im} z = \frac{z-\overline{z}}{2i}$$

podemos elegir en la expresión (7.18) las combinaciones lineales con $C_1 = C_2 = \frac{1}{2}$:

$$\frac{\Phi_1(t) + \overline{\Phi_1(t)}}{2} = \operatorname{Re}\Phi_1(t)$$

y con $C_1 = \frac{1}{2i}$, $C_2 = \frac{-1}{2i}$:

$$\frac{\Phi_1(t) - \overline{\Phi_1(t)}}{2i} = \operatorname{Im}\Phi_1(t)$$

La independencia lineal de $\operatorname{Re}\Phi_1(t)$ e $\operatorname{Im}\Phi_1(t)$ se deduce de la independencia de Φ_1 y Φ_2 y se dejan los detalles como ejercicio. □

Ejercicio.

7.11 Encuentre una base del espacio de soluciones del siguiente sistema de ecuaciones diferenciales

$$x_1'(t) = 2x_1(t) - x_2(t)$$
$$x_2'(t) = x_1(t) + 2x_2(t)$$

7.9 Sistemas no homogéneos

Al comenzar a estudiar sistemas de ecuaciones diferenciales vimos que si la matriz de coeficientes es diagonalizable, entonces un cambio de variable permite resolver el caso homogéneo. Veremos ahora que ese mismo cambio resuelve el caso no homogéneo

$$X' = AX + F$$

como fue definido en 7.5. Para ello, consideramos nuevamente la factorización $A = PDP^{-1}$ y reemplazamos en el sistema

$$X' = PDP^{-1}X + F$$

Al multiplicar a izquierda por P^{-1} resulta el sistema equivalente

$$\left(P^{-1}X\right)' = D\left(P^{-1}X\right) + P^{-1}F$$

Con el cambio de variable $Y = P^{-1}X$ resulta

$$Y' = DY + P^{-1}F$$

Como $P^{-1}F$ es un vector de funciones conocidas, lo que queda por resolver son n ecuaciones desacopladas en las incógnitas $y_1(t), \ldots, y_n(t)$, para luego deshacer el cambio de variable.

Ejemplo 7.11

Resolver el sistema de ecuaciones

$$X' = \underbrace{\begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}}_{A} X(t) + \underbrace{\begin{bmatrix} e^t \\ 0 \end{bmatrix}}_{F(t)}$$

Los autovalores de la matriz A son $\lambda_1 = 1$ y $\lambda_2 = 3$. Sus correspondientes autoespacios son

$$S_1 = \operatorname{gen}\left\{[1\ 1]^t\right\} \qquad S_3 = \operatorname{gen}\left\{[1\ \ -1]^t\right\}$$

Por lo tanto

$$A = \underbrace{\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}}_{P} \underbrace{\begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix}}_{P^{-1}}$$

Entonces y como ya habíamos anticipado, el sistema original puede escribirse en la forma

$$Y' = DY + P^{-1}F \Leftrightarrow \begin{bmatrix} y_1'(t) \\ y_2'(t) \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} \begin{bmatrix} y_1(t) \\ y_2(t) \end{bmatrix} + \begin{bmatrix} 1/2 & 1/2 \\ 1/2 & -1/2 \end{bmatrix} \begin{bmatrix} e^t \\ 0 \end{bmatrix}$$

y llegamos a las dos ecuaciones desacopladas

$$y_1'(t) = y_1(t) + \frac{1}{2}e^t$$
$$y_2'(t) = 3y_2(t) + \frac{1}{2}e^t$$

Se trata de ecuaciones diferenciales lineales, cuyas soluciones son

$$y_1(t) = C_1 e^t + \frac{te^t}{2}$$
$$y_2(t) = C_2 e^{3t} - \frac{e^t}{4}$$

con $C_1, C_2 \in \mathbb{R}$. Para conocer la solución del problema original deshacemos el cambio de variables:

$$X(t) = PY(t) = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} C_1 e^t + \frac{te^t}{2} \\ C_2 e^{3t} - \frac{e^t}{4} \end{bmatrix} = \begin{bmatrix} C_1 e^t + C_2 e^{3t} + \frac{(2t-1)e^t}{4} \\ C_1 e^t - C_2 e^{3t} + \frac{(2t+1)e^t}{4} \end{bmatrix} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

Podemos reescribir la solución como

$$X(t) = C_1 e^t \begin{bmatrix} 1 \\ 1 \end{bmatrix} + C_2 e^{3t} \begin{bmatrix} 1 \\ -1 \end{bmatrix} + \begin{bmatrix} \frac{(2t-1)e^t}{4} \\ \frac{(2t+1)e^t}{4} \end{bmatrix} \qquad \text{con } C_1, C_2 \in \mathbb{R}$$

para apreciar que la solución consta, como era de esperarse, de la solución general del problema homogéneo $X' = AX$ sumada con una solución particular.[a]

[a] Los lectores interesados en conocer la generalización de este comentario podrán consultar la bibliografía.

Ejercicio.

7.12 Encuentre la solución general del siguiente sistema de ecuaciones diferenciales

$$x_1' = x_1 - 3x_2$$
$$x_2' = -3x_1 + x_2 - e^t$$

Soluciones

Soluciones del Capítulo 1

1.1 Lo primero que observamos es que las operaciones que se han definido son cerradas, porque tanto el producto de números positivos como las potencias de un número positivo son números positivos. Con respecto a las propiedades de la suma y el producto, es sencillo comprobarlas usando los resultados conocidos sobre números reales. Así, por ejemplo, para comprobar la propiedad 5 consideramos

$$(\alpha+\beta)(x_1,x_2)=\left(x_1^{(\alpha+\beta)},x_2^{(\alpha+\beta)}\right)=\left(x_1^{\alpha}x_1^{\beta},x_2^{\alpha}x_2^{\beta}\right)=\left(x_1^{\alpha},x_2^{\alpha}\right)+\left(x_1^{\beta},x_2^{\beta}\right)=\alpha\,(x_1,x_2)+\beta\,(x_1,x_2)$$

Por último, el elemento neutro para la suma es $0_V=(1,1)$ y el opuesto del vector $u=(x_1,x_2)$ es $-u=(x_1^{-1},x_2^{-1})$.

1.2 Planteando la combinación lineal $w=\alpha_1 v_1+\alpha_2 v_2$ resulta $(1+i,2)=\alpha_1(1,0)+\alpha_2(0,1)=(\alpha_1,\alpha_2)$. Esto implica que necesariamente $\alpha_1=1+i\wedge\alpha_2=2$.

1.3 Considerando la equivalencia trigonométrica $\operatorname{sen}(a+b)=\operatorname{sen}a\cos b+\cos a\operatorname{sen}a$ resulta para todo $x\in\mathbb{R}$:

$$\operatorname{sen}\left(x+\frac{\pi}{4}\right)=\operatorname{sen}\left(\frac{\pi}{4}+x\right)=\operatorname{sen}\left(\frac{\pi}{4}\right)\cos x+\cos\left(\frac{\pi}{4}\right)\operatorname{sen}x=\frac{\sqrt{2}}{2}\cos x+\frac{\sqrt{2}}{2}\operatorname{sen}x$$

Por lo tanto $f=\frac{\sqrt{2}}{2}f_1+\frac{\sqrt{2}}{2}f_2$.

1.4 Como se trata de una doble implicación, supondremos en primer lugar que los vectores $\{v_1,v_2,...,v_r\}$ constituyen un conjunto linealmente dependiente. Esto significa que existen escalares $\alpha_1,\alpha_2=...=\alpha_r$ no todos nulos tales que $\alpha_1 v_1+\alpha_2 v_2+...+\alpha_r v_r=0_V$. Podemos suponer, sin perder generalidad, que $\alpha_1\neq 0$. Entonces

$$\alpha_1 v_1=-\alpha_2 v_2-...-\alpha_r v_r\Rightarrow v_1=-\frac{\alpha_2}{\alpha_1}v_2-...-\frac{\alpha_r}{\alpha_1}v_r$$

y esto exhibe un vector que es combinación lineal de los demás.
Recíprocamente, si un vector es combinación lineal de los demás, podemos suponer una vez más que es $v_1=\beta_2 v_2+...+\beta_r v_r$. Pero entonces se tiene una combinación lineal no trivial que da cero:

$$0_V=(-1)v_1+\beta_2 v_2+...+\beta_r v_r$$

y se trata de un conjunto linealmente dependiente.

1.5 $W\left(f_1,f_2,f_3\right)(x)=2x$, con lo cual basta con elegir $x\neq 0$ y aplicar el teorema 1.3.

1.6 Es el enunciado contrarrecíproco del teorema 1.3, y por lo tanto es equivalente.

1.7 Una alternativa es demostrar que es un conjunto generador y aplicar el teorema 1.8. Para ello alcanza con mostrar que los polinomios de la base canónica pueden obtenerse como combinación lineal de $1-t$, $1+t$ y t^2.

1.8 Por ser rectas distintas, existen dos vectores linealmente independientes u, v tales que $S_1 = \text{gen}\{u\}$ y $S_2 = \text{gen}\{v\}$. Por lo tanto $S_1 + S_2$ contiene a todas las combinaciones lineales de u y v. Luego si las rectas están en $\mathbb{R}^2$ es $S_1 + S_2 = \mathbb{R}^2$, mientras que si las rectas están en $\mathbb{R}^3$, $S_1 + S_2$ es el plano que las contiene.

1.9 El plano S_1 admite una base $\{u, v\}$, mientras que la recta S_2 admite una base $\{w\}$. El vector w no es combinación lineal de u y v (¿por qué?). Luego el conjunto $\{u, v, w\}$ es un base de $\mathbb{R}^3$. De esta forma, para cada vector $v \in \mathbb{R}^3$ existen y son únicos los escalares α, β, γ tales que

$$v = \alpha u + \beta v + \gamma w = \underbrace{\left(\alpha u + \beta v\right)}_{\in S_1} + \underbrace{\gamma w}_{\in S_2}$$

1.10 $B_1 = \left\{t^2 - 1, t^3 - t\right\}$, $B_2 = \left\{t^3 + 8\right\}$ y $B_3 = \left\{2t^3 + t^2 + 15\right\}$ son respectivamente bases de S_1, S_2 y S_3. No se verifica la ecuación (1.18), puesto que $\dim(S_1 + S_2 + S_3) = 3$ mientras que $\dim(S_1) + \dim(S_2) + \dim(S_3) = 4$ (probarlo).

Soluciones del Capítulo 2

2.1 Es posible que para algunos lectores no sea clara la definición de la función f. Cuando eso ocurre, antes de lanzarse a la resolución del planteo propiamente dicho, es recomendable hacer un ejemplo sencillo de uso de la definición. En este caso, consideramos el polinomio $p(t) = 3t^2 - 5$. Entonces la transformación consiste en evaluar al polinomio en cero: $f(p) = p(0) = -5$.
Ahora sí, para demostrar que es una transformación verificamos que se cumplen las dos condiciones.
❶ dados $p, q \in \mathscr{P}_2 : f(p+q) = (p+q)(0) = p(0) + q(0) = f(p) + f(q)$ (donde hemos usado la definición de suma en $\mathscr{P}_2$).
❷ dados $p \in \mathscr{P}_2, \alpha \in \mathbb{R} : f(\alpha p) = (\alpha p)(0) = \alpha\big(p(0)\big) = \alpha f(p)$ (donde hemos usado la definición de producto por escalar en $\mathscr{P}_2$).

2.2 Alcanza con aplicar el teorema 2.2 a $f(V)$ y $f^{-1}(0_W)$.

2.3

- Una posibilidad es $f[(x_1, x_2)] = (x_1 + x_2, x_2)$. ¿Hay otra posibilidad?
- Una posibilidad es $f[(x_1, x_2)] = (x_1, 0)$. ¿Hay otra posibilidad?

2.4

- Para que f sea un isomorfismo, debe cumplirse que $\dim(\text{Nu}\, f) = 0$ y a la vez que $\text{Im}\, f = W$, con lo cual el teorema de la dimensión queda para este caso $\dim(V) = 0 + \dim(W)$.
- $\text{Im}\, f \subseteq W$ y por ende $\dim(\text{Im}\, f) \leq \dim(W)$. En el teorema de la dimensión queda

  $$\dim(V) = \dim(\text{Nu}\, f) + \dim(\text{Im}\, f) \leq \dim(\text{Nu}\, f) + \dim(W)$$

  con lo cual, teniendo en cuenta que $\dim(V) > \dim(W) \Leftrightarrow \dim(V) - \dim(W) > 0$

  $$0 < \dim(V) - \dim(W) \leq \dim(\text{Nu}\, f)$$

  y el núcleo no es trivial, por lo que f no puede ser un monomorfismo.

Soluciones del Capítulo 3

3.1 Dadas matrices arbitrarias $A, B, C \in \mathbb{R}^{m\times n}$ y $k \in \mathbb{R}$, comprobamos que se verifican las condiciones de la definición.
❶ $(A, B+C) = \operatorname{tr}\left(A^t(B+C)\right) = \operatorname{tr}\left(A^tB + A^tC\right) = \operatorname{tr}\left(A^tB\right) + \operatorname{tr}\left(A^tC\right) = (A,B) + (A,C)$
❷ $(kA, B) = \operatorname{tr}\left((kA)^tB\right) = \operatorname{tr}\left(kA^tB\right) = k\operatorname{tr}\left(A^tB\right) = k\,(A,B)$
Para las dos últimas condiciones apelamos al hecho de que $\operatorname{tr}\left(A^tB\right) = \sum a_{ij}b_{ij}$ donde $1 \le i \le m \ \wedge\ 1 \le j \le n$.
❸ $(A, B) = \operatorname{tr}\left(A^tB\right) = \operatorname{tr}\left(B^tA\right) = (B, A)$
❹ $(A, A) = \operatorname{tr}\left(A^tA\right) \ge 0 \ \wedge\ (A, A) = 0 \Leftrightarrow A = 0$

3.2 En el espacio $\mathscr{C}([0,1])$ es $(f,g) = \frac{1}{2}$ y $\|g\| = \frac{\sqrt{3}}{3}$. En el espacio $\mathscr{C}([-1,1])$ es $(f,g) = 0$ y $\|g\| = \frac{\sqrt{6}}{3}$.

3.3 $(u, v) = -4 + 4i$, $\|u\| = \sqrt{6}$, $\|v\| = \sqrt{14}$.

3.4 La igualdad surge de aplicar la definición de norma junto con las propiedades del teorema 3.1:

$$\begin{aligned}\|u+v\|^2 + \|u-v\|^2 &= (u+v, u+v) + (u-v, u-v)\\ &= (u,u) + (v,u) + (u,v) + (v,v) + (u,u) - (u,v) - (v,u) + (v,v)\\ &= 2\|u\|^2 + 2\|v\|^2\end{aligned}$$

Si u y v son vectores de $\mathbb{R}^2$, determinan un paralelogramo cuyas diagonales se corresponden con $u+v$ y $u-v$, por lo que se puede decir que "la suma de los cuadrados de las diagonales de un paralelogramo es igual a la suma de los cuadrados de los lados". (Notar que en el caso de los rectángulos se deduce inmediatamente del teorema de Pitágoras).

3.5 Las dos primeras son inmediatas a partir de la definición y las propiedades de la norma. En cuanto a la tercera, es una consecuencia de la desigualdad triangular 3.4:

$$\mathrm{d}(u,v) = \|u-v\| = \|(u-w)+(w-v)\| \le \|u-w\| + \|w-v\| = \mathrm{d}(u,w) + \mathrm{d}(w,v)$$

3.6 Desarrollando $\|u+v\|^2$ como en la demostración de la desigualdad triangular 3.4:

$$\|u+v\|^2 = \|u\|^2 + \|v\|^2 + 2\operatorname{Re}(u,v)$$

Como se supone que $\|u+v\|^2 = \|u\|^2 + \|v\|^2$, la cancelación arroja

$$\operatorname{Re}(u,v) = 0$$

En el caso de un $\mathbb{R}$ espacio (y sólo en ese caso), concluimos que $(u,v) = 0$.

3.7 Como ya se probó en un ejercicio anterior, las funciones son ortogonales para el producto interno definido. Como además no son nulas, constituyen un conjunto linealmente independiente de tres elementos en $\mathscr{P}_2$ y por lo tanto son una base ortogonal. Como las normas no son uno (comprobarlo), la base no es ortonormal.

3.8 $P_S(v) = \left(\frac{2}{3}, -\frac{1}{3}, -\frac{1}{3}\right)$, $\mathrm{d}(v,S) = \frac{1}{\sqrt{3}}$.

3.9 $(\operatorname{Fil} A)^\perp = \operatorname{Nul}\overline{A}$, $(\operatorname{Col} A)^\perp = \operatorname{Nul} A^H$

3.10 $A = \begin{bmatrix} -1/\sqrt{2} & 1/\sqrt{3} & -1/\sqrt{6} \\ 1/\sqrt{2} & 1/\sqrt{3} & -1/\sqrt{6} \\ 0 & 1/\sqrt{3} & \sqrt{6}/3 \end{bmatrix} \begin{bmatrix} \sqrt{2} & \sqrt{2} & -\sqrt{2} \\ 0 & \sqrt{3} & 1/\sqrt{3} \\ 0 & 0 & \sqrt{6}/3 \end{bmatrix}$

3.11 Un paso previo a la construcción de la matriz es obtener una base ortonormal: $B = \left\{ \frac{1}{\sqrt{2}} [1\ 0\ -1\ 0]^t, \frac{1}{\sqrt{6}} [1\ 2\ 1\ 0]^t \right\}$. Luego

$$[P_S]_E = \frac{1}{3} \begin{bmatrix} 2 & 1 & -1 & 0 \\ 1 & 2 & 1 & 0 \\ -1 & 1 & 2 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

3.12 $y = \frac{1}{60} x^2 + \frac{37}{300} x + \frac{53}{50}$

3.13

- $v - R(v) = \left(P_S(v) + P_{S^\perp}(v)\right) - \left(P_S(v) - P_{S^\perp}(v)\right) = 2P_{S^\perp}(v) \in S^\perp$.
- $\frac{1}{2}(v + R(v)) = \frac{1}{2}\left(P_S(v) + P_{S^\perp}(v) + P_S(v) - P_{S^\perp}(v)\right) = P_S(v)$.
- $\mathrm{d}(v, S) = \left\| P_{S^\perp}(v) \right\| = \left\| -P_{S^\perp}(v) \right\| = \mathrm{d}(R(v), S)$.

Soluciones del Capítulo 4

4.1

- $([R]_B)^n = [R]_B$ si n es impar, mientras que $([R]_B)^n = I$ si n es par. Geométricamente: al aplicar dos veces la misma reflexión a un vector, se obtiene el vector original.
- $\det [R]_E = \det (C_{BE} [R]_B C_{EB}) = \det C_{BE} \det [R]_B \det (C_{BE})^{-1} = \det [R]_B$

4.2 $S_3 = \mathrm{gen}\{[1\ 1]^t\}$; $S_1 = \mathrm{gen}\{[-1\ 1]^t\}$; $m_3 = m_1 = 1$. Geométricamente, la transformación produce una dilatación de factor 3 en la dirección del autovector $[1\ 1]^t$, mientras que en la dirección perpendicular los vectores quedan fijos. Se puede ilustrar considerando que el cuadrado de vértices $[0\ 0]^t, [1\ 1]^t, [0\ 2]^t, [-1\ 1]^t$ se transforma en un rectángulo (¿cuál?).

4.3 La verificación de las propiedades es inmediata. En cuanto a la semejanza, si A fuera semejante a I, existiría una matriz inversible P tal que $A = PIP^{-1} = I$ lo cual es absurdo. Por lo tanto, las propiedades enunciadas en el teorema 4.2 no son suficientes para que exista semejanza entre matrices.

4.4

$$P = \begin{bmatrix} i & -i \\ 1 & 1 \end{bmatrix} \wedge D = \begin{bmatrix} \frac{1+i\sqrt{3}}{2} & 0 \\ 0 & \frac{1-i\sqrt{3}}{2} \end{bmatrix} \wedge P^{-1} = \frac{1}{2} \begin{bmatrix} -i & 1 \\ i & 1 \end{bmatrix}$$

Esta no es la única posibilidad. Se puede alterar, por ejemplo, el orden de los autovalores y usar múltiplos (no nulos) de los autovectores.

4.5 $p_A(t) = -t^3 + t^2 + 33t + 63 = -(t-7)(t+3)^2$. Llamando $\lambda_1 = 7$, el autoespacio asociado es $S_{\lambda_1} = \mathrm{gen}\{[1\ 1\ 0]^t\}$. Para $\lambda_2 = -3$, tenemos que $m_{\lambda_2} = 2$ y

$$\mathrm{Nul}(A - (-3)I) = \mathrm{Nul} \begin{bmatrix} 5 & 5 & 3 \\ 5 & 5 & c \\ 0 & 0 & 0 \end{bmatrix} = \mathrm{Nul} \begin{bmatrix} 5 & 5 & 3 \\ 0 & 0 & c-3 \\ 0 & 0 & 0 \end{bmatrix}$$

Por ende, si $c = 3$ el autoespacio asociado es $S_{\lambda_2} = \mathrm{gen}\{[-3\ 0\ 5]^t, [-1\ 1\ 0]^t\}$ y será $\mu_{\lambda_2} = m_{\lambda_2} = 2$, con lo cual A será diagonalizable. Si $c \neq 3$ el autoespacio asociado a λ_2 será de dimensión 1, es decir que $\mu_{\lambda_2} < m_{\lambda_2}$ y la matriz no será diagonalizable.

4.6 La interpretación geométrica como composición de rotaciones permite anticipar que

$$A^n = \begin{bmatrix} \cos(n\theta) & -\operatorname{sen}(n\theta) \\ \operatorname{sen}(n\theta) & \cos(n\theta) \end{bmatrix}$$

Para demostrarlo formalmente, conviene estudiar primero el caso $n = 2$, que se demuestra usando las equivalencias trigonométricas $\operatorname{sen}(\alpha+\beta) = \operatorname{sen}\alpha\cos\beta + \cos\alpha\operatorname{sen}\beta$ y $\cos(\alpha+\beta) = \cos\alpha\cos\beta - \operatorname{sen}\alpha\operatorname{sen}\beta$.

4.7

$$A = PDP^{-1} \quad \text{con} \quad P = \begin{bmatrix} 3 & 1 \\ 1 & 0 \end{bmatrix} \quad \text{y} \quad D = \begin{bmatrix} 1 & 0 \\ 0 & 2/3 \end{bmatrix}$$

Luego los vectores $v \in \mathbb{R}^2$ pueden escribirse en la forma

$$v = \alpha \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \beta \begin{bmatrix} 1 \\ 0 \end{bmatrix}$$

para α y β adecuados. Por lo tanto, *para cualquier* $v \in \mathbb{R}^2$

$$A^k v = \alpha A^k \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \beta A^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} = \alpha \begin{bmatrix} 3 \\ 1 \end{bmatrix} + \beta \left(\frac{2}{3}\right)^k \begin{bmatrix} 1 \\ 0 \end{bmatrix} \to \alpha \begin{bmatrix} 3 \\ 1 \end{bmatrix}$$

a medida que $k \to \infty$.

4.8 De $\dim[\operatorname{Nul}(A-3I)] = 2$ se deduce que 3 es un autovalor de multiplicidad geométrica 2. Se sabe que para el polinomio $p(t) = t^2+2t$ resulta $p(A)$ singular, es decir que un autovalor de $p(A)$ es cero (¿por qué?). Entonces, de acuerdo con el teorema 4.7, existe un autovalor λ de A tal que $p(\lambda) = 0$. Esto es, $p(\lambda) = \lambda^2+2\lambda = \lambda(\lambda+2) = 0$. Ahora bien, λ no puede ser cero porque A es inversible, luego $\lambda = -2$. En resumen, los autovalores de A son -2 y 3, este último de multiplicidad algebraica y geométrica doble. Luego es una matriz diagonalizable.

4.9 Si A es diagonalizable existe una base $\{v_1, v_2, \ldots v_n\}$ de autovectores. De acuerdo con (1) estos mismos autovectores serán autovectores de $p(A)$ y por ende $p(A)$ será diagonalizable.

Soluciones del Capítulo 5

5.1 Por tratarse de la matriz asociada a una reflexión con respecto a un plano, resulta $S_1 = \operatorname{gen}\{[1\ 1\ -2]^t, [1\ 0\ 1]^t\}$ y $S_{-1} = \operatorname{gen}\{[-1\ 3\ 1]^t\}$. Una base ortonormal de autovectores es

$$B = \left\{ \frac{1}{\sqrt{6}}[1\ 1\ -2]^t, \frac{1}{\sqrt{66}}[7\ 1\ 4]^t, \frac{1}{\sqrt{11}}[-1\ 3\ 1]^t \right\}$$

y una diagonalización posible es

$$A = \begin{bmatrix} 1/\sqrt{6} & 7/\sqrt{66} & -1/\sqrt{11} \\ 1/\sqrt{6} & 1/\sqrt{66} & 3/\sqrt{11} \\ -2/\sqrt{6} & 4/\sqrt{66} & 1/\sqrt{11} \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix} \begin{bmatrix} 1/\sqrt{6} & 1/\sqrt{6} & -2/\sqrt{6} \\ 7/\sqrt{66} & 1/\sqrt{66} & 4/\sqrt{66} \\ -1/\sqrt{11} & 3/\sqrt{11} & 1/\sqrt{11} \end{bmatrix}$$

5.2 En general, no se cumple la relación $Q(x+y) = Q(x)+Q(y)$ ni $Q(kx) = kQ(x)$. (Sin embargo, hay una matriz A para la cual $Q(x) = x^t Ax$ define una transformación lineal: ¿cuál es?).

5.3

- $Q(x) = 2x_1^2 + x_1x_2 - 2x_1x_3 + 3x_2^2 + 3x_3^2$
- $A = \begin{bmatrix} 2 & 3/2 & -1/2 \\ 3/2 & 4 & -1 \\ -1/2 & -1 & -1 \end{bmatrix}$

5.4 Con el cambio de variables como en el ejemplo 5.8 queda la ecuación $y_1^2 + 3y_2^2 = k$. Es decir que se trata de elipses si $k > 0$, un punto si $k = 0$ y un conjunto vacío si $k < 0$.

5.5

$$A = \begin{bmatrix} 1 & -4 \\ -4 & -5 \end{bmatrix} = \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & 1 \\ -1 & 2 \end{bmatrix} \cdot \begin{bmatrix} 3 & 0 \\ 0 & -7 \end{bmatrix} \cdot \frac{1}{\sqrt{5}} \begin{bmatrix} 2 & -1 \\ 1 & 2 \end{bmatrix}$$

Con el cambio de variable $x = Py$ la forma es $\tilde{Q}(y) = 3y_1^2 - 7y_2^2$. Por lo tanto, los conjuntos de nivel $\mathcal{N}_1(Q)$, $\mathcal{N}_{-1}(Q)$ son hipérbolas (con distintos ejes) y $\mathcal{N}_0(Q)$ es un par de rectas.

5.6 Definida positiva, semidefinida positiva, indefinida, definida negativa, semidefinida negativa.

5.7 Como A es simétrica y sus autovalores son 7, 2 y -3, tenemos que $Q(x)$ es indefinida.

5.8 Como $Q(x) = x^t Ax$ con $A = \begin{bmatrix} 1 & c \\ c & 1 \end{bmatrix}$, basta con determinar para qué valores de c los autovalores de A son positivos. Estudiando la ecuación $\det(A - tI) = 0$ se deduce que debe ser $|c| < 1$.

5.9 El valor mínimo es -7 y se realiza en $x = \pm \left[\frac{3}{2\sqrt{5}} \ \frac{6}{5\sqrt{5}}\right]^t$.

5.10 (Para resolver el problema se usa el cuadrado de la distancia, es decir la norma). La distancia mínima al origen es $\frac{1}{\sqrt{8}}$ y se realiza en los vectores $x = \pm \left[\frac{1}{4} \ \frac{1}{4}\right]^t$. La distancia máxima al origen es $\frac{1}{2}$ y se realiza en los vectores $x = \pm \frac{1}{\sqrt{8}} [-1 \ 1]^t$.

5.11 (La restricción ya fue estudiada en el ejemplo 5.12). El máximo es 2 y se realiza en los vectores $x = \pm\sqrt{2}\,[1 \ 1]^t$. El mínimo es $-\frac{2}{3}$ y se realiza en los vectores $x = \pm\sqrt{\frac{2}{3}}\,[-1 \ 1]^t$.

5.12 La forma cuadrática se puede escribir como $Q(x) = (3x_1 - x_2)^2$. Luego los conjuntos de nivel $\mathcal{N}_k = \{x \in \mathbb{R}^2 / Q(x) = k\}$ están vacíos si $k < 0$, son un par de rectas si $k > 0$, una recta si $k = 0$. Se trata de una forma semidefinida positiva. El valor mínimo sujeto a la restricción $\|x\|^2 = 1$ es cero y se alcanza en los vectores $\pm\frac{1}{\sqrt{10}}\,[1 \ 3]^t$; el máximo es 10 y se alcanza en los vectores $\pm\frac{1}{\sqrt{10}}\,[-3 \ 1]^t$.

Soluciones del Capítulo 6

6.1 Es una consecuencia de que $\operatorname{Nul} A = \operatorname{Nul}\left(A^t A\right)$ como probamos en el teorema 3.17.

6.2 Los autovalores de $A^t A$ son $\lambda_1 = 360$, $\lambda_2 = 90$ y $\lambda_3 = 0$. Luego los valores singulares de A son $\sigma_1 = \sqrt{360} = 6\sqrt{10}$, $\sigma_2 = \sqrt{90} = 3\sqrt{10}$ y $\sigma_3 = 0$. Los autoespacios de $A^t A$ son

$$S_{\lambda_1} = \operatorname{gen}[1 \ 2 \ 2]^t \qquad S_{\lambda_2} = \operatorname{gen}[-2 \ -1 \ 2]^t \qquad S_{\lambda_3} = \operatorname{gen}[2 \ -2 \ 1]^t$$

Luego $B = \{v_1, v_2, v_3\} = \left\{\frac{1}{3}[1 \ 2 \ 2]^t, \frac{1}{3}[-2 \ -1 \ 2]^t, \frac{1}{3}[2 \ -2 \ 1]^t\right\}$ es la base ortonormal buscada. Dado que $Av_3 = 0$ una base de $\operatorname{Nul} A$ es $\{v_3\}$. Una base ortonormal para $\operatorname{Col} A = \mathbb{R}^2$ es

$$\left\{\frac{Av_1}{\sigma_1}, \frac{Av_2}{\sigma_2}\right\} = \left\{\frac{1}{\sqrt{10}}[3 \ 1]^t, \frac{1}{\sqrt{10}}[1 \ -3]^t\right\}$$

6.3 De acuerdo con las observaciones hechas en los ejemplos 6.1 y 6.2, los vectores definidos en (6.4) constituyen una base ortonormal de $\mathbb{R}^2$ formada por autovectores de A^tA, mientras que una base ortonormal de $\operatorname{Col} A$ es la que ya se describió en la ecuación (6.5). En coordenadas $\begin{bmatrix} y_1 & y_2 \end{bmatrix}^t$ con respecto a esta base, los transformados de la circunferencia unitaria verifican la ecuación pedida.

6.4 $A^tA = \begin{bmatrix} 3 & -3 \\ -3 & 3 \end{bmatrix}$ con los autoespacios $S_6 = \operatorname{gen}\{[-1\ 1]^t\}$ y $S_0 = \operatorname{gen}\{[1\ 1]^t\}$. Razonando como en el ejemplo 6.2, se deduce que los vectores de $\mathbb{R}^2$ con norma uno se transformarán en vectores del subespacio generado por $A[-1\ 1]^t$. Más aún, describirán un segmento de longitud $2\sqrt{6}$.

6.5 Pueden adaptarse los argumentos del ejemplo 6.3 para concluir que en el primer caso es una matriz de rango 3 y T transforma a la esfera unitaria en un elipsoide, mientras que en el segundo caso es una matriz de rango 2 y T transforma a la esfera unitaria en una región plana de borde elíptico.

6.6

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 6\sqrt{10} & 0 & 0 \\ 0 & 3\sqrt{10} & 0 \end{bmatrix} \frac{1}{3} \begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & 2 \\ 2 & -2 & 1 \end{bmatrix}$$

6.7 La base ortonormal de $\mathbb{R}^2$

$$\left\{ \frac{1}{\sqrt{10}} [3\ \ 1]^t, \frac{1}{\sqrt{10}} [1\ \ -3]^t \right\}$$

es también una base de $\operatorname{Col} A$. Es inmediato que $P_{\operatorname{Col} A}(v) = v$.
Usando la base ortonormal de $\mathbb{R}^3$ $\left\{ \frac{1}{3}[1\ \ 2\ \ 2]^t, \frac{1}{3}[-2\ \ -1\ \ 2]^t, \frac{1}{3}[2\ \ -2\ \ 1]^t \right\}$ se obtienen las bases

$$\left\{ \frac{1}{3}[1\ \ 2\ \ 2]^t, \frac{1}{3}[-2\ \ -1\ \ 2]^t \right\} \qquad \text{y} \qquad \left\{ \frac{1}{3}[2\ \ -2\ \ 1]^t \right\}$$

de $\operatorname{Fil} A$ y $\operatorname{Nul} A$ respectivamente. $P_{\operatorname{Fil} A}(w) = \frac{1}{9}[-1\ \ 1\ \ 4]^t$.

6.8

$$A = \begin{bmatrix} 4 & 11 & 14 \\ 8 & 7 & -2 \end{bmatrix} = \begin{bmatrix} \frac{3}{\sqrt{10}} & \frac{1}{\sqrt{10}} \\ \frac{1}{\sqrt{10}} & -\frac{3}{\sqrt{10}} \end{bmatrix} \begin{bmatrix} 6\sqrt{10} & 0 \\ 0 & 3\sqrt{10} \end{bmatrix} \frac{1}{3} \begin{bmatrix} 1 & 2 & 2 \\ -2 & -1 & 2 \end{bmatrix}$$

6.9

$$A^\dagger = \begin{bmatrix} -1/180 & 13/180 \\ 1/45 & 2/45 \\ 1/18 & -1/18 \end{bmatrix}$$

6.10 El sistema es compatible determinado y $x^\dagger = A^\dagger b = \begin{bmatrix} \frac{1}{15} & \frac{1}{15} & 0 \end{bmatrix}^t$ es la única solución.

Soluciones del Capítulo 7

7.1 $\phi : (0, +\infty) \to \mathbb{R}$, $\phi(t) = \frac{1}{\operatorname{senh} t}\left(1 - \frac{1}{2}t^2\right)$

7.2

- $y(t) = 3t^2 + Ct$ con $C \in \mathbb{R}$

- $\operatorname{Nu} L = \operatorname{gen}\{t\}$.

7.3 $y(t) = -\frac{1}{8}\left(1+2t+2t^2\right) + \frac{5}{8}e^{2t} + \frac{1}{2}e^{-2t}$

7.4 $y(t) = \frac{1}{7}e^{-4t} + \frac{6}{7}e^{3t}$

7.5 $y(t) = 2te^{-2t} + e^{-2t}$

7.6 El primer problema admite como solución cualquier función de la forma $y = C_2 \operatorname{sen}(4t)$, mientras que el segundo no tiene solución.

7.7 $y(t) = e^t\left(2\cos(2t) - \frac{3}{2}\sin(2t)\right)$

7.8 $y(t) = e^{3t} - 4te^{3t} + \frac{1}{2}t^2e^{3t}$; $y(t) = \frac{97}{192}e^{-3t} + \frac{13}{24}e^{3t} - \frac{3}{64}e^t - \frac{1}{16}te^t - \frac{1}{8}t^2e^t$

7.9 $x_1(t) = -50e^{4t} + 100e^t$, $x_2(t) = 50e^{4t} + 200e^t$

7.10 $X(t) = C_1e^t\begin{bmatrix}1\\2\\1\end{bmatrix} + C_2\begin{bmatrix}0\\1\\1\end{bmatrix} + C_3\begin{bmatrix}1\\0\\0\end{bmatrix} = \begin{bmatrix}C_1e^t + C_3\\2C_1e^t + C_2\\C_1e^t + C_2\end{bmatrix}$ donde $C_1, C_2, C_3 \in \mathbb{R}$ (destacamos que el autoespacio asociado al cero es de dimensión dos, por lo que existen muchas alternativas para escribir la solución general).

7.11 Una base posible es $\{\Phi_1, \Phi_2\}$, donde $\Phi_1(t) = e^{2t}\begin{bmatrix}-\operatorname{sen} t\\ \cos t\end{bmatrix}$ y $\Phi_2(t) = e^{2t}\begin{bmatrix}\cos t\\ \operatorname{sen} t\end{bmatrix}$.

7.12 $X(t) = \begin{bmatrix}-C_1e^{4t} + C_2e^{-2t} - \frac{e^t}{3}\\ C_1e^{4t} + C_2e^{-2t}\end{bmatrix}$ con $C_1, C_2 \in \mathbb{R}$

Índice alfabético

Esta página fue dejada intencionalmente en blanco

Bibliografía

[1] Juan de Burgos, *Álgebra Lineal*, McGraw-Hill, 1996.

[2] C. Henry Edwards, *Ecuaciones Diferenciales y problemas con valores en la frontera*, Pearson Educación, cuarta edición, 2009.

[3] Stanley Grossman, *Álgebra Lineal con Aplicaciones*, McGraw-Hill, tercera edición, 1990.

[4] Kenneth Hoffman y Ray Kunze, *Álgebra Lineal*, Prentice Hall, 1984.

[5] David Lay, *Álgebra Lineal y sus Aplicaciones*, Addison Wesley Longman, 1999.

[6] Ben Noble y James Daniel, *Álgebra Lineal Aplicada*, Prentice Hall, tercera edición, 1989.

[7] David Poole, *Álgebra Lineal: una introducción moderna*, Thomson, 2004.

[8] Jesús Rojo, *Álgebra Lineal*, McGraw-Hill, 2001.

[9] Gilbert Strang, *Álgebra Lineal y sus Aplicaciones*, Addison Wesley Iberoamericana, 1989.

[10] Dennis G. Zill, *Ecuaciones Diferenciales con aplicaciones de modelado*, Cengage Learning, novena edición, 2009.